Microeconometrics

Isabel Casas
Office: V5-206a-2
icasas@sam.sdu.dk

What is Econometrics?

Econometrics (From Economics and -metrics). It is part of the economics science which applies mathematical and statistical techniques to economic theories to verify them and to solve economic problems through models.

1. Economic model (constructed from observation or mathematical deduction)
2. Data
3. Methodology
4. Software
5. Economic understanding and an open mind

Difference between Statistics and Econometrics

- Economics generates problems that do not appear in other fields
- Special statistical methodologies applied to these problems
- Analysis of financial data
- Macroeconomic analysis
- Microeconomic problems

Empirical analysis

- Causality
- Types of economical data
- The multivariate linear model

What is a causal relationship?

Causal relationship: One variable causes changes in another variable

$$\text{Correlation} \neq \text{Causality}$$

Example 1

- A zip code of a person(X) may give us an idea of his/her wealth(Y).
- Are X and Y correlated?

What is a causal relationship?

Causal relationship: One variable causes changes in another variable

$$\text{Correlation} \neq \text{Causality}$$

Example 1

- A zip code of a person(X) may give us an idea of his/her wealth(Y).
- Are X and Y correlated?
- Does the zip code imply a bulgy bank account?

What is a causal relationship?

Causal relationship: One variable causes changes in another variable

$$\text{Correlation} \neq \text{Causality}$$

Example 1

- A zip code of a person(X) may give us an idea of his/her wealth(Y).
- Are X and Y correlated?
- Does the zip code imply a bulgy bank account?
- X does not imply Y
- The two variables are correlated but there is not a causal relationship between them.

What is a causal relationship?

Causal relationship: One variable causes changes in another variable

$$\text{Correlation} \neq \text{Causality}$$

Example 2

- There is a positive correlation between wine quality(Y) and a hot August (X).
- Are X and Y correlated?

What is a causal relationship?

Causal relationship: One variable causes changes in another variable

$$\text{Correlation} \neq \text{Causality}$$

Example 2

- There is a positive correlation between wine quality(Y) and a hot August (X).
- Are X and Y correlated?
- Does the quantity of sun during August implies a good quality wine?

What is a causal relationship?

Causal relationship: One variable causes changes in another variable

$$\text{Correlation} \neq \text{Causality}$$

Example 2

- There is a positive correlation between wine quality(Y) and a hot August (X).
- Are X and Y correlated?
- Does the quantity of sun during August implies a good quality wine?
- X implies Y: causal relationship

What is a causal relationship?

Causal relationship: One variable causes changes in another variable

$$\text{Correlation} \neq \text{Causality}$$

Example 2

- There is a positive correlation between wine quality(Y) and a hot August (X).
- Are X and Y correlated?
- Does the quantity of sun during August implies a good quality wine?
- X implies Y: causal relationship
- Does a good quality wine implies a good summer?
- One should ask the right question

What is a causal relationship?

- Most of the knowledge we deal with is correlational knowledge.
- As econometricians, we want to find causal knowledge: cause–effect.
- The conditional expectation is our tool.

Econometric process?

1. Data
2. Model
3. Estimation
4. Testing
5. Interpretation

Data structures

1. Cross–sectional data
2. Time series data
3. Panel data

Cross–sectional data

- It is a data set that consists of a unities such as individuals, homes, companies, cities, prices, etc.
- Commonly the data of each unity does not correspond to the exact same time period.
- They are obtained as a random sample of certain population.
- For example: A sample of 500 people can be taken from the employed population. A sample of their salaries, education and experience.

Cross sectional data example

Observaction	wages $ per hour	Years of education	Years of experience	Status (married=1) another=0)	Sex (F=1/M=0)
1	3.10	11	2	1	1
2	3.24	12	6	0	1
3	6	14	7	1	0
⋮	⋮	⋮	⋮	⋮	⋮
580	4.10	8	10	1	0

Cross sectional data

- However, sometime the random sample strategy is not a good way to go to obtain the cross–sectional data.
- For example: we are interested in studying the factors that influence the accumulation of family wealth. A random sample of families might not be a good idea. It could happen that the richest families do not want to inform on their patrimony. Therefore the resulting sample would not be a random sample of the families population.
- The cross sectional data is commonly used in economics: labor economics, public finance, industrial organisation, urban economy, demographic economics, health economics...

Cross sectional data

- Sometime, in the cross sectional data, different variables correspond to different time periods.
- For example: To determine the effects of certain governmental policies in the long term economical growth, economists have studied the relationship between the growth of the GDP during the time period 1960–1985 and certain variables which were determined by the policy in 1960 (percentage of GDP dedicated to government expenditure and secondary education rates of adults).

Time series, out of the scope of this course

- Time series data set consists of a variable or a set of variables along a time span. For example, asset prices, GDP, yearly number of homicides, yearly number of traffic accidents, etc.
- Differently to the cross sectional data, one cannot assume that the observations of time series are independent. For example, the GDP of the last trimester is going to give an approximation of the GDP that is expected in this trimester.

Time series data

- The frequency of the time series data is one of their important characteristics. The most common frequency in economics are daily, montly, trimester, yearly...
- High frequency data from high frequency trading that can happen at intervals of nanoseconds.
- Example: asset prices (daily, seconds), exchange rate of euro with respect to the dollar (daily), inflation and unemployment (monthly), GDP (three months), child mortality (yearly).
- Time series data tend to show a stationarity which is an important factor in their analysis.

Example of time series data

Observation	years	Minimum salary average year ($)	Unemployment rate yearly	Natality rate
1	1950	45	15.4	4.8
2	1951	89	16	3.6
⋮	⋮	⋮	⋮	⋮
38	1987	320	16.8	2.5

Panel data

- Panel data sets consists of time series for each cross sectional unity.
- Example: Data containing salary, job type and education for a number of people during the last ten years.
- Example: Financial data of investments of a number of companies during a period of 5 years.

Panel data

- Panel data records contain the same unities during a determined period of time.
- The best way to organise these data is to show the same type of observation for a given unity of the same year.

Panel data example

Observation	City	Year	Mean $ loss	% criminal in population
1	1	1998	900000	1.5
2	1	1999	1200000	2.1
3	2	1998	600000	5.6
4	2	1999	700000	7.1
⋮	⋮	⋮	⋮	⋮
199	100	1998	20000000	2.8
200	100	1999	25000000	2.5

Panel data

- Having several observation of the same unity allows us to control over certain characteristics of these unities. This can facility the causality inference.
- Another advantage of these type of data is that it allows to study the importance of delays in the decision making or delays in behaviours.
- Its importance in economic has been recognised lately.
- These type of data can also be used as a cross sectional data.

Data resources

- http://elsa.berkeley.edu
- http://www.fedstats.gov
- http://www.bls.gov/data/
- http://www.who.int/research/en/
- http://www.unece.org/stats/trends/
- "http://www.sdu.dk/Bibliotek/Soegning/Databaser"

Examples of models

- $\mathbf{y}$ is a random vector: *explained variable*
- $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ is a $1 \times k$ random vector of *explanatory variables*

If $E(|\mathbf{y}|) < \infty$ then there is a function:

$$E(\mathbf{y}|\mathbf{X}) = \mu(\mathbf{X})$$

the expected value of $\mathbf{y}$ given the values $\mathbf{X}$.

We can determine how this average value changes when $\mathbf{X}$ changes.

Review of multivariate linear regression (MLR)

- Structural model
- OLS estimator
- Mean and variance of the estimator
- OLS estimator in R.

MLR

- The multivariate linear regression is an extension of the simple regression for k explanatory variables.
- In general, the response variable or endogenous variable $\mathbf{y}$ depends of other variables $\mathbf{x}_1, \ldots, \mathbf{x}_k$,
- although some of those can be unobservable or even unknown to the econometrician.
- The linear model includes the effect of the most important ones.
- Naturally, the more useful factors included in the model will result in a better explanation of the variation of y and therefore a better prediction of its expected value.

Ceteris paribus

"Holding other things constant": Everything outside the economic environment that our model describes is held constant.

$E(\mathbf{y}|\mathbf{X}, \mathbf{c})$: expected value of $\mathbf{y}$ conditional to $\mathbf{X}$ and $\mathbf{c}$.

- $\mathbf{y}$ is the dependant variable, endogenous variable
- $\mathbf{X}$ is the matrix of regressors, exogenous variables
- $\mathbf{c}$ is a set of control variables

Ceteris paribus

"Holding other things constant": Everything outside the economic environment that our model describes is held constant.

$E(\mathbf{y}|\mathbf{X}, \mathbf{c})$: expected value of $\mathbf{y}$ conditional to $\mathbf{X}$ and $\mathbf{c}$.

- $\mathbf{y}$ is the dependant variable, endogenous variable
- $\mathbf{X}$ is the matrix of regressors, exogenous variables
- $\mathbf{c}$ is a set of control variables

Study the effect of $\mathbf{X}$ on the expected value of $\mathbf{y}$ once the variables $\mathbf{c}$ are kept fixed.

Model specification

Let assume that there is a linear relationship between $\mathbf{y}$ and k variables $\mathbf{x}_i,\ i = 1, \ldots, k$.

$$y_j = \beta_0 + \beta_1 x_{1j} + \ldots, \beta_k x_{kj} + \epsilon_j, \quad j = 1, \ldots, n \tag{1}$$

where

- y_j is the j–value of the response variable, dependant variable, endogenous variable or explained variable.
- x_{ij} is the j–value of the explanatory variable, exogenous variable, independent variable or regressor x_i
- $\beta_0, \beta_1, \ldots, \beta_k$ are the parameters or coefficients associated to each model variable.
- ϵ_j are the random errors which are assumed independent and normal with mean 0 and variance σ^2, $N(0, \sigma^2)$

Model specification

The relationship (1) is linear on the parameters, but it does not have to be linear on the variables. For example, there may be a square relationship of y with the variable x_2:

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{2j}^2 + \epsilon_j$$

Some other times, it might be necessary to take the logarithm of certain variables, both of the response variable and of the regressors:

$$\ln y_j = \beta_0 + \beta_1 \ln x_{1j} + \ldots + \beta_k \ln x_{kj} + \epsilon_j \quad \text{(log-log model)}$$

Q: Why?

Model specification

We refer to the conditional expected value of y_j given all $\mathbf{x}_i$:

$$E(y_j|\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) = \beta_0 + \beta_1 x_{1j} + \ldots + \beta_k x_{kj}$$

The multivariate linear model may be written in matrix form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{k1} \\ 1 & x_{12} & \ldots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \ldots & x_{kn} \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Examples of models

In our course we are going to study only *parametric models* for the conditional expectation. They are defined by a set of parameters.

For example, for $k = 2$:

$$E(\mathbf{y}|\mathbf{X}) = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 \quad \text{(linear in } \mathbf{x}_1 \text{ and } \mathbf{x}_2\text{)} \quad (2)$$

$$E(\mathbf{y}|\mathbf{X}) = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_3 \mathbf{x}_2^2 \quad \text{(nonlinear in } \mathbf{x}_2\text{)} \quad (3)$$

$$E(\mathbf{y}|\mathbf{X}) = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_3 \mathbf{x}_1 \mathbf{x}_2 \quad \text{(nonlinear in } \mathbf{x}_1 \text{ and } \mathbf{x}_2\text{)} \quad (4)$$

$$E(\mathbf{y}|\mathbf{X}) = \exp[\beta_0 + \beta_1 \log(\mathbf{x}_1) + \beta_2 \mathbf{x}_2] \quad \text{(nonlinear in } \beta_j \text{ and } \mathbf{x}_j\text{)} \quad (5)$$

Interpretation: Partial effects

The parameters of the linear model are interpret as **partial/marginal effects**:

$$\frac{\partial E(\mathbf{y}|\mathbf{X})}{\partial \mathbf{x}_1}$$

Marginal effect: The expected value of $\mathbf{y}$ when x_1 changes by 1 unit.

Exercise (5 minutes): Work with the person on your right. Find the partial effect of $\mathbf{x}_1$ on the expected value of $\mathbf{y}$ for models (1), (2), (3) and (4)

Interpretation: Partial elasticity

The *partial elasticity* of $E(Y|\mathbf{X})$ with respect to $\mathbf{x}_1$:

$$\frac{\partial E(\mathbf{y}|\mathbf{X})}{\partial \mathbf{x}_1} \cdot \frac{\mathbf{x}_1}{E(Y|\mathbf{X})}$$

Partial elasticity: The percentage change in the expected value of $\mathbf{y}$ when x_1 changes by 1%.

Q: Find elasticities of models (1), (2), (3) and (4)

Simple examples

Model	Interpretation
$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \epsilon$	If you change $\mathbf{x}_1$ by one, the expected value of $\mathbf{y}$ changes β_1
$\log(\mathbf{y}) = \beta_0 + \beta_1 \mathbf{x}_1 + \epsilon$	If you change $\mathbf{x}_1$ by one, the expected value of $\mathbf{y}$ changes $100 \times \beta_1 \%$
$\log(\mathbf{y}) = \beta_0 + \beta_1 \log(\mathbf{x}_1) + \epsilon$	If you change $\mathbf{x}_1$ by 1%, the expected value of $\mathbf{y}$ changes $\times \beta_1 \%$

Classical MLR

MLR.1 The population model is linear:

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \ldots \beta_k \mathbf{x}_k + \boldsymbol{\epsilon}$$

MLR.2 We have a random sample of n observations

$$\{(x_{1j}, x_{2j}, \ldots, x_{kj}, y_j) : j = 1, 2, \ldots, n$$

MLR.3 No perfect collinearity: none of the $\mathbf{x}$'s are constant and one cannot be written as a linear relationship of the others. The rank of $E(\mathbf{X}'\mathbf{X})$ is $k + 1$.

MLR.4 Strict exogeneneity:

$$E(\boldsymbol{\epsilon}|\mathbf{x}_1, \ldots, \mathbf{x}_k) = 0$$

MLR.5 Error term is homokedastic:

$$Var(\boldsymbol{\epsilon}|\mathbf{x}_1, \ldots, \mathbf{x}_k) = \sigma^2$$

MLR.6 The error term is independent of $\mathbf{x}_1, \ldots, \mathbf{x}_k$ and normally distributed

Least Squares

Let assume that we have two random variables such that

$$E(\mathbf{y}|\mathbf{X};\boldsymbol{\theta}) = h(\mathbf{X};\boldsymbol{\theta})$$

From a given sample $(y_1, \ldots, y_n, \mathbf{x}_1, \ldots, \mathbf{x}_n)$, the least squares estimator of θ is the value that minimises the distance between $\mathbf{y}$ and $h(\mathbf{X}, \boldsymbol{\theta})$:

$$d(\theta) = (\mathbf{y} - h(\mathbf{X}, \boldsymbol{\theta}))'(\mathbf{y} - h(\mathbf{X}, \boldsymbol{\theta}))$$

If $\mathbf{y}$ is a vector $(n \times 1)$ and $\mathbf{X}$ is a matrix $(n \times k)$, this distance can be written as a sum of squares

$$d(\boldsymbol{\theta}) = \sum_{i=1}^{n} (y_i - h(\mathbf{x}_i, \boldsymbol{\theta}))^2$$

Least Squares

- The estimators depend on the function of distance chosen. For example if $h(\cdot)$ is linear (like in the regression model), $E(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}$, then the objective function is:

$$d(\boldsymbol{\theta}) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i\boldsymbol{\theta})^2$$

  which gives us the ordinary least squares (OLS) estimator.
- The weighted least squares (WLS) estimator is found when the distance function is more general than the above. It contains a matrix $\mathbf{W}(n \times n)$ with the weights for each variable:

$$d(\boldsymbol{\theta}) = (\mathbf{y} - h(\mathbf{X}, \boldsymbol{\theta}))'\mathbf{W}(\mathbf{y} - h(\mathbf{X}, \boldsymbol{\theta}))$$

OLS Assumptions

OLS.1 $E(\mathbf{X}'\boldsymbol{\epsilon}) = 0$.

OLS.2 No perfect collinearity: none of the x's are constant and one cannot be written as a linear relationship of the others.

OLS.3 $E(\boldsymbol{\epsilon}^2\mathbf{X}'\mathbf{X}) = \sigma^2 E(\mathbf{X}'\mathbf{X})$.

If any of these assumptions are violated, the estimators of β_i and σ^2 could be biased, inconsistent or not efficient.

Estimation of by OLS

This is the same that solving a system of $k+1$ equations called *normal equations*:

$$\begin{aligned}
n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{1i} + \ldots + \hat{\beta}_k \sum_{i=1}^n x_{ki} &= \sum_{i=1}^n y_i \\
\hat{\beta}_0 \sum_{i=1}^n x_{1i} + +\hat{\beta}_1 \sum_{i=1}^n x_{1i}^2 \ldots + \hat{\beta}_k \sum_{i=1}^n x_{ki} &= \sum_{i=1}^n y_i x_{1i} \\
\ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots \\
\hat{\beta}_0 \sum_{i=1}^n x_{ki} + +\hat{\beta}_1 \sum_{i=1}^n x_{1i} \ldots + \hat{\beta}_k \sum_{i=1}^n x_{ki}^2 &= \sum_{i=1}^n y_i x_{ki}
\end{aligned} \tag{6}$$

Estimation of MLR by OLS

- The MLR has $k+2$ unknowns: the regression coefficients $\beta_0, \beta_1, \ldots, \beta_k$ and the variance of the perturbation σ^2.
- For the model of k independent variables, the estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are obtained as the values that minimise the following object function:

$$\min_{\boldsymbol{\beta}} (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \ldots \beta_k x_{ki})^2$$

5 minutes

Compare MLR assumptions with OLS assumptions

Estimation of MLR by OLS

- The fitted values $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \ldots + \hat{\beta}_k x_{ki}$,
- Define $\hat{\epsilon}_i = y_i - \hat{y}_i$ as the least square residuals,
- the OLS parameters are those that minimise $RSS = \sum_{i=1}^{n} \hat{\epsilon}_i^2 = 0$
- Also $\sum_{i=1}^{n} x_{ji}\hat{\epsilon}_i = 0$ for $j = 2, \ldots, k$.
- That is, the OLS residuals are orthogonal to all the regressors of the model.

Q: Can you put these expressions in matrix and vector form?

Estimation of MLR by OLS

The matrix form of the normal equations is:

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}^{\text{OLS}} - \mathbf{X}'\mathbf{y} = 0 \tag{7}$$

If $rank(\mathbf{X}'\mathbf{X}) = k + 1 < n$, then there is an unique solution to the system of equations (7): the ordinary least squares estimator which is the parameters vector:

$$\hat{\boldsymbol{\beta}}^{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{8}$$

Estimation of MLR by OLS

Another way to see that is obtaining the estimator by the analogy problem.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon} \Rightarrow \mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta}+\mathbf{X}'\boldsymbol{\epsilon} \Rightarrow E(\mathbf{X}'\mathbf{y}) = \boldsymbol{\beta}E(\mathbf{X}'\mathbf{X})+E(\mathbf{X}'\boldsymbol{\epsilon})$$

Then,

$$\boldsymbol{\beta} = [E(\mathbf{X}'\mathbf{X})]^{-1}E(\mathbf{X}'\mathbf{y})$$

The analogy problems says that an estimator of $\boldsymbol{\beta}$ can be found by substituting the expectations for their corresponding sample mean.

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i'y_i\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Variance estimation

If OLS.3 is true, the estimators $\hat{\beta}_i$ are efficient (in the sense that have the minimum variance amongst all the linear unbiased linear estimators – Gauss–Markov Theorem).

In fact their variance is:

$$Var(\hat{\beta}_i) = diag(\sigma^2(\mathbf{X}'\mathbf{X})^{-1})[i+1], \quad i = 0, \ldots, k) \tag{9}$$

20 minutes

```
#Understand the code
X1=rnorm(200)
X2=cos(X1)
X3 = rchisq(200, 2)
X4= rchisq(200, 4)
epsilon= rnorm(200, mean=0, sd= 0.5)
Y= 0.5*X1+ 0.3*X2+ 1.2 *X3+ epsilon

#What estimators are consistent? Reason
#Which model has smaller variance?
model1= lm (Y~-1+X1)
summary(model1)
model2= lm( Y~-1+ X1+ X2)
summary(model2)
model3= lm( Y~-1+ X1+ X2+ X3)
summary(model3)
model4<-lm(Y~-1 + X1+ X2+ X3+ X4)
summary(model4)
```

Variance estimation ($\hat{\sigma}$) by OLS

- The variance of the residuals plays a part in the variance of the estimator as it can be seen in equation (9).
- σ^2 is also unknown.
- It is estimated by the *residual variance*:

$$s^2 = \frac{\sum_{i=1}^{n} \hat{\epsilon}_i^2}{n-k-1} = \frac{SSR}{n-k-1}$$

$$\hat{\sigma}^2_{\text{OLS}} = s^2$$

Variance estimation ($\hat{\sigma}$) by OLS

Notation:

- $SST = \sum(y_i - \bar{y})^2$, total sum of squares
- $SSE = \sum(\hat{y}_i - \bar{y})^2$, the explained sum of squares
- $SSR = \sum(y_i - \hat{y}_i)^2 = \sum \epsilon_i^2$, the residual sum of squares
- $SST = SSE + SSR$: the total variation of y_i is equal to the total variation of the fitted values $\{\hat{y}_i\}$ and the total variation of the residuals $\{\hat{\epsilon}_i\}$.

Q: Is $\hat{\sigma}^2$ biased?

Q: What is the standard deviation of $\hat{\beta}_i$

Estimation by OLS

- $Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ can only be estimated
- $\hat{V} = \widehat{Var}(\hat{\boldsymbol{\beta}}) = s^2(\mathbf{X}'\mathbf{X})^{-1}$
- The standard error of $\hat{\beta}_i$ is given by:

$$se(\hat{\beta}_i) = \sqrt{diag(\hat{V})[i+1]}$$

- Note the difference between the standard deviation and the standard error of the OLS estimator.

What is R? Model fitness

- The *determination coefficient*: is denoted by R^2 and it is a measure of goodness of fit
- how well is the model fitted to the data.
- how well future outcomes are going to be predicted by the model.
- It is a number between 0 and 1, the closer to one, the better fit of the line.
- Its equation may be written as:

$$R^2 = \frac{\left(\sum_{i=1}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2}{\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)\left(\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2\right)} = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

OLS estimator properties

- As our estimator depends on the sample and therefore of the sample size, it is possible that its statistical properties are different for different sample sizes.
- There are problems where it is very difficult or impossible to determine the estimator properties for small sample sizes.
- In this cases, the asymptotic properties (when $n \to \infty$) of the estimator can help us to make a choice.

Q: Is desirable that an estimator is consistent?

Unbiased

- We say that $\hat{\boldsymbol{\beta}}$ is unbiased if $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.
- This condition is satisfied if MLR.3 is satisfied, that is if $E(\boldsymbol{\epsilon}|\mathbf{X}) = 0$ and if $rank E(\mathbf{X}'\mathbf{X}) = k + 1$.
- The OLS estimator is not necessarily unbiased if OLS.1 and OLS.2 are satisfied. However, it is if we impose MLR.4 and OLS.2.
- Some estimators might be biased but **asymptotically unbiased**. This means that their bias tends to zero as $n \to \infty$.

Q: Find $E(\hat{\boldsymbol{\beta}})$

Question:

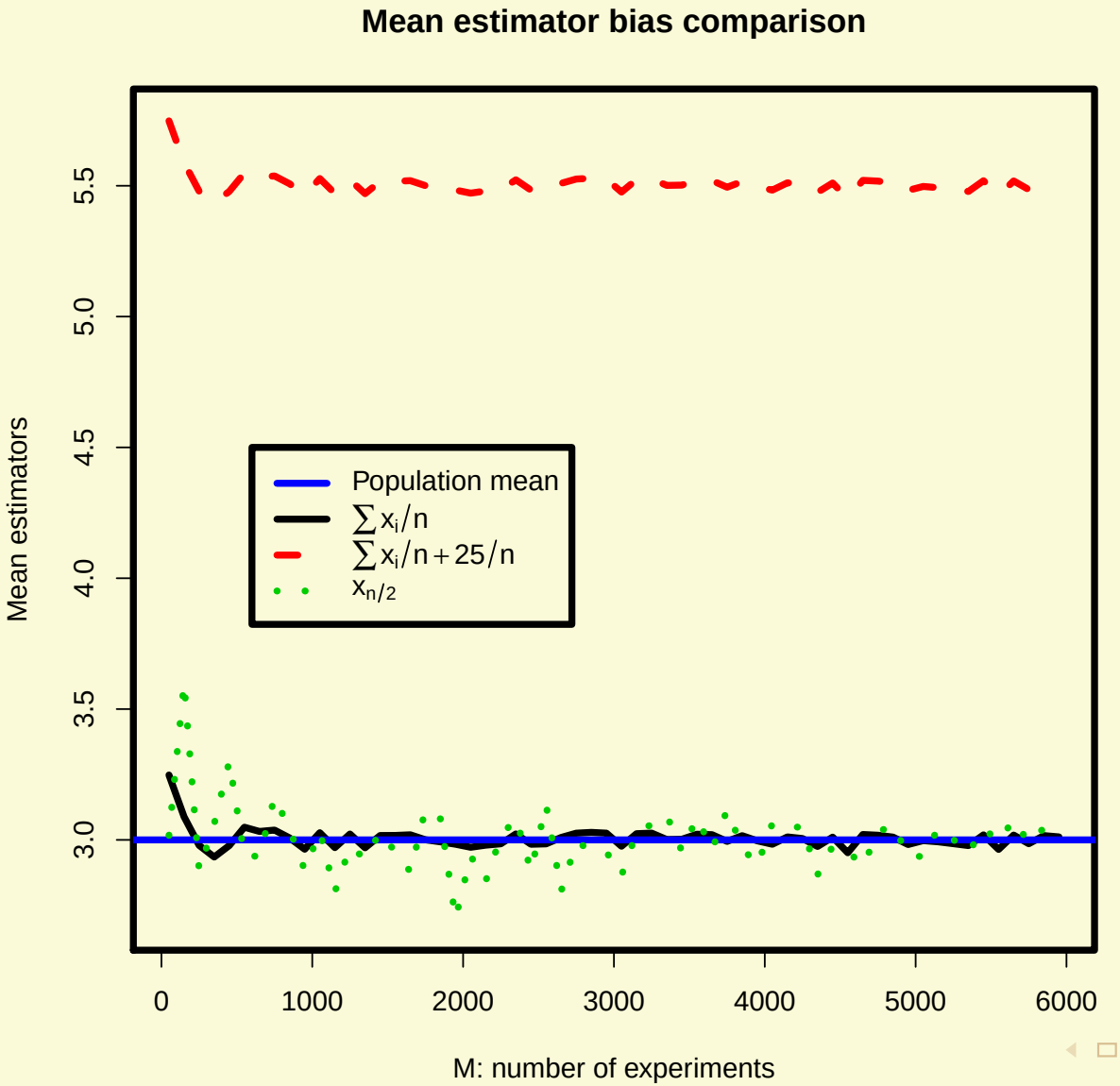
Let $\{x_1, x_2, \ldots, x_n\}$ a random sample of the variable $X \sim N(3, 25)$ and $n = 10$.

Which of the following three estimators of μ is unbiased:

1. $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{25}{n}$
3. $\hat{\mu}_3 = x_{[n/2]}$: the middle point of the sample

And asymptotically unbiased?

Solution:


Mean estimator bias comparison
Mean estimators
3.0
3.5
4.0
4.5
5.0
5.5
0
1000
2000
3000
4000
5000
6000
M: number of experiments
Population mean
$\sum x_i/n$
$\sum x_i/n + 25/n$
$x_{n/2}$

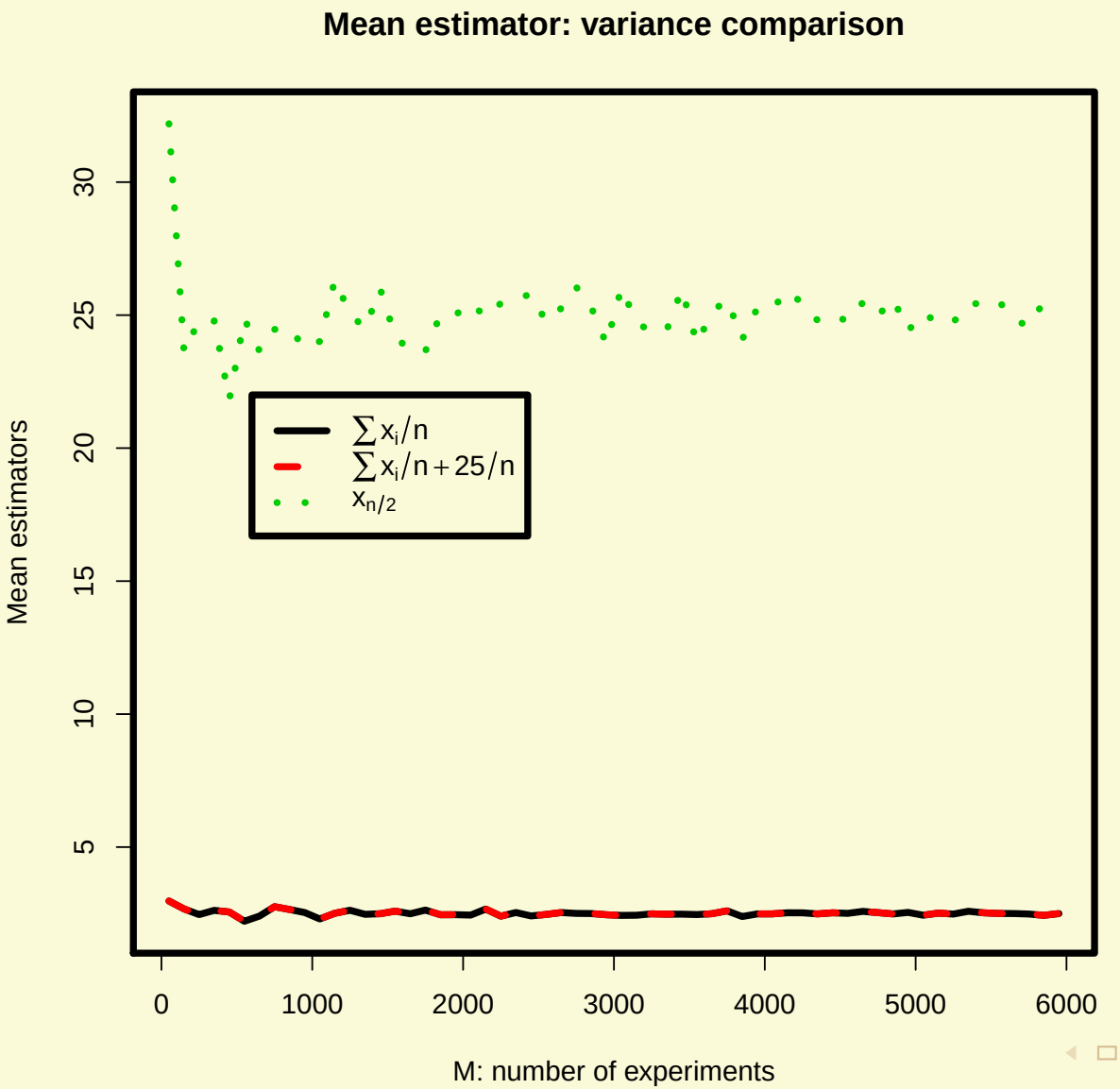
Question on variance

Let $\{x_1, x_2, \ldots, x_n\}$ a random sample of the variable $X \sim N(3, 25)$ and $n = 10$.

Which of the following three estimators have greater variance:

1. $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{25}{n}$
3. $\hat{\mu}_3 = x_{[n/2]}$: the middle point of the sample

Solution:


Mean estimator: variance comparison
Mean estimators
M: number of experiments
$\sum x_i/n$
$\sum x_i/n + 25/n$
$x_{n/2}$
5
10
15
20
25
30
0
1000
2000
3000
4000
5000
6000

Consistency

- The property of consistency assures that the estimator is, with a very high probability, very close to the real value of the parameter if the sample size if sufficiently large.
- An estimator $\hat{\theta}_n$ is consistent if and only if

$$\forall \epsilon > 0 \quad \lim_{n \to \infty} P\{|\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}| \leq \epsilon\} = 1,$$

that is, the succession of $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ (which depends on n) converges in probability to the real parameter θ. It is denoted by plim $_{n \to \infty} \hat{\boldsymbol{\theta}}_{\boldsymbol{n}} = \boldsymbol{\theta}$ or simply plim$\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} = \boldsymbol{\theta}$.

Consistency

- Intuitively, if an estimator is consistent, as the size of the sample grows, the probability that the estimation is close to the real value also grows.
- The OLS parameter $\hat{\boldsymbol{\beta}}$ is consistent, so the probability that is close to $\boldsymbol{\beta}$ is large when $n \to \infty$.
- This condition is satisfied if OLS.1 and OLS.2 are satisfied.

Consistency

How do we proof consistency?

- Using the definition of consistency, this tends to be difficult.
- Try to show that the estimator is asymptotically unbiased and its variance tends to zero as the sample size goes to ∞. This consdition is sufficient but not absolutely necessary.

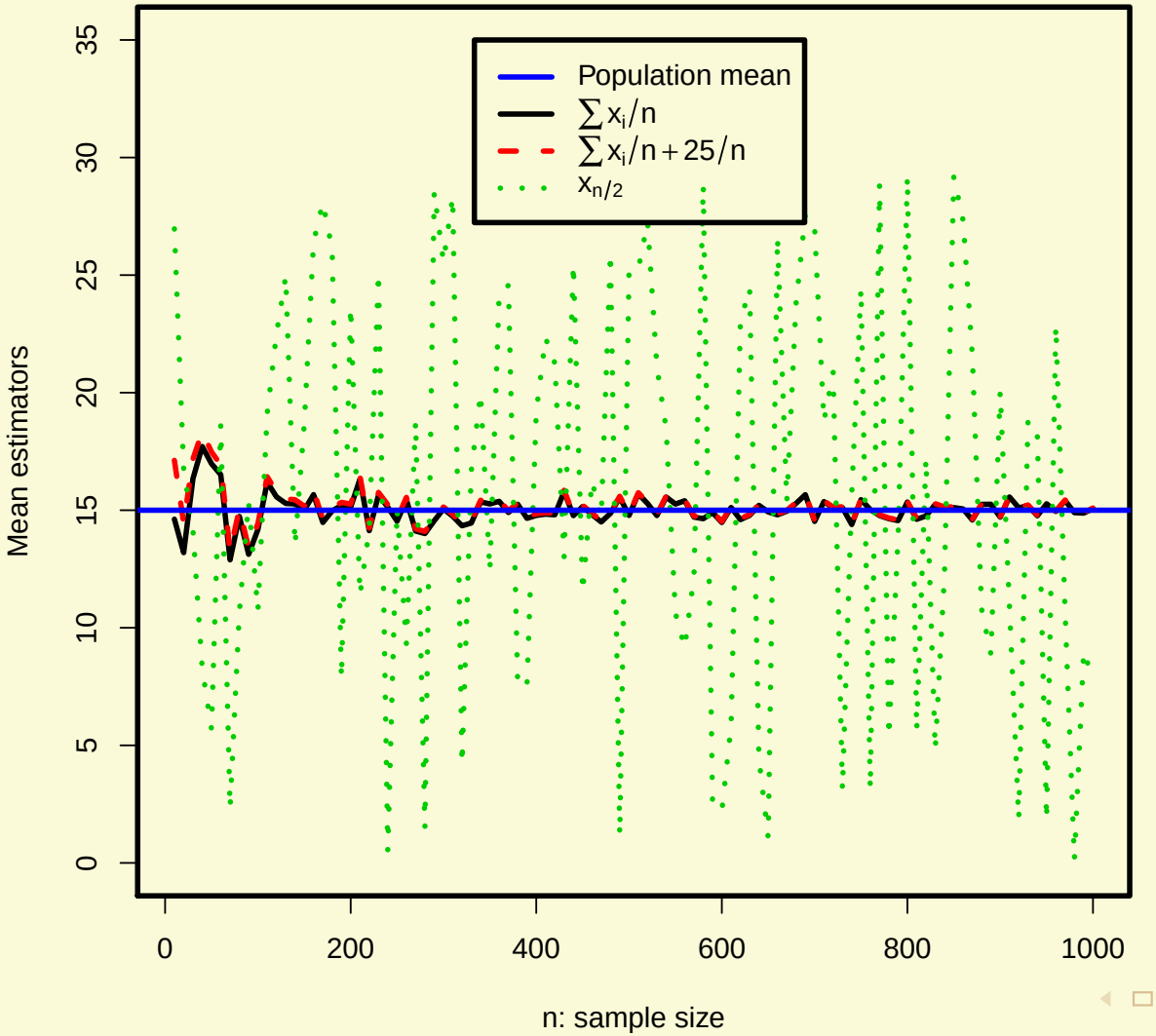
Question:

Let $\{x_1, x_2, \ldots, x_n\}$ a random sample of the variable $X \sim U(0, 30)$ with mean $\mu = 15$ and variance $\sigma^2 = 75$.

Which of the following three estimators of μ are consistent:

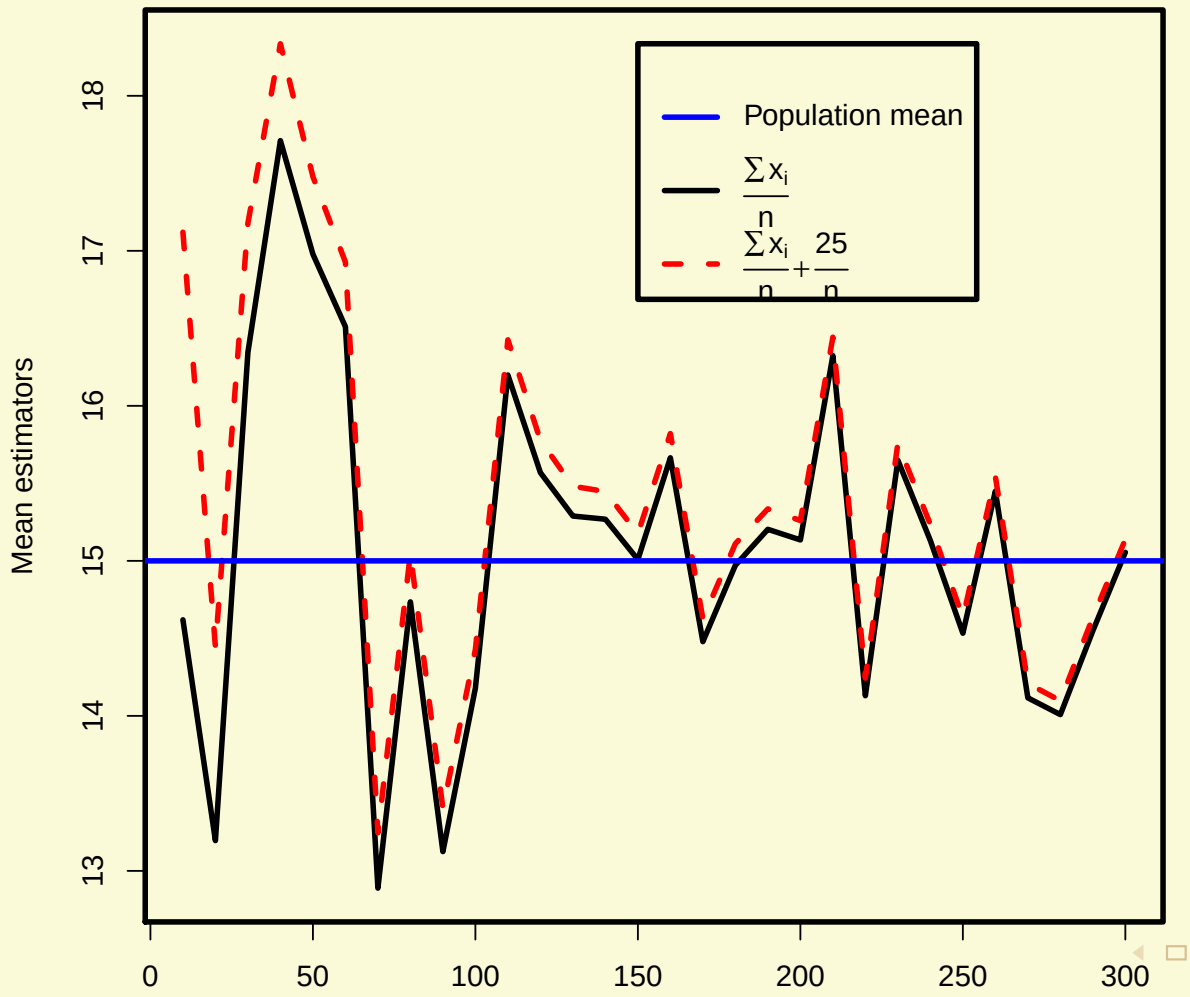
1. $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{25}{n}$
3. $\hat{\mu}_3 = x_{[n/2]}$: the middle point of the sample

Solution:


Population mean
$\sum x_i/n$
$\sum x_i/n + 25/n$
$x_{n/2}$
Mean estimators
n: sample size
0
5
10
15
20
25
30
35
0
200
400
600
800
1000

Solution:

A closer look



Exercise: $\hat{\boldsymbol{\beta}}$ in terms of $\boldsymbol{\beta}$

Mnemotechnic rule:
The OLS estimator $\hat{\boldsymbol{\beta}}^{\text{OLS}}$ is BLUE= Best Linear Unbiased Estimator of $\boldsymbol{\beta}$.

Asymptotically normal

- Commonly it is difficult to derive analytically the exact distribution of an estimator.
- A solution to this problem is to look at the distribution probability in the infinity.
- If the distribution of our parameter gets close to one of the known distributions as the sample size increases, then we can use this known distribution as an approximation of the distribution of our estimator for large samples.
- If the error term is normally distributed, then the OLS estimator is normally distributed
- If the error term is not normally distributed, then the OLS estimator is *asymptotically* normally distributed
- Under conditions OLS.1–OLS.3, the OLS estimator is asymptotically normal:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim^{a} N(0, \sigma^2 \Sigma^{-}1)$$

Asymptotically normal, intuition

$$\hat{\boldsymbol{\beta}} - \beta = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\epsilon})$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \beta) = \sqrt{n}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon} = \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\left(\frac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\epsilon}\right) \quad (10)$$

Asymptotically normal, intuition

Theorem

Central Limit Theorem (CLT)

Let $\{\mathbf{x}_t, t = 0, 1, 2, \ldots\}$ *be a sequence of independent identically distributed random variables with mean* μ *and variance* σ^2*; then the asymptotic distribution of the variable*

$$Z_T = \frac{\bar{X} - \mu}{\sigma/\sqrt{T}}$$

is $N(0, 1)$.

Asymptotically normal, intuition

- The first term of (10) is a positive definite matrix which converges to zero as $n \to \infty$
- The second term is a sum of random variables, the CLT says that if they are iid then it converges to a normal variable
- The mean of (10) is zero because the $\mathbf{X}$ and $\boldsymbol{\epsilon}$ are uncorrelated and the mean of the epsilon is zero. Therefore the asymptotic normal variable has mean zero.
- What is the mean of $\hat{\boldsymbol{\beta}}$?
- As the first term is a non-stochastic matrix, then the variance is driven by the second term of the equation.
-

$$
\begin{aligned}
Var(\tfrac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\epsilon}) = & \quad \tfrac{1}{n}E(\mathbf{X}'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{X}) - \tfrac{1}{n}E^2(\mathbf{X}'\boldsymbol{\epsilon}) \\
= & \quad \tfrac{1}{n}E(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}) \\
= & \quad \tfrac{1}{n}E(\mathbf{X}'\textstyle\sum_j(\epsilon_j^2)\mathbf{X}) \text{Because the homokedasticity} \\
= & \quad \tfrac{1}{n}E(\mathbf{X}'\mathbf{X})\textstyle\sum_j(E(\epsilon_j^2)) = \sigma^2 E(\mathbf{X}'\mathbf{X})
\end{aligned}
$$

Asymptotically normal

Therefore the asymptotic distribution of the OLS estimator is

$$\hat{\boldsymbol{\beta}}^{\text{OLS}} \sim N\left(\boldsymbol{\beta}, \frac{\sigma^2}{n}\boldsymbol{\Sigma}^{-1}\right)$$

- Now that we know the asymptotic distribution of our estimator, we can do inference and test economical hypothesis.
- Find confident intervals

Exercise

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$ be a random sample of size T. The sample moment of order h is written as

$$m_h = \frac{1}{T} \sum_{i=1}^{T} \mathbf{x}_i^h$$

and it corresponding population moment, which we suppose finite, is

$$\mu_h = E(\mathbf{x}_i^h) \text{for all} i = 1, \ldots, T$$

a) Show that m_h is a consistent estimator of μ_h.

b) Obtain the asymptotic distribution of m_h.

Solution

a) m_h is a consistent estimator of μ_h if

$$\lim_{T\to\infty} P\{|m_h - \mu_h| > \epsilon\} = 0 \quad \forall \epsilon > 0.$$

A sufficient condition (but not necessary) to obtain this result is that the bias and the variance of m_h go to zero as T grows, i.e.,

$$(E(m_h) - \mu_h) \to_{T\to\infty} 0 \quad \text{and} \quad Var(m_h) \to_{T\to\infty} 0$$

Solution

In our example

$$E(m_h) = \frac{1}{T}\sum_{i=1}^{T} E(\mathbf{x}_i^h) = \mu_h \text{ (unbiased)}$$

$$Var(m_h) = E\left(m_h - E(m_h)\right)^2 = E\left(m_h - \mu_h\right)^2 = E(m_h^2) - \mu_h^2 = \frac{\mu_{2h} - }{T}$$

because

$$\begin{aligned} E(m_h^2) = & \frac{1}{T^2}\left(\sum_{i=1}^{T} E(\mathbf{x}_i^{2h}) + \sum\sum_{i\neq j} E(\mathbf{x}_i^h)E(\mathbf{x}_j^h)\right) \\ = & \frac{1}{T^2}\left(\sum_{i=1}^{T} E(\mathbf{x}_i^{2h}) + \sum_{i,i\neq j} E(\mathbf{x}_i^h)\sum_j E(\mathbf{x}_j^h)\right) \\ = & \frac{1}{T}\mu_{2h} + \frac{T(T-1)\mu_h^2}{T^2} \end{aligned}$$

Solution

Therefore, $Var(m_h) \rightarrow_{T \rightarrow \infty} 0$.
b) The Central Limit Theorem (CLT) says that:
Let $\{\mathbf{x}_t, t = 0, 1, 2, \ldots\}$ be a sequence of independent identically distributed random variables with mean μ and variance σ^2; then the asymptotic distribution of the variable

$$Z_T = \frac{\bar{X} - \mu}{\sigma/\sqrt{T}}$$

is $N(0, 1)$.
In out case, $\mathbf{x}_1^h, \mathbf{x}_2^h, \ldots, \mathbf{x}_T^h$ are independent identically distributed variables with mean μ_h and by the CLT, its mean $m_h = \bar{X}^h$ follows a normal distribution when $T \rightarrow \infty$.

Solution

Therefore

$$Z_T = \frac{m_h - \mu_h)}{\sqrt{(\mu_{2h} - \mu_h^2)/T}} \sim N(0,1)\ T \to \infty.$$