Econometrics II

Linear Regression with Time Series Data

Morten Nyboe Tabor

Learning Goals

- Give an interpretation of the linear regression model for stationary time series and explain the model's identifying assumptions.
- Derive the method of moments (MM) estimator and state the assumptions used to derive the estimator. Estimate and interpret the parameters.
- Explain the sufficient conditions for consistency, unbiasedness, and asymptotic normality of the method of moments estimator in the linear regression model.
- Construct misspecification tests and analyze to what extent a statistical model is congruent with the data.
- Explain the consequences of autocorrelated, heteroskedastic, and non-normal residuals for the properties of the MM estimator.

Outline

1. The Linear Regression Model

Time Series Regression Models

$$\underset{K\times 1}{X_t} = \begin{pmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{Kt} \end{pmatrix}, \quad \underset{K\times 1}{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}$$

- Consider the linear regression model

$$\boxed{y_t = x_t'\beta + \epsilon_t} = x_{1t}\beta_1 + x_{2t}\beta_2 + ... + x_{kt}\beta_k + \epsilon_t,$$

for $t = 1, 2, ..., T$. Note that y_t and ϵ_t are 1×1, while x_t and β are $k \times 1$.

- The interpretation of the model depends on the variables in x_t.

TOPIC 1

1. If x_t contains contemporaneously dated variables, it is denoted a static regression:

$$y_t = x_t'\beta + \epsilon_t.$$

X_t ENTERS AT TIME t.

TOPIC 2

2. A simple model for y_t given the past is an autoregressive model:

$$y_t = \theta y_{t-1} + \epsilon_t.$$

UNIVARIATE

$X_t = y_{t-1}$

TOPIC 3

3. More complicated dynamics in the autoregressive distributed lag (ADL) model:

$$y_t = \theta_1 y_{t-1} + \tilde{x}_t'\varphi_0 + \tilde{x}_{t-1}'\varphi_1 + \epsilon_t.$$

$$X_t = \begin{pmatrix} y_{t-1} \\ \tilde{x}_t \\ \tilde{x}_{t-1} \end{pmatrix}$$

2. How do we identify and interpret parameters of the model?

Socrative Question 1

Consider the linear regression model for time series,

$$y_t = x_t'\beta + \epsilon_t, \quad t = 1, 2, ..., T. \qquad (*)$$

Q: Which one of the following assumptions identifies the parameters β in $(*)$?

- A $E(\epsilon_t) = 0$. Zero unconditional mean of ε_t.
- 65% B $E(\epsilon_t | x_t) = 0$. Predeterminedness.
- C $E(\epsilon_t | x_{1t}, x_{2t}, ..., x_t, ..., x_T) = 0$. Strict exogeneity.
- D $E(x_t \epsilon_t) = 0$. Moment condition
- E Don't know.

PREDETERMINEDNESS

$$\boxed{E(\varepsilon_t | x_t) = 0}$$

(1) No feedback $x_t \rightarrow y_t$, $y_t \not\rightarrow x_t$

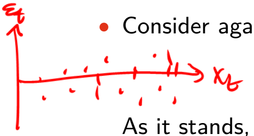
(2) Omitted variables are uncorr. with x_t.

All relevant info about $x_t \rightarrow y_t$ is included in the model.

Please go to www.socrative.com, click Student login, and enter room id Econometrics2.

Interpretation of Regression Models

$x_t \cdot 2 = E(y_t|x_t)$



- Consider again the regression model

$$y_t = x_t'\beta + \epsilon_t, \quad t = 1, 2, ..., T. \qquad (*)$$

As it stands, the equation is a tautology: not informative on β!
Why?... For any β, we can find a residual ϵ_t so that $(*)$ holds.

- We have to impose restrictions on ϵ_t to ensure a unique solution to $(*)$.
This is called identification in econometrics.
Assume that $(*)$ represents the conditional expectation, $\boxed{E(y_t \mid x_t) = x_t'\beta,}$
so that

$$\boxed{E(\epsilon_t \mid x_t) = 0.} \qquad (**)$$

X_t CONTAINS NO INFO ABOUT THE VALUE OF ϵ_t

This condition states a zero-conditional-mean. We say that x_t is predetermined.

- Under assumption $(**)$ the coefficients are the partial (ceteris paribus) effects

$$\frac{\partial E(y_t \mid x_t)}{\partial x_{jt}} = \beta_j.$$

Socrative Question 2

Consider the linear regression model for time series,

$$y_t = x_t'\beta + \epsilon_t, \quad t = 1, 2, ..., T, \qquad (*)$$

where we assume the regressors are predetermined,

$$E(\epsilon_t | x_t) = 0.$$

Q: Which of the following assumptions is *not* implied by predeterminedness?

17% A. $E(\epsilon_t) = 0$.

61% B. $E(\epsilon_t | x_1, x_2, ..., x_t, ..., x_T) = 0$.

0% C. $E(x_t \epsilon_t) = 0$.

17% D. $E(x_t^2 \epsilon_t) = 0$.

E. Don't know.

Strict exo. $\Rightarrow$ Predetermin.

UE

TWO IMPLICATIONS

① $E(\varepsilon_t | x_t) = 0 \quad (*) \quad \Longrightarrow \quad E(f(x_t)\varepsilon_t) = 0$ for any $f(x_t)$

PROOF:

$$E(f(x_t)\varepsilon_t) \overset{\text{LIE}}{=} E\big(E(f(x_t)\varepsilon_t \,|\, x_t)\big)$$

$$= E\Big(f(x_t)\underbrace{E(\varepsilon_t | x_t)}_{=0 \text{ BY } (*)}\Big) = E(f(x_t)\cdot 0)$$

$$= E(0) = \underline{\underline{0}}$$

IN PARTICULAR:

$f(x_t) = x_t$

MOMENT COND

$$(*) \Longrightarrow \boxed{E(x_t\varepsilon_t) = 0}$$

② $E(\varepsilon_t | x_t) = 0 \quad \Longrightarrow \quad E(\varepsilon_t) = 0$

PROOF: $E(\varepsilon_t) = E\big(\underbrace{E(\varepsilon_t | x_t)}_{=0 \text{ BY } (*)}\big) = E(0) = \underline{\underline{0}}$

Identification

- Predeterminedness implies the so-called moment condition:

$$\boxed{E(x_t \epsilon_t) = 0,} \qquad (***)$$

  stating that x_t and ϵ_t are uncorrelated.

- Now insert the model definition, $\epsilon_t = y_t - x_t'\beta$, in $(***)$ to obtain

$$\begin{aligned} E(x_t(y_t - x_t'\beta)) &= 0 \\ E(x_t y_t) - E(x_t x_t')\beta &= 0. \end{aligned}$$

  This is a system of k equations in the k unknown parameters, β, and if $E(x_t x_t')$ is non-singular we can find the so-called population estimator

$$\beta = E(x_t x_t')^{-1} E(x_t y_t),$$

  k×1

  which is unique.

- The parameters in β are identified by $(***)$ and the non-singularity condition.
  The latter is the well-known condition for no perfect multicollinearity.

Method of Moments (MM) Estimation

- From a given sample $(y_t, x_t')'$, $t = 1, 2, ..., T$, we cannot compute expectations. In practice we replace with sample averages and obtain the MM estimator

$$\widehat{\beta} = \left(T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} x_t y_t \right).$$

Note that the MM estimator coincides with the OLS estimator.

- For MM to work, i.e., $\widehat{\beta} \to \beta$, we need a law of large numbers (LLN) to apply, i.e.,

SAMPLE POPULATION

$$T^{-1} \sum_{t=1}^{T} x_t y_t \to E(x_t y_t) \quad \text{and} \quad T^{-1} \sum_{t=1}^{T} x_t x_t' \to E(x_t x_t').$$

- Note the two distinct conditions for OLS to converge to the true value:
  1. The moment condition $(***)$ should be satisfied.
  2. A law of large numbers should apply.

A central part of econometric analysis is to ensure these conditions.

Socrative Question 3

LLN TO CALCULATE $\hat{\beta} = (\sum x_t x_t')^{-1} (\sum x_t y_t)$

STATIONARITY $\Rightarrow$ LLN.

LLN ENSURES CONSISTENCY

$\hat{\beta} \longrightarrow \beta$ FOR $T \to \infty$

Consider the linear regression model for time series,

$$y_t = x_t'\beta + \epsilon_t, \quad \text{for } t = 1, 2, ..., T.$$

Q: Is the following statement true or false?

We can only calculate the method of moments (MM) estimator, $\widehat{\beta}$, if the time series y_t and x_t are stationary (and weakly dependent).

66% A True.

30% B False.

C .

D .

E Don't know.

CALCULATING $\hat{\beta}$ FROM DATA (y_t, x_t) $\neq$ NICE PROPERTIES

IF NO MULTICOLL. IN x_t WE CAN ALWAYS CALCULATE $\hat{\beta}$.

BUT FOR $\hat{\beta}$ TO HAVE AN INTERPRETATION + NICE PROPERTIES (WHICH WE USE FOR EG. TESTING) WE NEED ASS. OF THE MODEL TO BE FULFILLED

CHECK
YOUR
MODEL!

THEORY MODEL	EMPIRICAL MODEL (REAL WORLD)
• STOCHASTIC PROCESS: $y_1, y_2, \dots, y_t, \dots, y_T$	• OBSERVATIONS $y_1, \dots, y_T$
• MAKE ASSUMPTIONS $\Rightarrow$ DERIVE PROPERTIES	• CALCULATE NUMBERS
EXP. $E(y_t) = \mu$ VAR: $V(y_t) = \gamma_0$	EMPIRICAL AVERAGE $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ EMP. VARIANCE $\hat{\gamma}_0 = \frac{1}{T}\sum (y_t - \bar{y})^2$
POP. ESTIMATOR $\beta = E(x_t x_t')^{-1} E(x_t y_t)$	SAMPLE EST. $\hat{\beta} = \left(\sum x_t x_t'\right)^{-1} \sum x_t y_t$

LLN

Main Assumption

- We impose assumptions to ensure that a LLN applies to the sample averages.

Main Assumption

Consider a time series y_t and the $k \times 1$ vector time series x_t. We assume:

1. The process $z_t = (y_t, x_t')'$ has a joint stationary distribution.
2. The process z_t is weakly dependent, so that z_t and z_{t+h} becomes approximately independent for $h \to \infty$.

- Interpretation:
  Think of (1) as replacing identical distributions for IID data.
  Think of (2) as replacing independent observations for IID data.

- Under the main assumption, most of the results for linear regression on random samples carry over to the time series case.

Matrix Notation

EX. 1.2

$$\underset{K\times 1}{x_t} = \begin{pmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{Kt} \end{pmatrix}, \quad \underset{K\times 1}{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}$$

- MODEL: $y_t = x_t'\beta + \varepsilon_t$ for $t = 1, 2, \dots, T$.

- WRITE OUT THE T EQUATIONS:

$$\begin{aligned} y_1 &= x_1'\beta + \varepsilon_1 \\ y_2 &= x_2'\beta + \varepsilon_2 \\ &\vdots \\ y_t &= x_t'\beta + \varepsilon_t \\ &\vdots \\ y_T &= x_T'\beta + \varepsilon_T \end{aligned}$$

$$\underset{T\times 1}{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, \quad \underset{T\times K}{X} = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{pmatrix}, \quad \underset{T\times 1}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

COLLECT IN MATRICES:

$$\boxed{\underset{T\times 1}{Y} = \underset{T\times K}{X}\,\underset{K\times 1}{\beta} + \underset{T\times 1}{\varepsilon}}$$

- ESTIMATOR:

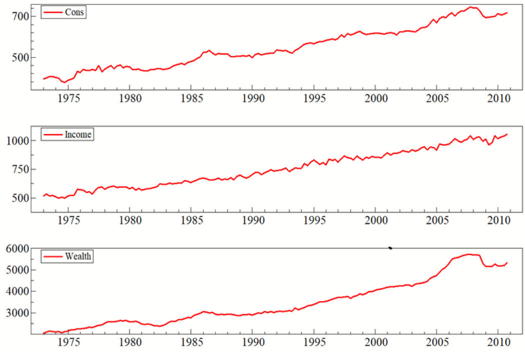
$$\boxed{\underset{K\times 1}{\hat{\beta}} = \big(\underbrace{\underset{K\times T}{X'}\,\underset{T\times K}{X}}_{K\times K}\big)^{-1}\underbrace{\underset{K\times T}{X'}\,\underset{T\times 1}{Y}}_{K\times 1}}$$

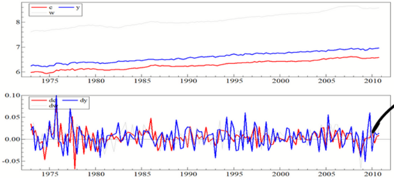
Empirical Example: Consumption, Income, and Wealth

Say that we want to estimate the effect of households' income and wealth on consumption. We have quarterly Danish data from 1973(1) to 2010(4).

Where do we start?

1. Graphs





LOGS

STATIONARITY?
NO

1. DETREND.
2. FIRST-DIFF.
3. COINTEGRATION.

$\Delta c_t = \log(c_t) - \log(c_{t-1})$
Growth rates.
Stationarity ✓

3. Properties of the Estimator

Monte Carlo Illustration: Static Regression

1. We simulate M replications from the data generating process,

$$\boxed{y_t = b_1 z_{1t} + e_t, \quad t = 1, 2, ..., T,} \tag{1}$$

where $z_{1t} \sim N(0,1)$, $e_t \sim N(0,1)$, and $\underline{b_1 = 1}$.

2. For each replication of (y_t, z_{1t}), we estimate the static linear regression model by OLS,

$$\boxed{y_t = \beta_0 + \beta_1 z_{1t} + \epsilon_t, \quad t = 1, 2, ..., T,} \tag{2}$$

which gives the estimates $\widehat{\beta}_T^{(m)}$ for $m = 1, 2, ..., M$ and given the sample size T.

We plot,

- The distributions of $\widehat{\beta}_T^{(m)}$.
- $M^{-1} \sum_{m=1}^{M} \widehat{\beta}_T^{(m)}$ plus-minus two Monte Carlo standard deviations (MCSD).

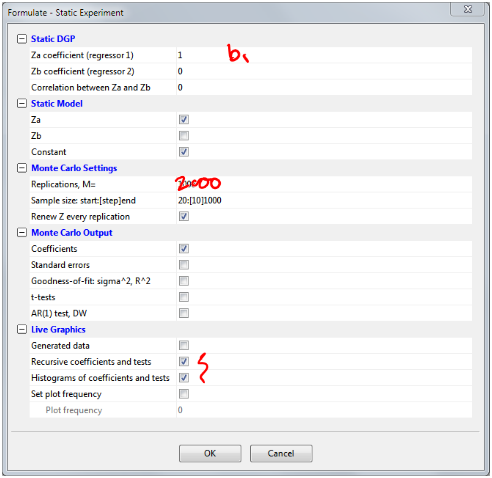
for an increasing sample size $T \in \{20, 30, ..., 1000\}$.

What do you see?

Simulation Settings in OxMetrics

We run the simulations in `OxMetrics`. Choose `Model`, select the `PcGive` module with the category `Monte Carlo` and model class `Static Experiment` using `PcNaive`.

We use the settings:



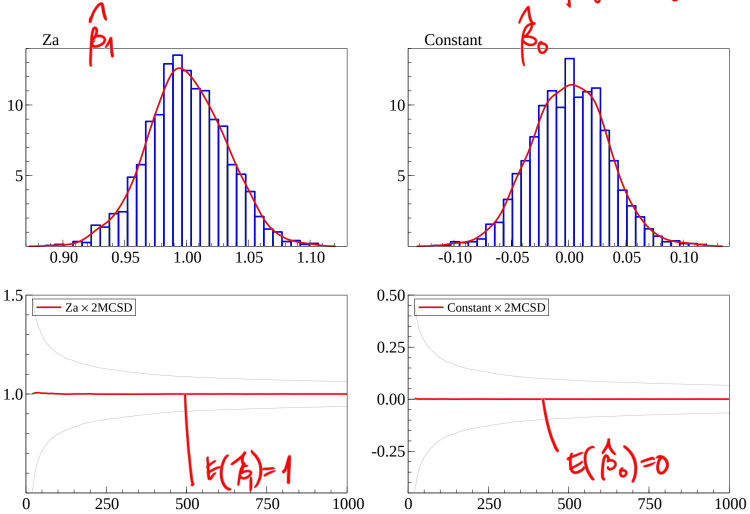
DGP

ESTIMATED MODEL

What Do You See?

1. CONSISTENCY: $\hat{\beta}_1 \to b_1 = 1$

   $\hat{\beta}_0 \to b_0 = 0$



2. UNBIASED: $E(\hat{\beta}) = \beta$

3. ASYMPTOTIC NORMALITY
   $\sqrt{T}(\hat{\beta} - \beta) \longrightarrow N(0, V)$

The Linear Regression Model and the MM Estimator

LINEAR REGRESSION MODEL:

$$y_t = x_t'\beta + \varepsilon_t$$

- MAIN ASSUMPTION: (y_t, x_t) STATIONARY + WEAK DEP.

- PREDETERMINEDNESS $\Longrightarrow$ MOMENT COND.

$$E(\varepsilon_t | x_t) = 0 \quad \Longrightarrow \quad E(x_t \varepsilon_t) = 0$$

- MM ESTIMATOR:

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t y_t$$

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t$$

$$= \left(\sum x_t x_t'\right)^{-1} \sum x_t \left(x_t'\beta + \varepsilon_t\right)$$

$$= \cancel{\left(\sum x_t x_t'\right)^{-1} \sum x_t x_t'} \beta + \left(\sum x_t x_t'\right)^{-1} \sum x_t \varepsilon_t$$

$$\boxed{\underset{\text{ESTIMATOR}}{\hat{\beta}} = \underset{\text{TRUE PARAMETER}}{\beta} + \underbrace{\left(\sum x_t x_t'\right)^{-1} \sum x_t \varepsilon_t}_{\text{SAMPLING ERROR}}}$$

Socrative Question 4

Consider the linear regression model,

$$y_t = x_t'\beta + \epsilon_t, \quad \text{for } t = 1, 2, ..., T,$$

where we assume that x_t is a single explanatory with $E(x_t) = 0$.

The MM estimator can be written as

$$\widehat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} = \beta + \frac{\sum_{t=1}^{T} x_t \epsilon_t}{\sum_{t=1}^{T} x_t^2}. \rightarrow 0$$

AS $T \to \infty$

Q: In addition to stationarity, what is required for consistency of the MM estimator, $\widehat{\beta} \to \beta$ for $T \to \infty$?

- A Strict exogeneity: $E(\epsilon_t | x_1, x_2, ..., x_T) = 0$.
- 19% B Predeterminedness: $E(\epsilon_t | x_t) = 0$.
- 70% C Moment conditions: $E(x_t \epsilon_t) = 0$.
- D Conditional homoskedasticity and no serial correlation: $E(\epsilon_t^2 | x_t) = \sigma^2$ and $E(\epsilon_t \epsilon_s | x_t, x_s) = 0$.
- E Don't know.

Please go to www.socrative.com, click Student login, and enter room id Econometrics2.

$$\hat{\beta} = \beta + \frac{\frac{1}{T}\sum x_t \varepsilon_t}{\frac{1}{T}\sum x_t^2} \xrightarrow{p} \beta + \frac{0}{\sigma_x^2} = \beta$$

FOR $T \to \infty$

$$(1) \quad \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t^2 = E(x_t^2) = \sigma_x^2$$

STATIONARITY

LLN

MOMENT COND.

$$(2) \quad \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t = E(x_t \varepsilon_t) = 0$$

Consistency

- Consistency is the first requirement for an estimator: $\widehat{\beta}$ should converge to β.

Result 1: Consistency

Let y_t and x_t obey the main assumption. If the regressors obey the moment condition,

$$E(x_t \epsilon_t) = 0,$$

then the OLS estimator is consistent, i.e., $\widehat{\beta} \to \beta$ as $T \to \infty$.

- The OLS estimator is consistent if the regression represents the conditional expectation of $y_t \mid x_t$. In this case $E(\epsilon_t | x_t) = 0$, and β_i has a "ceteris paribus" interpretation.

- The conditions in Result 1 are sufficient but not necessary. We will see examples of estimators that are consistent even if the conditions are not satisfied (related to unit roots and cointegration).

Illustration of Consistency

- Consider the regression model with a single explanatory variable, $k = 1$,

$$y_t = x_t\beta + \epsilon_t, \quad t = 1, 2, ..., T, \quad \text{with} \quad E(x_t\epsilon_t) = 0 \text{ and } E(x_t) = 0.$$

- Write the OLS estimator as

$$\widehat{\beta} = \frac{T^{-1}\sum_{t=1}^{T} y_t x_t}{T^{-1}\sum_{t=1}^{T} x_t^2} = \frac{T^{-1}\sum_{t=1}^{T}(x_t\beta + \epsilon_t)x_t}{T^{-1}\sum_{t=1}^{T} x_t^2} = \beta + \frac{T^{-1}\sum_{t=1}^{T} \epsilon_t x_t}{T^{-1}\sum_{t=1}^{T} x_t^2},$$

and look at the terms as $T \to \infty$:

$$\underset{T\to\infty}{\text{plim}}\ T^{-1}\sum_{t=1}^{T} x_t^2 = \sigma_x^2, \qquad 0 < \sigma_x^2 < \infty \qquad (\nabla)$$

$$\underset{T\to\infty}{\text{plim}}\ T^{-1}\sum_{t=1}^{T} \epsilon_t x_t = E(\epsilon_t x_t) = 0 \qquad (\nabla\nabla)$$

- Where LLN applies under Assumption 1; and
  (∇) holds for a stationary process (σ_x^2 is the limiting variance of x_t).
  ($\nabla\nabla$) follows from predeterminedness.

Proof of Unbiasedness

The MM estimator in the linear regression model for stationary time series can be written as

$$\widehat{\beta} = \beta + \frac{\sum_{t=1}^{T} x_t \epsilon_t}{\sum_{t=1}^{T} x_t x_t'}.$$

Assuming strict exogeneity,

$$E(\epsilon_t | x_1, x_2, ..., x_T) = 0,$$

the MM estimator is conditionally and unconditionally unbiased.

1. Use the law of iterated expectations to prove that the assumption of strict exogeneity implies that the MM estimator is conditionally unbiased:

$$E(\epsilon_t | x_1, x_2, ..., x_T) = 0 \quad \Rightarrow \quad E(\widehat{\beta} | x_1, x_2, ..., x_T) = \beta.$$

2. Use the law of iterated expectations to prove that conditional unbiasedness implies unconditional unbiasedness:

$$E(\widehat{\beta} | x_1, x_2, ..., x_T) = \beta \quad \Rightarrow \quad E(\widehat{\beta}) = \beta$$

Unbiasedness

- A stronger requirement for an estimator is unbiasedness: $E(\widehat{\beta}) = \beta$.

Result 2: Unbiasedness

Let y_t and x_t obey the main assumption. If the regressors are strictly exogenous,

$$E\left(\epsilon_t \mid x_1, x_2, ..., x_t, ..., x_T\right) = 0, \qquad (*)$$

then the OLS estimator is unbiased, i.e., $E(\widehat{\beta} \mid x_1, x_2, ..., x_T) = \beta$.

- CONDITIONAL UNBIASED:

$$E(\hat{\beta} \mid x_1, ..., x_T) = E\left(\beta + \frac{\sum x_t \varepsilon_t}{\sum x_t x_t'} \,\Big|\, x_1, ..., x_T\right)$$

$$= \beta + \frac{\sum' x_t \underbrace{E(\varepsilon_t \mid x_1, ..., x_T)}_{=0 \text{ UNDER } (*)}}{\sum' x_t x_t'} = \beta + 0 = \underline{\beta}$$

- UNCONDITIONAL UNBIASED:

$$E(\hat{\beta}) \overset{\text{LIE}}{=} E(\underbrace{E(\hat{\beta} \mid x_1, ..., x_T)}_{=\beta \text{ FROM ABOVE}}) = E(\beta) = \underline{\beta}$$

Unbiasedness

- A stronger requirement for an estimator is unbiasedness: $E(\widehat{\beta}) = \beta$.

Result 2: Unbiasedness

Let y_t and x_t obey the main assumption. If the regressors are strictly exogenous,

$$E\left(\epsilon_t \mid x_1, x_2, ..., x_t, ..., x_T\right) = 0,$$

then the OLS estimator is unbiased, i.e., $E(\widehat{\beta} \mid x_1, x_2, ..., x_T) = \beta$.

- Unbiasedness requires strict exogeneity, which is not fulfilled in a dynamic regression.
  Consider the first order autoregressive model

$\varepsilon_t \to y_t \to y_{t+1}$

$\to y_{t+2}$

$\to \cdots$

(1) $E(y_{t-1}\varepsilon_t)=0$ √ OK

AR(1)

$$y_t = \theta y_{t-1} + \epsilon_t.$$

ε

Here y_t is function of ϵ_t, so ϵ_t cannot be uncorrelated with $y_t, y_{t+1}, ..., y_T$.

Result 3: Estimation bias in dynamic models

In general, the OLS estimator is not unbiased in regressions with lagged dependent variables.

Example: Finite Sample Bias in an AR(1)

- Consider a first-order autoregressive, AR(1), model,

$$y_t = \theta y_{t-1} + \epsilon_t, \qquad t = 1, 2, ..., T, \qquad (*)$$

where $\theta = 0.9$. The time series y_t satisfy Assumption 1 of stationarity and weak dependence (we will return to the properties of y_t later).
The OLS estimator is,

$$\widehat{\theta} = \left(\sum_{t=1}^{T} y_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^{T} y_{t-1} y_t \right).$$

We are often interested in $E(\widehat{\theta})$ to check for bias for a given T. This is typically difficult to derive analytically.

- But if we could draw realizations of $\widehat{\theta}$, then we could estimate $E(\widehat{\theta})$.

Finite Sample Bias in an AR(1)

MC simulation:

1. Construct (randomly) M artificial data sets from the AR(1) model $(*)$.
2. Estimate the model $(*)$ for each data set and get the estimates, $\widehat{\theta}^{(m)}$, for $m = 1, 2, ..., M$.
3. Compute the Monte Carlo mean, the Monte Carlo standard deviation, and an estimate of the bias:

$$\begin{aligned}
\text{MEAN}(\widehat{\theta}) &= M^{-1} \sum_{m=1}^{M} \widehat{\theta}^{(m)} \\
\text{MCSD}(\widehat{\theta}) &= \sqrt{M^{-1} \sum_{m=1}^{M} \left(\widehat{\theta}^{(m)} - \text{MEAN}(\widehat{\theta}) \right)^2} \\
\text{Bias}(\widehat{\theta}) &= \text{MEAN}(\widehat{\theta}) - \theta.
\end{aligned}$$

Note that by the LLN (for independent observations, since the $\widehat{\theta}^{(m)}$'s are independent):

$$\text{MEAN}(\widehat{\theta}) = M^{-1} \sum_{m=1}^{M} \widehat{\theta}^{(m)} \rightarrow E(\widehat{\theta}) \quad \text{for} \quad M \rightarrow \infty.$$

Finite Sample Bias in an AR(1): Monte Carlo Simulations in PcNaive

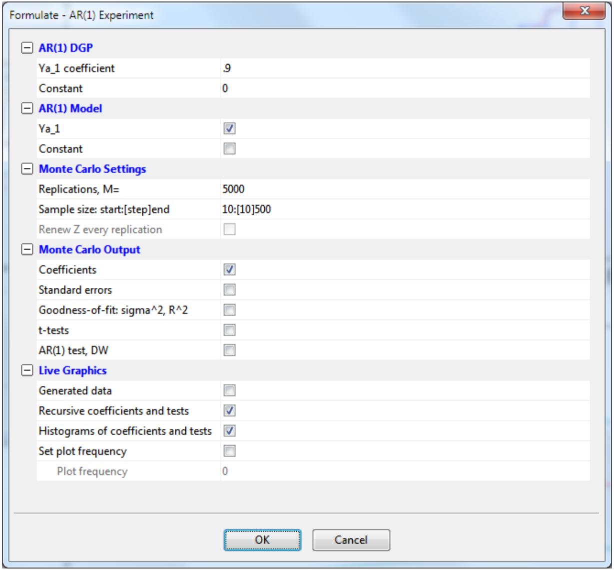
We consider Monte Carlo simulations to illustrate potential bias of the OLS estimator for θ in the AR(1) model, $(*)$. We consider the parameter values:

$$\theta \in \{0.0, 0.5, 0.9\}.$$

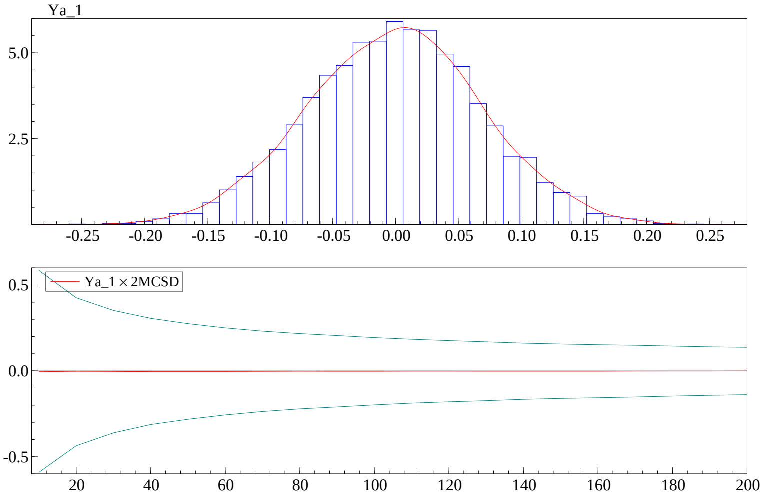
This can be done quite easily in the module PcNaive in OxMetrics.

1. In OxMetrics, click the "Model" menu and select "Model".
2. In the window that pops up, select the "PcGive" module. In "Category" select "Monte Carlo" and in "Model class" select "AR(1) Experiment using PcNaive".
3. Click "Formulate".
4. In the "Formulate" window you must specify the DGP, the estimating model, the number of replications, the sample length(s), and some output settings. See next slide.

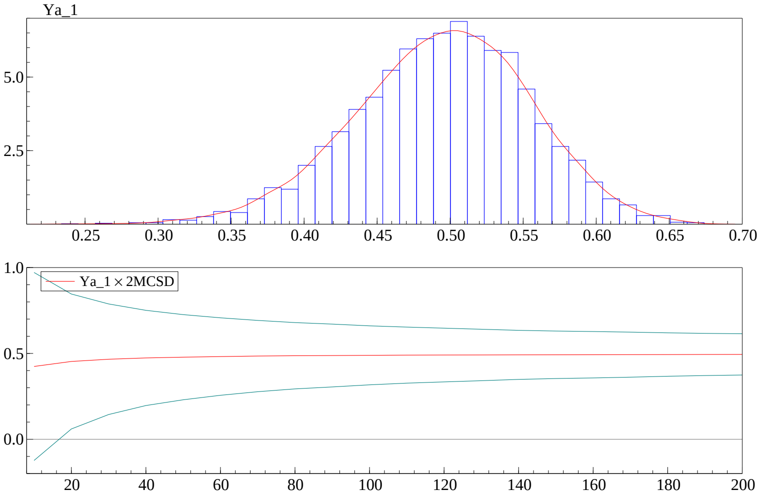
Finite Sample Bias in an AR(1): PcNaive



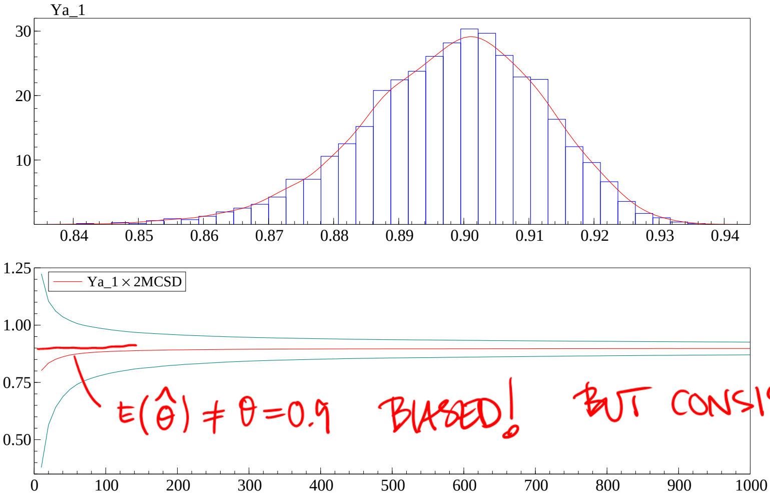
Finite Sample Bias in an AR(1): PcNaive Output for $\theta = 0$



Finite Sample Bias in an AR(1): PcNaive Output for $\theta = 0.5$



Finite Sample Bias in an AR(1): PcNaive Output for $\theta = 0.9$

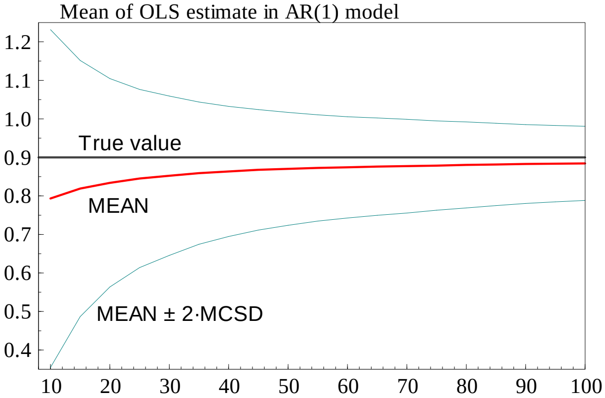


$E(\hat{\theta}) \neq \theta = 0.9$ BIASED! BUT CONSISTENT

Finite Sample Bias in an AR(1)

- In a Monte Carlo simulation, we take an AR(1) as the DGP and estimation model:

$$y_t = 0.9 \cdot y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim N(0,1).$$



Asymptotic Distribution

- To derive the asymptotic distribution we need a CLT; additional restrictions on ϵ_t.

Result 4: Asymptotic distribution and OLS variance

Let y_t and x_t obey the main assumption and let x_t and ϵ_t be uncorrelated, $E(\epsilon_t x_t) = 0$. Furthermore, assume (conditional) **homoskedasticity** and **no serial correlation**, i.e.,

$$\begin{aligned} (\#) \qquad E\left(\epsilon_t^2 \mid x_t\right) &= \sigma^2 \\ (\#\#) \qquad E\left(\epsilon_t \epsilon_s \mid x_t, x_s\right) &= 0 \quad \text{for all } t \neq s. \end{aligned}$$

Then as $T \to \infty$, the OLS estimator is **asymptotically normal** with usual variance:

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) \xrightarrow{D} N(0, \sigma^2 E(x_t x_t')^{-1}).$$

- Inserting estimators, we can test hypotheses using

$$\widehat{\beta} \overset{a}{\sim} N\left(\beta, \widehat{\sigma}^2 \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}\right).$$

- Recall that $E(\epsilon_t x_t) = 0$ is implied by $E[\epsilon_t | x_t] = 0$.

$$\sqrt{T}(\hat{\beta}-\beta) = \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1}}_{\xrightarrow{p} E(x_t x_t') = \Sigma_{xx}} \underbrace{\frac{\sqrt{T}}{T}\sum_{t=1}^{T} x_t \varepsilon_t}_{\xrightarrow{d} N(0,S)} \xrightarrow{d} N(0, \Sigma_{xx}^{-1} S \Sigma_{xx}^{-1})$$

Σ_{xx}: $k \times k$

SYMMETRIC

$$\begin{aligned} S &= E((x_t\varepsilon_t)(x_t\varepsilon_t)') \\ &= E(x_t\varepsilon_t\varepsilon_t' x_t') \\ &= E(\varepsilon_t^2 x_t x_t') \\ &= E(E(\varepsilon_t^2 x_t x_t' \mid x_t)) \\ &= E(E(\varepsilon_t^2 \mid x_t) x_t x_t') \\ &= E(\sigma^2 x_t x_t') \\ &= \sigma^2 \Sigma_{xx} \end{aligned}$$

$$\Sigma_{xx}^{-1} S \Sigma_{xx}^{-1} = \Sigma_{xx}^{-1} \sigma^2 \cancel{\Sigma_{xx}} \cancel{\Sigma_{xx}^{-1}} = \sigma^2 \Sigma_{xx}^{-1} = \sigma^2 E(x_t x_t')^{-1}$$

$$\sqrt{T}(\hat{\beta}-\beta) \xrightarrow{d} N(0, \sigma^2 E(x_t x_t')^{-1})$$

$$\hat{\beta}-\beta \overset{a}{\sim} N(0, T^{-1} \sigma^2 E(x_t x_t')^{-1})$$

$$\boxed{\hat{\beta} \overset{a}{\sim} N(\beta, T^{-1}\sigma^2 E(x_t x_t')^{-1})}$$

Empirical Example: The Linear Regression Model

Empirial Example: Estimates

4. Dynamic Completeness and Autocorrelation

A Dynamically Complete Model

$E(\varepsilon_t \varepsilon_s | X_t, X_s) = 0$
for $t = 1, 2, \dots, T$, $s = 1, 2, \dots, T$ $t \neq s$.

- Result 4 (Asymptotic distribution of the OLS estimator) required no serial correlation in the error terms.
  The precise condition for no serial correlation looks strange. Often we disregard conditioning and consider whether ϵ_t and ϵ_s are uncorrelated.

- We say that a model is dynamically complete if

  $E(y_t | I_t)$

  $$E(y_t \mid x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \dots, y_1, x_1) = E(y_t \mid x_t) = x_t'\beta.$$

  SET OF EXPL. VARIABLES

  x_t contains all relevant information in the available information set.
  No-serial-correlation is practically the same as dynamic completeness.

- All systematic information in the past of y_t and x_t is used in the regression model.
  This is often taken as an important design criteria for a dynamic regression model.
  We should always test for no-autocorrelation in time series models.

Autocorrelation of the Error Term

- If $\text{Cov}(\epsilon_t, \epsilon_s) \neq 0$ for some $t \neq s$, we have autocorrelation of the error term.

ε_t

- This is detected from the estimated residuals and $\text{Cov}(\widehat{\epsilon_t}, \widehat{\epsilon_s}) \neq 0$ is referred to as residual autocorrelation. Often used synonymously.

$\hat{\varepsilon}_t$

- Residual autocorrelation does not imply that the DGP has autocorrelated errors. Typically, autocorrelation is taken as a signal of misspecification. Different possibilities:

  1. Autoregressive errors in the DGP.
  2. Dynamic misspecification.
  3. Omitted variables and non-modelled structural shifts.
  4. Misspecified functional form

The solution to the problem depends on the interpretation.

Consequences of Autocorrelation

$E(x_t \varepsilon_t) = 0$

CASE 1
- Autocorrelation will not violate the assumptions for **Result 1** (consistency) in general.

CASE 2
- But $E(x_t \epsilon_t) = 0$ is violated if the model includes a lagged dependent variable. Look at an AR(1) model with error autocorrelation, i.e., the two equations

$y_{t-1} = \theta y_{t-2} + \varepsilon_{t-1}$

$\varepsilon_{t-1} \to \varepsilon_t$, $\varepsilon_{t-1} \to y_{t-1}$

$$\begin{aligned} y_t &= \theta y_{t-1} + \epsilon_t \\ \epsilon_t &= \rho \epsilon_{t-1} + v_t, \quad v_t \sim \text{IID}(0, \sigma_v^2). \end{aligned}$$

$E(y_{t-1} \varepsilon_t) \neq 0$ $\longrightarrow$ NOT CONSISTENT EST.

Both y_{t-1} and ϵ_t depend on ϵ_{t-1}, so $E(y_{t-1}\epsilon_t) \neq 0$.

Result 5: Inconsistency of OLS

In a regression model including the lagged dependent variable, the OLS estimator is not consistent in the presence of autocorrelation of the error term.

- Even if OLS is consistent, the standard formula for the variance in **Result 4** is no longer valid. It is possible to derive the variance, the so-called heteroskedasticity-and-autocorrelation-consistent (HAC) standard errors.

- STATIC REGR. WITH AUTOCORR. ERRORS:

$$y_t = x_t' \beta + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t, \quad v_t \sim IID(0,\sigma^2)$$

$|\rho|<1$

$E(\varepsilon_t)=0$

$V(\varepsilon_t) = E(\varepsilon_t^2) = \dfrac{\sigma^2}{1-\rho^2}$

- MOMENT CONDITIONS NOT VIOLATED

$E(x_t \varepsilon_t) = 0 \implies \hat{\beta}$ IS CONSISTENT.

SERIAL CORR. IN ε_t:

- $E(\varepsilon_t \varepsilon_s | x_t, x_s) \neq 0$

$\implies$ USUAL OLS ASYMPT. VAR. NOT VALID

THIS IS WHAT SOFTWARE REPORTS!

Simulation Illustration: Inconsistency in Dynamic Model

Data-generating Process

We simulate time series from the stationary AR(2) model with autocorrelated errors,

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \epsilon_t, \qquad (*)$$

$$\epsilon_t = \rho \epsilon_{t-1} + v_t, \qquad (**)$$

for $t = 1, 2, ..., T$, where v_t is assumed IIDN(0,1) and the initial values are $y_0 = y_{-1} = 0$. We set $\theta_1 = 0.7$, $\theta_2 = 0.2$, and $\rho = 0.9$. We simulate $M = 10,000$ replications with sample size $T = 200, 1000$.

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 0.7 \\ 0.2 \end{pmatrix}$$

Estimated Model

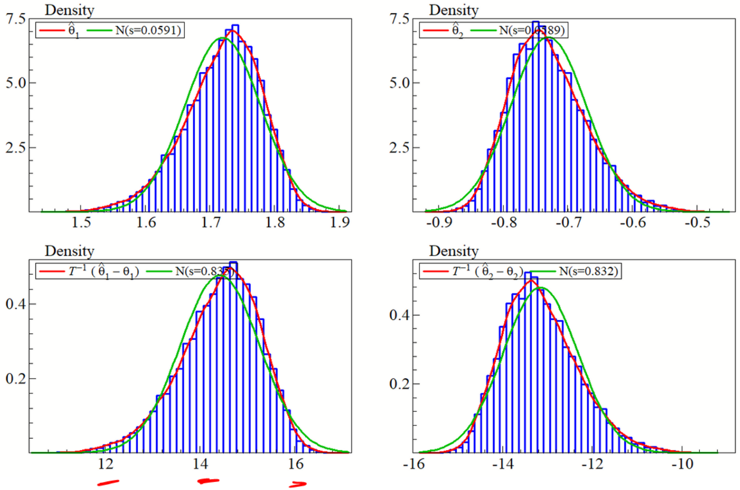
For the simulated data, we estimate the AR(2) model in $(*)$ by MM/OLS:

$$\begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{pmatrix} = \left(\sum_{t=1}^{T} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} \begin{pmatrix} y_{t-1} & y_{t-2} \end{pmatrix} \right)^{-1} \left(\sum_{t=1}^{T} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} y_t \right)$$

We plot the distributions of $\widehat{\theta}_1$, $\widehat{\theta}_2$, $\sqrt{T}(\widehat{\theta}_1 - \theta_1)$, and $\sqrt{T}(\widehat{\theta}_2 - \theta_2)$.

Consistency of the estimator requires the moment condition, which is violated due to the autocorrelation of the error-term.

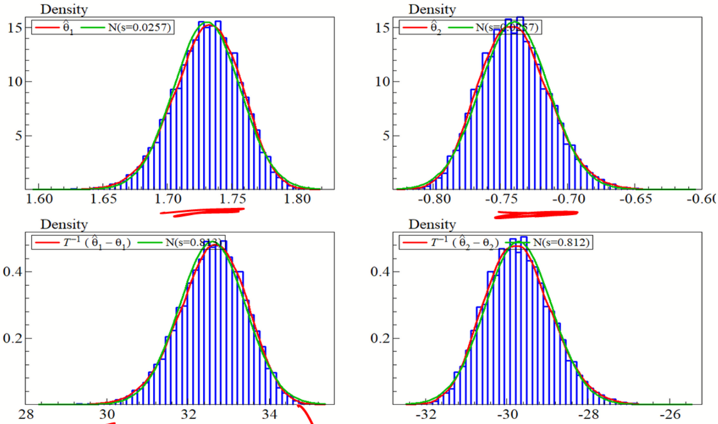
Simulation Illustration: Inconsistency in Dynamic Model



$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 0.7 \\ 0.2 \end{pmatrix}$$

$$\hat{\theta}_1 \not\to \theta_1$$

$$\hat{\theta}_2 \not\to \theta_2$$



$\sqrt{T}(\hat{\theta}_1 - \theta_1)$

NOT A STABLE DISTR.

Simulation Illustration: Inconsistency in Dynamic Model

(1) Autoregressive Errors in the DGP

- Consider the case where the errors are truly autoregressive:

$$y_t = x_t'\beta + \epsilon_t$$
$$\epsilon_t = \rho\epsilon_{t-1} + v_t, \qquad v_t \sim \text{IID}(0, \sigma_v^2).$$

- If ρ is known we can write

$$\begin{aligned} (y_t - \rho y_{t-1}) &= \left(x_t' - \rho x_{t-1}'\right)\beta + (\epsilon_t - \rho\epsilon_{t-1}) \\ y_t &= \rho y_{t-1} + x_t'\beta - x_{t-1}'\rho\beta + v_t. \end{aligned}$$

  The transformation is analog to GLS transformation in the case of heteroskedasticity.

- The GLS model is subject to a so-called common factor restriction: three regressors but only two parameters, ρ and β. Estimation is non-linear and can be carried out by maximum likelihood.

- Consistent estimation of the parameters in the GLS model requires

$$E((x_t - \rho x_{t-1})(\epsilon_t - \rho\epsilon_{t-1})) = 0.$$

But then ϵ_{t-1} should be uncorrelated with x_t, i.e., $E(\epsilon_t x_{t+1}) = 0$.
Consistency of GLS requires stronger assumptions than consistency of OLS.

- The GLS transformation is rarely used in modern econometrics.
  1. Residual autocorrelation does not imply that the error term is autoregressive.
     There is no a priori reason to believe that the transformation is correct.
  2. The requirement for consistency of GLS is strong.

(2) Dynamic Misspecification

- Residual autocorrelation indicates that the model is not dynamically complete, so

$$E(y_t \mid x_t) \neq E(y_t \mid x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, ..., y_1, x_1).$$

  The (dynamic) model is misspecified and should be reformulated.
  Natural remedy is to extend the list of lags of x_t and y_t.

- If autocorrelation seems of order one, then a starting point is the GLS transformation.
  The AR(1) structure is only indicative and we look at the unrestricted

TOPIC 2 ADL model

$$y_t = \alpha_0 y_{t-1} + x_t' \alpha_1 + x_{t-1}' \alpha_2 + \eta_t.$$

  Finding of first-order autocorrelation is only used to extend the list of regressors. The common factor restriction is removed.

(3) Omitted Variables

- Omitted variables in general can also produce autocorrelation. Let the DGP be

$$y_t = x_{1t} \cdot \beta_1 + x_{2t} \cdot \beta_2 + \epsilon_t, \qquad (\diamondsuit)$$

and consider the estimation model

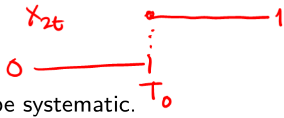
$$y_t = x_{1t} \cdot \beta_1 + u_t. \qquad (\diamondsuit\diamondsuit)$$

OMITTED X_{2t}.

Then the error term is $u_t = x_{2t} \cdot \beta_2 + \epsilon_t$, which is autocorrelated if x_{2t} is persistent.

- An example is if the DGP exhibits a level shift, e.g., $(\diamondsuit)$ includes the dummy variable

$$x_{2t} = \begin{cases} 0 & \text{for} \quad t < T_0 \\ 1 & \text{for} \quad t \geq T_0 \end{cases}.$$



If x_{2t} is not included in $(\diamondsuit\diamondsuit)$ then the residual will be systematic.

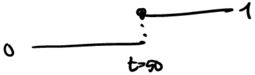
- Again the solution is to extend the list of regressors, x_t.

PRACTICE: FIND T_0. ADD X_{2t}.

1. SIMULATE FROM DGP $\quad b_1 = 1, \; b_2 = 2$

$$y_t = b_1 X_{1t} + b_2 X_{2t} + \varepsilon_t, \quad t = 1, 2, \dots, 100$$

X_{2t} LEVEL-SHIFT: $\quad X_{2t} = \begin{cases} 0 & \text{for } t < 50 \\ 1 & \text{for } t \geq 50 \end{cases}$



2. ESTIMATE:

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t$$

~ IDENTICAL TO DGP.

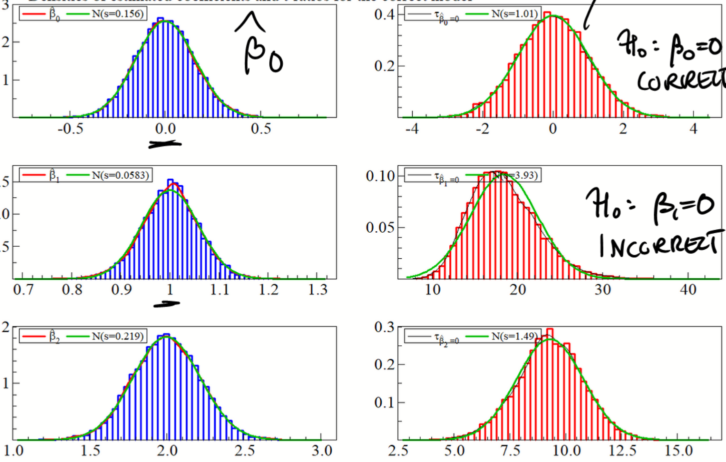
3. ESTIMATE:

$$y_t = \beta_0 + \beta_1 X_{1t} + \varepsilon_t$$

②

$$t_{\beta_0=0} = \frac{\hat{\beta}_0 - 0}{se(\hat{\beta}_0)}$$

Densities of estimated coefficients and t-ratios for the correct model



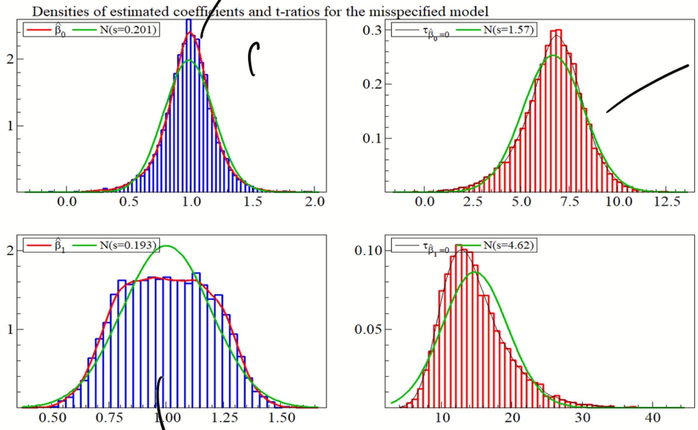
$\hat{\beta}_0$

$\mathcal{H}_0: \beta_0 = 0$ CORRECT $\sim N(0,1)$ under the null

$\mathcal{H}_0: \beta_1 = 0$ INCORRECT

$b_0 = 0$

Centred on 1.



NOT ASYMPTOTICALL NORMAL!

(4) Misspecified Functional Form

- If the true relationship

$$y_t = g(x_t) + \epsilon_t$$

  is non-linear, the residuals from a linear regression will typically be autocorrelated.

- The solution is to try to reformulate the functional form of the regression line.

5. Model Formulation and Misspecification Testing

Model Formulation and Properties of the Estimator

- We consider the linear regression model for time series,

$$y_t = x_t'\beta + \varepsilon_t. \qquad (*)$$

If there is no perfect multicolleniarity in x_t we can always compute the OLS estimator $\widehat{\beta}$ and consider the estimated model,

$$y_t = x_t'\widehat{\beta} + \widehat{\varepsilon}_t, \qquad \text{for the sample } t = 1, 2, ... T. \qquad (**)$$

- Under specific assumptions can we derive properties of the estimator $\widehat{\beta}$:
  1. Consistency: $\widehat{\beta} \to \beta$ as $T \to \infty$.
  2. Unbiasedness: $E[\widehat{\beta}|x_1, x_2, ..., x_T] = \beta$.
  3. Asymptotic distribution: $\widehat{\beta} \overset{a}{\sim} N(\beta, \widehat{\sigma}^2(\sum_{t=1}^{T} x_t x_t')^{-1})$.

- However, if $(**)$ does not satisfy the specific assumptions we cannot use the results in (1)-(3) to interpret the estimated coefficients and to do statistical inference (e.g. test the hypothesis $\mathcal{H}_0 : \beta_i = 0$).

Misspecification Testing

- We can never show that the model is well specified;
  but we can certainly find out if it is not in a specific direction!
  Misspecification testing.

- We consider the following standard tests on the estimated errors, $\widehat{\varepsilon}_t$:
  1. Test for no-autocorrelation.
  2. Test for no-heteroskedasticity.
  3. Test for normality of the error term.

- If all tests are passed, we may think of the model as representing the main features of the data and we can use the results in (1) – (3), e.g. for testing hypotheses on estimated parameters.

(1) Tests for No-Autocorrelation

- Let $\widehat{\epsilon}_t$ $(t = 1, 2, ..., T)$ be the estimated residuals from the original regression model

$$y_t = x_t'\beta + \epsilon_t. \longrightarrow \hat{\beta},\ \hat{\varepsilon}_t$$

  A test for no-autocorrelation (of first order) is based on the hypothesis $\gamma = 0$ in the auxiliary regression

$$\boxed{\widehat{\epsilon}_t = x_t'\delta + \overset{=0}{\gamma}\widehat{\epsilon}_{t-1} + u_t,}$$

  where x_t is included because it may be correlated with ϵ_{t-1}.

- A valid test is $t-$ratio for $\gamma = 0$. Alternatively there is the Breusch-Godfrey LM test

  \# OF OBS.

  FROM AUX. REGR.

$$\text{LM} = T \cdot R^2 \sim \chi^2(1).$$

  Note that x_t and ϵ_t are orthogonal, and any explanatory power is due to $\widehat{\epsilon}_{t-1}$.

- The Durbin Watson (DW) test is derived for finite samples. Based on strict exogeneity. Not valid in many models.

Statistical Testing

1. Null hypothesis:

   NO AUTOCORR. $H_0: \gamma = 0$

   SIGNIFICANCE OF β_1 $H_0: \beta_1 = 0$

2. Test statistics:

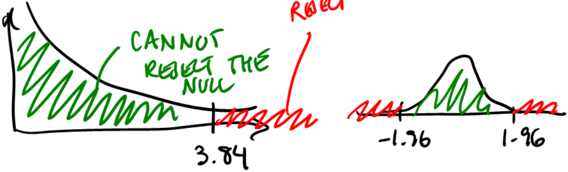
   $LM = T \cdot R^2$

   $t_{\beta_1 = 0} = \dfrac{\hat{\beta}_1 - 0}{se(\hat{\beta}_1)}$

3. Asymptotic distribution of the test statistics under the null:

   $LM \to \chi^2(1)$

   $t_{\beta_1 = 0} \to N(0,1)$



(2) Test for No-Heteroskedasticity

- Consider the auxiliary regression

$$\widehat{\epsilon_t^2} = \gamma_0 + x_{1t}\gamma_1 + ... + x_{kt}\gamma_k + x_{1t}^2\delta_1 + ... + x_{kt}^2\delta_k + u_t.$$

- A test for unconditional homoskedasticity is based on the null hypothesis:

$$\gamma_1 = ... = \gamma_k = \delta_1 = ... = \delta_k = 0.$$

  The alternative is that the variance of ϵ_t depends on x_{it} or the squares x_{it}^2 for some $i = 1, 2, ...k$.

- The LM test,

$$\mathrm{LM} = T \cdot R^2 \sim \chi^2(2k).$$

(3) Test for Normality of the Error Term

- Skewness (S) and kurtosis (K) are the estimated third and fourth central moments of the standardized estimated residuals $u_t = (\widehat{\varepsilon}_t - \bar{\varepsilon})/\widehat{\sigma}$.:

$$S = T^{-1} \sum_{t=1}^{T} u_t^3 \quad \text{and} \quad K = T^{-1} \sum_{t=1}^{T} u_t^4.$$

- The normal distribution is symmetric and has $S = 0$, while $K = 3$ ($K - 3$ is often referred to as excess kurtosis). If $K > 3$, the distribution has "fat tails".
- Under the assumption of normality,

$$\begin{aligned} \xi_S &= \frac{T}{6} S^2 && \to \chi^2(1) \\ \xi_K &= \frac{T}{24} (K-3)^2 && \to \chi^2(1). \end{aligned}$$

- It turns out that ξ_S and ξ_K are independent: Jarque-Bera joint test of normality:

$$\xi_{JB} = \xi_S + \xi_K \to \chi^2(2).$$

Notes on Normality of the Error Term

- Often normality of the error terms is rejected due to the presence of a few large outliers that the model cannot account for (i.e. they are captured by the error term).
- Start with a plot and a histogram of the (standardized) estimated residuals. Any large outliers of, say, more than three standard deviations?
- A big residual at time T_0 may be accounted for by including a dummy variable ($x_t = 1$ for $t = T_0$, $x_t = 0$ for all other t).
- Note that the results for the linear regression model hold without assuming normality of ε_t. However, the normal distribution is a natural benchmark and given normality:
  - $\widehat{\beta}$ converges faster to the asymptotic normal distribution.
  - The OLS estimator coincides with the maximum likelihood (ML) estimator.

Model Formulation and Misspecification Testing in Practice

- Finding a well-specified model for the data of interest can be hard in practice! Potential reasons include:
  - Which variables to include in x_t? Economic theory can often guide us.
  - How many lags to include to obtain a dynamically complete model?
  - Are the model parameters constant over the sample or do we need to include structural breaks?
  - Do we need to include dummy variables to capture large outliers?

  Think about which model and which formulation is useful to represent the characteristics of the data!

- In practice we use an iterative procedure:
  1. Estimate the model for a specific formulation.
  2. Do misspecification tests on the estimated errors.
  3. If the misspecification tests are rejected, re-formulate the model. Repeat (1) and (2) until the misspecification tests are not rejected.

- General-to-Specific: Start with a larger model and simplify it by removing insignificant variables.

Socrative Question 5

For the estimated model for consumption, we get the following output for the test for no autocorrelation of order 1-2.

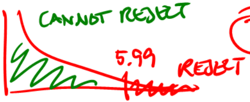
```
Error autocorrelation coefficients in auxiliary regression:
       Coefficient  Std.Error    t-value
Lag 1     -0.18842     0.1305     -1.444
Lag 2     0.092372      0.103     0.8972
RSS = 0.0304648  sigma = 0.000210102

Testing for error autocorrelation from lags 1 to 2
 Chi^2(2)   =    5.7742 [0.0557]   and F-form F(2,145)   =   2.8826 [0.0592]
```

OXMETRICS REPORTS F-FORM BETTER SMALL SAMPLE PROP.

DISTR. TEST STAT P-VALUE

Q: What is your conclusion to the LM test for no autocorrelation of order 1-2?

- A We cannot reject the null of autocorrelation.
- 70% B We cannot reject the null of no autocorrelation.
- C We reject the null of autocorrelation.
- D We reject the null of no autocorrelation.
- E Don't know.



(1) NULL: NO autocorr of order 1-2. H_0

(2) TEST STAT: $LM = T \cdot R^2 = 5.77$

(3) ASYMPT. DISTR. $LM \to \chi^2(2)$

CV. 5% = 5.99

Please go to www.socrative.com, click Student login, and enter room id Econometrics2.

CONSEQUENCES OF AUTOCORRELATION

① AUTOCORR. IN STATIC MODEL → $E(x_t \varepsilon_t) = 0$ ✓ ; $E(\varepsilon_t \varepsilon_s | x_t, x_s) \neq 0$ →

- CONSISTENCY AS $\frac{1}{T} \sum x_t \varepsilon_t \to E(x_t \varepsilon_t) = 0$
- ASYMPTOTIC NORMALITY $\sqrt{T}(\hat{\beta} - \beta) \to N(0, V)$ BUT $V \neq \sigma^2 E(x_t x_t')^{-1}$!

② AUTO CORR. IN DYNAMIC MODEL → $E(x_t \varepsilon_t) \neq 0$; $E(\varepsilon_t \varepsilon_s | x_t, x_s) \neq 0$ →

- INCONSISTENTLY ! $\hat{\beta} \not\to \beta$
- $\sqrt{T}(\hat{\beta} - \beta)$ NOT ASYMPT. NORMAL !

WHAT TO DO?

- ADD LAGS ! TOPIC 2.
- LOOK FOR LARGE LEVEL SHIFTS. INCLUDE SHIFT-DUMMIES.
- HAC STD ERRORS (ROBUST) (+ COMPARE!)

Socrative Question 6

Let the Jarque-Berra test for normality of the estimated residuals give the following test statistics,

$$\xi_{JB} = \xi_S + \xi_K = 4.9.$$

Q: What is your conclusion to the test based on a 5 percent significance level?

- A We cannot reject the null of normality of the estimated residuals.
- B We reject the null of normality of the estimated residuals.
- C .
- D .
- E Don't know.

Socrative Question 7

Q: Which of the following statements is correct if the error term is heteroskedastic?

6% (A) $\widehat{\beta}$ is inconsistent.

8% (B) $\widehat{\beta}$ is inconsistent, but $\sqrt{T}(\widehat{\beta}-\beta)$ is asymptotically normal.

76% (C) $\widehat{\beta}$ is consistent, but $\sqrt{T}(\widehat{\beta}-\beta)$ is not asymptotically normal.

6% (D) $\widehat{\beta}$ is consistent and $\sqrt{T}(\widehat{\beta}-\beta)$ is asymptotically normal.

3% (E) Don't know.

BUT $V \neq \sigma^2 E(x_t x_t')^{-1}$

MOMENT COND:

$E(x_t \varepsilon_t) = 0$ ✓

HETEROSKEDASTICITY

$\longrightarrow$ $E(\varepsilon_t^2) \neq \sigma^2$ $\longrightarrow$ ASYMPTOTIC NORMALITY

$\left(E(\varepsilon_t^2 | x_t) \neq \sigma^2\right)$

$\sqrt{T}(\hat{\beta} - \beta) \to N(0, V)$

BUT $V \neq \sigma^2 E(x_t x_t')^{-1}$

WHAT TO DO?

1. LOOK FOR OUTLIERS! $\to$ DUMMIES.
2. MODEL TIME-VARYING HETEROSKEDASTICITY
   TOPIC 4: ARCH.
3. HAC STD. ERRORS.

Socrative Question 8

$\varepsilon_t \sim N(0,\sigma^2) \Longrightarrow \hat{\beta}_{MM} = \hat{\beta}_{OLS} = \hat{\beta}_{ML}$

+ independent

MAXIMUM LIKELIHOOD EST

Q: Why do we care about normality of the error term?

57% **A** If $\epsilon_t \sim Niid(0, \sigma^2)$, then the MM estimator converges faster to the asymptotic normal distribution and it is efficient.

6% **B** The MM estimator is inconsistent if ϵ_t is not normally distributed and independent over time.

24% **C** The t-test of the null that $\beta_i = 0$, $t_{\beta_i=0} = \widehat{\beta}_i / s.e.(\widehat{\beta}_i)$ is only asymptotically $N(0,1)$ if $\epsilon_t \sim Niid(0, \sigma^2)$.

$t_{\beta_i=0} \overset{a}{\sim} N(0,1)$

consequence of consistency + asympt. normality

4% **D** The MM estimator is not asymptotically normally distributed unless ϵ_t is normally distributed.

10% **E** Don't know.

NORMALITY IS NICE, BUT NOT REQUIRED!

Empirical Example: Consumption, Income, and Wealth

We consider the model for consumption

$$\Delta c_t = \delta + \gamma_1 \Delta y_t + \gamma_2 \Delta w_t + \alpha \mathrm{ECM}_{t-1} + \epsilon_t,$$

for $t = 1, 2, ..., T$, where ϵ_t is assumed IID(0,σ^2) and ECM_{t-1} is defined as

$$\mathrm{ECM}_{t-1} = c_{t-1} - 0.88 - 0.44 y_{t-1} - 0.31 w_{t-1}.$$

We estimate the model by MM/OLS using quarterly data for 1973(1)–2010(4).

Empirical Example: Consumption, Income, and Wealth

Empirical Example: Consumption, Income, and Wealth

Based on the graphs of the estimated residuals, discuss which of the following assumptions are satisfied for the estimated model.
Explain how you reach your conclusion.

(1) Stationarity of (y_t, x_t).

(2) Predeterminedness: $E(\epsilon_t | x_t) = 0$.
$\rightarrow$ Zero-mean: $E(\epsilon_t) = 0$.
$\rightarrow$ Moment-condition: $E(x_t \epsilon_t) = 0$.

(3) Homoskedasticity: $E(\epsilon_t^2 | x_t) = \sigma^2 \quad \rightarrow E(\epsilon_t^2) = \sigma^2$.

(4) No serial correlation: $E(\epsilon_t \epsilon_s | x_t, x_s) = 0 \quad \rightarrow E(\epsilon_t \epsilon_{t-k}) = 0$.

(5) Normally distributed errors: $\epsilon_t \sim N(0, \sigma^2)$.

Do you notice anything else we should think about?

Socrative Question 9

For the estimated model for consumption, we get the misspecification tests:

```
AR 1-5 test:      F(5,142)   =   1.9539 [0.0891]
Normality test:   Chi^2(2)   =   36.929 [0.0000]**
Hetero test:      F(6,144)   =   2.3407 [0.0346]*
```

Q. What do you conclude regarding the misspecification tests?

- A The estimated residuals are not autocorrelated, homoskedastic, and normal.
- B The estimated residuals are homoskedastic and normal, but autocorrelated.
- 90% C The estimated residuals are not autocorrelated, but heteroskedastic and non-normal.
- D The estimated residuals are autocorrelated, heteroskedastic, and non-normal.
- E Don't know.

Please go to www.socrative.com, click Student login, and enter room id Econometrics2.

6. The Frisch-Waugh-Lovell Theorem

Ex. #1.3

Motivation

Consider the linear regression model,

$$y_t = \mu + \beta_1 x_t + \varepsilon_t, \quad t = 1, 2, ..., T, \tag{3}$$

and the linear regression model for the de-meaned variables $\widetilde{y}_t = y_t - \bar{y}$ and $\widetilde{x}_t = x_t - \bar{x}$, given by

$$\widetilde{y}_t = b_1 \widetilde{x}_t + u_t, \quad t = 1, 2, ..., T. \tag{4}$$

Question: Would you expect $\hat{\beta}_1$ to equal $\hat{b}_1$?

Motivation: The Frisch-Waugh-Lovell Theorem

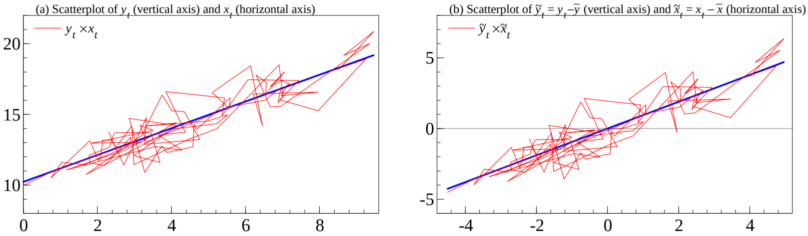
Consider the linear regression model

$$y_t = \mu + \beta_1 x_t + \varepsilon_t, \quad t = 1, 2, ..., T, \tag{5}$$

and the linear regression model for the de-meaned variables $\widetilde{y}_t = y_t - \bar{y}$ and $\widetilde{x}_t = x_t - \bar{x}$, given by

$$\widetilde{y}_t = b_1 \widetilde{x}_t + u_t, \quad t = 1, 2, ..., T. \tag{6}$$

Question: Would you expect $\hat{\beta}_1$ to equal $\hat{b}_1$?



The Frisch-Waugh-Lovell theorem tells us that $\hat{\beta}_1 = \hat{b}_1$.

Model

BEFORE: $x_{2t} = 1$.

Consider the more general model

$$y_t = x_{1t}'\beta_1 + x_{2t}'\beta_2 + \varepsilon_t, \quad t = 1, ..., T,$$

with x_{1t} $K_1 \times 1$ and x_{2t} $K_2 \times 1$.

The Frisch-Waugh-Lovell theorem tells us that we can obtain the OLS estimate of β_1 in two ways:

1. (The usual way) Regress y_t on $(x_{1t}', x_{2t}')'$ and obtain, $\hat{\beta}_1$.
2. (A three step procedure) Regress y_t on x_{2t} and obtain the residuals, $y_t^\star$. Next, regress x_{1t} on x_{2t} and obtain the residuals $x_{1t}^\star$. Lastly, regress $y_t^\star$ on $x_{1t}^\star$ and obtain $\hat{\beta}_1$.

The the two estimates are numerically identical.

$$y_t^* = x_{1t}^{*\prime} b_1 + u_t$$

A formal proof will be posted on Absalon.

FWL: $\hat{\beta}_1 = \hat{b}_1$.

Why is it a useful theorem?

In some cases, it will allow to greatly simplify the model:

- Demeaning variables: $x_{2t} = 1$.
- Detrending variables: See PS $\#1$.
- Removing fixed effects in a panel data model (Econometrics I).

7. Recap: Linear Regression Model with Time Series Data

Recap: $y_t = x_t'\beta + \varepsilon_t$

Main assumption: Cross-section vs. Time series

Cross-Section:	independent	and	identically distributed
Time Series:	weak dependence	and	stationarity

Technical requirements for the application of a **LLN** and a **CLT**.

OLS: Assumptions required for implementation and properties

	Computation $\widehat{\beta} = (X'X)^{-1}X'y$	Consistency $\text{plim}\,\widehat{\beta} = \beta$	Unbiasedness $E(\widehat{\beta}) = \beta$	Asympt. Distr. $\widehat{\beta} \overset{a}{\sim} Normal$
(1) No perfect collinearity in X variables				
(2) Strict exogeneity $E(\epsilon_t \mid x_1, x_2, \ldots, x_T) = 0$				
(3) Contemporaneous exogeneity (predeterminedness) $E(\epsilon_t \mid x_t) = 0$				
(4) Moment conditions $E(x_t \epsilon_t) = 0$				
(5) Conditional homoskedasticity $E(\epsilon_t^2 \mid x_t) = \sigma^2$				
(6) No serial correlation $E(\epsilon_t \epsilon_s \mid x_t, x_t) = 0 \quad \forall t \neq s$				

Note: Since (2) : $E(\epsilon_t \mid x_1, x_2, \ldots, x_T) = 0 \Rightarrow$ (3) : $E(\epsilon_t \mid x_t) = 0 \Rightarrow$ (4) : $E(\epsilon_t x_t) = 0$, but the reverse is not true, conditions ($*$) are stronger and therefore not necessary, as long as one of the less strong required assumptions mentioned in the table is fulfilled. ✓: In order obtain the "usual" OLS variance.

Recap: $y_t = x_t'\beta + \varepsilon_t$

Main assumption: Cross-section vs. Time series

Cross-Section:	independent	and	identically distributed
Time Series:	weak dependence	and	stationarity

Technical requirements for the application of a **LLN** and a **CLT**.

OLS: Assumptions required for implementation and properties

	Computation $\widehat{\beta} = (X'X)^{-1}X'y$	Consistency plim $\widehat{\beta} = \beta$	Unbiasedness $E(\widehat{\beta}) = \beta$	Asympt. Distr. $\widehat{\beta} \overset{a}{\sim}$ *Normal*
(1) No perfect collinearity in X variables	✓			
(2) Strict exogeneity $E(\epsilon_t \mid x_1, x_2, \ldots, x_T) = 0$				
(3) Contemporaneous exogeneity (predeterminedness) $E(\epsilon_t \mid x_t) = 0$				
(4) Moment conditions $E(x_t \epsilon_t) = 0$				
(5) Conditional homoskedasticity $E(\epsilon_t^2 \mid x_t) = \sigma^2$				
(6) No serial correlation $E(\epsilon_t \epsilon_s \mid x_t, x_t) = 0 \quad \forall t \neq s$				

Note: Since (2) : $E(\epsilon_t \mid x_1, x_2, \ldots, x_T) = 0 \Rightarrow$ (3) : $E(\epsilon_t \mid x_t) = 0 \Rightarrow$ (4) : $E(\epsilon_t x_t) = 0$, but the reverse is not true, conditions $(*)$ are stronger and therefore not necessary, as long as one of the less strong required assumptions mentioned in the table is fulfilled. ✓: In order obtain the "usual" OLS variance.

Recap: $y_t = x_t'\beta + \varepsilon_t$

Main assumption: Cross-section vs. Time series

Cross-Section:	independent	and	identically distributed
Time Series:	weak dependence	and	stationarity

Technical requirements for the application of a **LLN** and a **CLT**.

OLS: Assumptions required for implementation and properties

	Computation $\widehat{\beta} = (X'X)^{-1}X'y$	Consistency $\text{plim}\,\widehat{\beta} = \beta$	Unbiasedness $E(\widehat{\beta}) = \beta$	Asympt. Distr. $\widehat{\beta} \overset{a}{\sim} \textit{Normal}$
(1) No perfect collinearity in X variables	✓	✓		
(2) Strict exogeneity $E(\epsilon_t \mid x_1, x_2, \ldots, x_T) = 0$		$(*)$		
(3) Contemporaneous exogeneity (predeterminedness) $E(\epsilon_t \mid x_t) = 0$		$(*)$		
(4) Moment conditions $E(x_t\epsilon_t) = 0$		✓		
(5) Conditional homoskedasticity $E(\epsilon_t^2 \mid x_t) = \sigma^2$				
(6) No serial correlation $E(\epsilon_t\epsilon_s \mid x_t, x_t) = 0 \quad \forall t \neq s$				

Note: Since (2) : $E(\epsilon_t \mid x_1, x_2, \ldots, x_T) = 0 \Rightarrow$ (3) : $E(\epsilon_t \mid x_t) = 0 \Rightarrow$ (4) : $E(\epsilon_t x_t) = 0$, but the reverse is not true, conditions $(*)$ are stronger and therefore not necessary, as long as one of the less strong required assumptions mentioned in the table is fulfilled. ✓: In order obtain the "usual" OLS variance.

Recap: $y_t = x_t'\beta + \varepsilon_t$

Main assumption: Cross-section vs. Time series

Cross-Section:	independent	and	identically distributed
Time Series:	weak dependence	and	stationarity

Technical requirements for the application of a **LLN** and a **CLT**.

OLS: Assumptions required for implementation and properties

	Computation $\widehat{\beta} = (X'X)^{-1}X'y$	Consistency plim $\widehat{\beta} = \beta$	Unbiasedness $E(\widehat{\beta}) = \beta$	Asympt. Distr. $\widehat{\beta} \overset{a}{\sim} Normal$
(1) No perfect collinearity in X variables	✓	✓	✓	
(2) Strict exogeneity $E(\epsilon_t \mid x_1, x_2, \ldots, x_T) = 0$		$(*)$	✓	
(3) Contemporaneous exogeneity (predeterminedness) $E(\epsilon_t \mid x_t) = 0$		$(*)$		
(4) Moment conditions $E(x_t \epsilon_t) = 0$		✓		
(5) Conditional homoskedasticity $E(\epsilon_t^2 \mid x_t) = \sigma^2$				
(6) No serial correlation $E(\epsilon_t \epsilon_s \mid x_t, x_t) = 0 \quad \forall t \neq s$				

Note: Since (2) : $E(\epsilon_t \mid x_1, x_2, \ldots, x_T) = 0 \Rightarrow$ (3) : $E(\epsilon_t \mid x_t) = 0 \Rightarrow$ (4) : $E(\epsilon_t x_t) = 0$, but the reverse is not true, conditions $(*)$ are stronger and therefore not necessary, as long as one of the less strong required assumptions mentioned in the table is fulfilled. ✓: In order obtain the "usual" OLS variance.

Recap: $y_t = x_t'\beta + \varepsilon_t$

Main assumption: Cross-section vs. Time series

Cross-Section:	independent	and	identically distributed
Time Series:	weak dependence	and	stationarity

Technical requirements for the application of a **LLN** and a **CLT**.

OLS: Assumptions required for implementation and properties

	Computation $\widehat{\beta} = (X'X)^{-1}X'y$	Consistency plim $\widehat{\beta} = \beta$	Unbiasedness $E(\widehat{\beta}) = \beta$	Asympt. Distr. $\widehat{\beta} \overset{a}{\sim}$ *Normal*
(1) No perfect collinearity in X variables	✓	✓	✓	✓
(2) Strict exogeneity $E(\epsilon_t \mid x_1, x_2, \ldots, x_T) = 0$		$(*)$	✓	$(*)$
(3) Contemporaneous exogeneity (predeterminedness) $E(\epsilon_t \mid x_t) = 0$		$(*)$		$(*)$
(4) Moment conditions $E(x_t \epsilon_t) = 0$		✓		✓
(5) Conditional homoskedasticity $E(\epsilon_t^2 \mid x_t) = \sigma^2$				✓
(6) No serial correlation $E(\epsilon_t \epsilon_s \mid x_t, x_t) = 0 \quad \forall t \neq s$				✓

Note: Since (2) : $E(\epsilon_t \mid x_1, x_2, \ldots, x_T) = 0 \Rightarrow$ (3) : $E(\epsilon_t \mid x_t) = 0 \Rightarrow$ (4) : $E(\epsilon_t x_t) = 0$, but the reverse is not true, conditions $(*)$ are stronger and therefore not necessary, as long as one of the less strong required assumptions mentioned in the table is fulfilled. ✓: In order obtain the "usual" OLS variance.