Advanced Econometrics

Chapters 1 & 2: Introduction

Etienne Wijler

Econometrics and Data Science
Master Program

VU Vrije Universiteit Amsterdam
School of Business and Economics

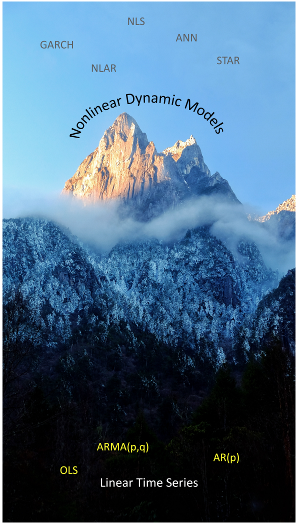
Agenda

Recap of the basics (start of our journey)

- Simple linear regression theory
- Linear time series theory

Looking up (road to the summit)

- Intractable estimators
- Properties of nonlinear dynamic models
- Model misspecification
- Predictive and causal analysis using complex models



Recap: The stuff you should already know

Simple linear regression:

- Deriving the OLS estimator
- Properties of the OLS estimator: Consistency
- Properties of the OLS estimator: Asymptotic distribution

Linear AR(1) model:

- Deriving the ML estimator
- Properties of the ML estimator: Consistency
- Properties of the ML estimator: Asymptotic distribution

Simple linear regression: optional literature

Optional literature

1. Wooldrige (2003), *"A Modern approach to econometrics"*
   - Chapter 2.1, 2.2, 5.1 and 5.2
2. Heij et al. (2004), *"Econometric methods with applications ..."*
   - Chapter 1.3.1, 2.1, 4.1.3 and 4.1.4
3. Stock and Watson. (2007), *"Introduction to Econometrics"*
   - Chapter 4.1, 4.2, 17.2 and 17.3
4. Verbeek. (2004), *"A guide to modern econometrics"*
   - Chapter 2.1, 2.2, 2.6.1 and 2.6.2

Simple linear regression

Linear regression model:

$$y_t = \alpha + \beta x_t + \epsilon_t$$

- y_t is the *dependent* or *endogenous* variable ("target" in machine learning)
- x_t is the *independent* or *explanotary* variable ("feature" in machine learning)
- ϵ_t is the *error term* or *innovation*
- (α, β) are the fixed *unknown parameters* (*intercept*,*slope*)

Important: withour further assumptions the model is meaningless!

Simple linear regression: exogeneity

Assumption (Exogenous regressors)

$\mathrm{E}(\epsilon_t | x_t) = 0 \; \forall t \in \mathbb{N}$.

Note: under this assumption, the linear regression model can be interpreted as a model of the *conditional expectation*

$$\mathrm{E}(y_t|x_t) = \mathrm{E}(\alpha + \beta x_t + \epsilon_t | x_t) = \alpha + \beta \underbrace{\mathrm{E}(x_t|x_t)}_{=x_t} + \underbrace{\mathrm{E}(\epsilon_t|x_t)}_{=0} = \alpha + \beta x_t$$

- The model tells us what value y_t will take *on average*, given a fixed value of x_t.
- The slope β measures the expected change in y_t, given a one unit change in x_t.
- The intercept α measures the average value of y_t when $x_t = 0$.

Simple Linear regression: estimation

Estimation: The parameters α and β are unknown, they need to be estimated.

Loss-functions: The most popular choice for estimating (α,β) is by minimizing a squared loss-function:

$$(\hat{\alpha}_T, \hat{\beta}_T) = \underset{(\alpha,\beta)}{\arg\min} \sum_{t=1}^{T} (y_t - \alpha - \beta x_t)^2$$

Ordinary Least Squares: This method is referred to as OLS. It was first published by Legendre in 1805, but later claimed to be invented by Gauss in 1795 at the age of 18!

Important: A squared loss-function is mathematically convenient, but not always the best choice. After this course, restricting yourself to OLS is not necessary anymore.

The regression line

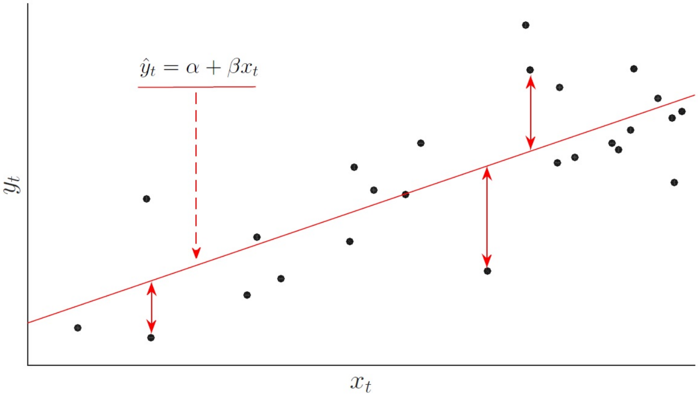


Figure: Observed data and the regression line with "fitted values"

Deriving the OLS estimator

For simplicity, assume that $\alpha = 0$, such that

$$\hat{\beta}_T = \underset{(\alpha,\beta)}{\arg\min} \sum_{t=1}^{T} (y_t - \beta x_t)^2$$

Then, the solution $\hat{\beta}_T$ is simply found by taking the derivate and equating it to zero:

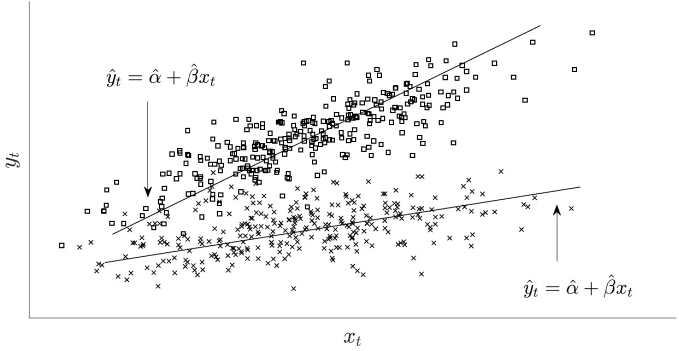
$$\frac{d \sum_{t=1}^{T} (y_t - \beta x_t)^2}{d\beta} = -2 \sum_{t=1}^{T} x_t (y_t - \beta x_t) \overset{s}{=} 0$$

$$\Rightarrow \hat{\beta}_T \sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T} x_t y_t \Rightarrow \hat{\beta}_T = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}$$

Note: I use $\overset{s}{=}$ to denote that we <u>set</u> the derivative equal to zero.

Sampling variability

- Different samples will lead to different estimates due to sampling variability.
- Formally, $\hat{\beta}_T$ is a random variable, because it is a function of the random data $\{y_1, \ldots, y_T\}$ and $\{x_1, \ldots, x_T\}$.



Question: What are the properties of the OLS estimator?

Properties of the OLS estimator: Consistency

Definition (Consistent Estimator)

An estimator $\hat{\beta}_T$ is said to be consistent for the parameter β if and only if $\hat{\beta}_T \xrightarrow{p} \beta$ as $T \to \infty$.

Note: $\hat{\beta}_T$ converges in probability to β if the probability that $\hat{\beta}_T$ is close to β increases to 1 as $T \to \infty$.

Assumption (Correct Specification)

The linear regression $y_t = \alpha + \beta x_t + \varepsilon_t$ is correctly specified.

Note: If the linear regression model is correctly specified. Then you know that each y_t is exactly given by $y_t = \alpha + \beta x_t + \varepsilon_t$, for some unknown $(\alpha, \beta) \in \mathbb{R}^2$.

Properties of the OLS estimator: Consistency

Example: Consider again the case where $\alpha = 0$.

Note: Under correct specification:

$$\begin{aligned} \hat{\beta}_T &= \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2} = \frac{\sum_{t=1}^{T} (\beta x_t + \varepsilon_t) x_t}{\sum_{t=1}^{T} x_t^2} \\ &= \underbrace{\beta}_{\text{true parameter}} + \underbrace{\frac{\sum_{t=1}^{T} x_t \varepsilon_t}{\sum_{t=1}^{T} x_t^2}}_{\text{remainder term}} \end{aligned}$$

Important: $\hat{\beta}_T \xrightarrow{p} \beta$ if the remainder term vanishes to zero!

Properties of the OLS estimator: Consistency

Note: To make the remainder term vanish, we just have to apply a *Law of Large Numbers* (LLN)!

Theorem (Bernoulli's Law of Large Numbers)

Let $z_1, \ldots, z_T$ *be iid random variables with* $\mathrm{E}\,|z_1| < \infty$. *Then*

$$\frac{1}{T}\sum_{t=1}^{T} z_t \xrightarrow{p} \mathrm{E}(z_t) \quad as \quad T \to \infty.$$

LLN is a remarkable piece of mathematics! First stated without proof by Cardano (1500's). First proof in Bernoulli's 1713 *"Ars Conjectandi"* (more than 20 years to complete the *Golden Theorem!*). LLN coined in 1837 by Poisson. Simpler proof derived by Chebyshev in 1874 using unproved inequality. Markov wrote complete proof in 1884.

Properties of the OLS estimator: Consistency

Applying the LLN to the remainder term:

1. Apply LLN to numerator and denominator

$$\frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t \xrightarrow{p} \mathrm{E}(x_t \varepsilon_t) \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T} x_t^2 \xrightarrow{p} \mathrm{E}(x_t^2) \quad \text{as } T \to \infty.$$

2. Apply continuous mapping theorem

$$\frac{\sum_{t=1}^{T} x_t \varepsilon_t}{\sum_{t=1}^{T} x_t^2} \xrightarrow{p} \frac{\mathrm{E}(x_t \varepsilon_t)}{\mathrm{E}\, x_t^2} = \frac{0}{\mathrm{E}\, x_t^2} = 0 \quad \text{as} \quad T \to \infty$$

Note: $\mathrm{E}(x_t \varepsilon_t) = \mathrm{E}\left(\mathrm{E}(x_t \varepsilon_t | x_t)\right) = \mathrm{E}\left(x_t\, \mathrm{E}(\varepsilon_t | x_t)\right) = 0.$

Properties of the OLS estimator: Consistency

Theorem (Consistency of OLS)

Let the observed data $\{y_t\}$ and $\{x_t\}$ be obtained from a correctly specified linear regression model

$$y_t = \beta x_t + \varepsilon_t \quad \forall\, t \in \mathbb{N}$$

with (i) *exogenous regressors* $\mathrm{E}(\varepsilon_t | x_t) = 0$,
(ii) *iid sequences* $\{x_t \varepsilon_t\}$ *and* $\{x_t^2\}$,
(iii) *finite moments* $\mathrm{E}\,|x_t \varepsilon_t| < \infty$ *and* $\mathrm{E}\,|x_t^2| < \infty$.

Then $\hat{\beta}_T$ *is consistent for* β, *i.e.* $\hat{\beta}_T \xrightarrow{p} \beta$ *as* $T \to \infty$.

Properties of the OLS estimator: Asymptotic Normality

Important: Econometricians are often interested in conducting statistical inference on the unknown parameter!

Example: IF the distribution of $\hat{\beta}_T$ is known under $H_0 : \beta = 4$. THEN we can calculate the probability of observing $\hat{\beta}_T = 3.5$.

- If probability of observing the estimate $\hat{\beta}_T = 3.5$ under the null hypothesis that $\beta = 4$ is very low, then we reject H_0.
- If the probability of observing $\hat{\beta}_T = 3.5$ under the null hypothesis H_0 is high, then we do not reject H_0.

Problem: In advanced settings the distribution of the estimator is not known exactly!

Solution: Derive 'approximate' distribution using a CLT.

Properties of the OLS estimator: Asymptotic Normality

Definition (Asymptotic normality)

An estimator $\hat{\beta}_T$ is said to be asymptotically normal for the parameter β if

$$\sqrt{T}\big(\hat{\beta}_T - \beta\big) \xrightarrow{d} N(0, \sigma^2) \quad \text{as} \;\; T \to \infty.$$

Note: it is easy to show that $\hat{\beta}_T$ is asymptotically normal!

Just rewrite $\sqrt{T}(\hat{\beta}_T - \beta)$ as

$$\sqrt{T}(\hat{\beta}_T - \beta) \;=\; \sqrt{T}\left(\beta + \frac{\sum_{t=1}^T x_t \varepsilon_t}{\sum_{t=1}^T x_t^2} - \beta\right) \;=\; \sqrt{T}\frac{\sum_{t=1}^T x_t \varepsilon_t}{\sum_{t=1}^T x_t^2}$$

and now apply a *Central Limit Theorem* to the remainder term!

Properties of the OLS estimator: Asymptotic Normality

Theorem (Lindeberg-Levy's Central Limit Theorem)

Let $z_1, \dots, z_T$ *be iid random variables with* $\mathrm{E}(z_1) = \mu$ *and* $\mathbb{V}\mathrm{ar}(z_1) = \sigma^2 < \infty$. *Then*

$$\sqrt{T}\Big(\frac{1}{T}\sum_{t=1}^{T} z_t - \mu\Big) \xrightarrow{d} N(0, \sigma^2) \quad \textit{as} \quad T \to \infty.$$

The idea: the standardized sample average converges in distribution to a normal random variable!

The CLT and LLN are the most important results in probability and statistics! de Moivre provided a remarkable first proof of the CLT in 1733. Unfortunately, the result passed unnoticed to everyone else.
Laplace rediscovered the CLT in 1812. Again, few understood it.
Finally, Lyapunov brought the theorem to fame in 1901!

Hyping up the Central Limit Theorem

"I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along."

- Sir Francis Galton* (1889)

*Unfortunately also famous for controversial and (nowadays) obscure ideas on eugenics. Needless to say, this quote ought to be read independent of those ideas.

Properties of the OLS estimator: Asymptotic Normality

Applying the CLT to the remainder term:

1. Multiply the numerator and denominator by $1/T$ to obtain

$$\sqrt{T}(\hat{\beta}_T - \beta) \;=\; \sqrt{T}\frac{\frac{1}{T}\sum_{t=1}^{T} x_t\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2} \;=\; \frac{\sqrt{T}\Big(\frac{1}{T}\sum_{t=1}^{T} x_t\varepsilon_t - \mathrm{E}(x_t\varepsilon_t)\Big)}{\frac{1}{T}\sum_{t=1}^{T} x_t^2}.$$

   Note: we can subtract $\mathrm{E}(x_t\varepsilon_t)$ because $\mathrm{E}(x_t\varepsilon_t) = 0$.

2. Apply a CLT to numerator and LLN to denominator:

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} x_t\varepsilon_t - \mathrm{E}(x_t\varepsilon_t)\right) \xrightarrow{d} N\Big(0, \underbrace{\mathbb{V}\mathrm{ar}(x_t\varepsilon_t)}_{=\sigma^2\,\mathrm{E}(x_t^2) \text{ if } \epsilon \perp\!\!\!\perp x_t}\Big) \quad \text{as} \quad T \to \infty,$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t^2 \xrightarrow{p} \mathrm{E}(x_t^2) \quad \text{as} \quad T \to \infty.$$

Properties of the OLS estimator: Asymptotic Normality

As a result: by *Slutsky's Theorem* we have

$$\sqrt{T}(\hat{\beta}_T - \beta) \;=\; \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2} \;\xrightarrow{d}\; \frac{N\big(0, \sigma_\varepsilon^2\,\mathrm{E}(x_t^2)\big)}{\mathrm{E}(x_t^2)}$$

Theorem (Asymptotic normality of OLS)

Let the observed data $\{y_t\}$ and $\{x_t\}$ be obtained from the correctly specified linear regression model

$$y_t = \beta x_t + \varepsilon_t \quad \forall\, t \in \mathbb{N},$$

with (i) exogenous regressors $\mathrm{E}(\varepsilon_t|x_t) = 0$,
(ii) iid sequences $\{x_t\varepsilon_t\}$ *and* $\{x_t^2\}$,
(iii) bounded moments $\mathrm{E}\,|x_t^2| < \infty$, $\mathrm{E}\,|x_t\varepsilon_t|^2 < \infty$, $\mathbb{V}\mathrm{ar}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$ *and* $x_t \perp\!\!\!\perp \varepsilon_t$.

$$\textit{Then} \quad \sqrt{T}(\hat{\beta}_T - \beta) \xrightarrow{d} N\Big(0, \sigma_\varepsilon^2\big[\,\mathrm{E}(x_t^2)\big]^{-1}\Big) \quad \textit{as} \;\; T \to \infty.$$

Properties of the OLS estimator: Asymptotic Normality

Problem: the true expectation $\mathrm{E}(x_t^2)$ is unknown!

In practice: we approximate it by $\frac{1}{T}\sum_{t=1}^{T} x_t^2$.

We use the approximate distribution:

$$\sqrt{T}(\hat{\beta}_T - \beta) \overset{\text{approx}}{\sim} N\Big(0\ ,\ \sigma_\varepsilon^2\big(\frac{1}{T}\sum_{t=1}^{T} x_t^2\big)^{-1}\Big).$$

$$\hat{\beta}_T \overset{\text{approx}}{\sim} N\Big(\beta\ ,\ \sigma_\varepsilon^2\big(\sum_{t=1}^{T} x_t^2\big)^{-1}\Big).$$

Note: This is very useful!

Example: Suppose $\hat{\beta}_T = 3.5$. Should we believe the hypothesis $H_0 : \beta = 0$?

Linear AR(1) model: further reading material

1. *Wooldrige (2003), "A Modern approach to econometrics"*
   - *Chapters 10.1, 10.2, 11.1 and 11.2*
2. *Heij et al. (2004), "Econometric methods with applications ..."*
   - *Chapters 1.3.1, 4.1.4, 4.3.1 – 4.3.3, 7.1.1 – 7.1.3 and 7.2.2*
3. *Stock and Watson. (2007), "Introduction to Econometrics"*
   - *Chapters 14.1–14.3*
4. *Verbeek. (2004), "A guide to modern econometrics"*
   - *Chapters 6.1, 8.1, 8.2 and 8.6.2*

Linear AR(1) model

Linear Gaussian AR(1) model:

$$x_t = \phi x_{t-1} + \varepsilon_t \quad \{\varepsilon_t\}_{t\in\mathbb{Z}} \sim \mathrm{NID}(0, \sigma^2).$$

Assumption (Exogenous regressors)

$\mathrm{E}(\varepsilon_t | x_{t-1}) = 0 \;\; \forall\, t \in \mathbb{Z}.$

Note: Conditional expectation $\mathrm{E}(x_t | x_{t-1}) = \phi x_{t-1}$.

In practice: AR(1) is very useful in modeling temporal dependence present in economic and financial time-series.

Temporal dependence: $x_t | x_{t-1} \sim N(\phi x_{t-1}, \sigma_\varepsilon^2)$.

Linear AR(1) model

Definition (Strict stationarity)

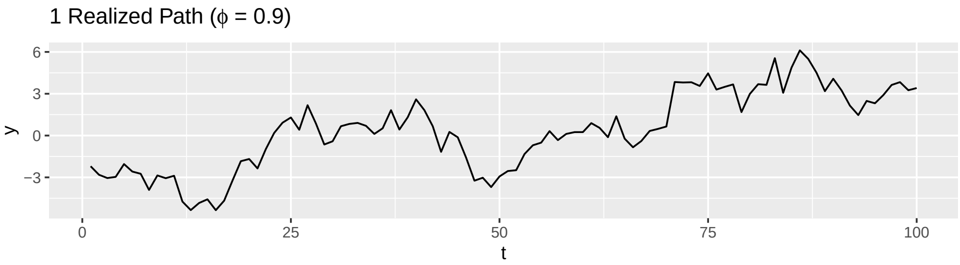
A time-series $\{x_t\}_{t \in \mathbb{Z}}$ is said to be strictly stationary if the distribution of every finite sub-vector is invariant in time

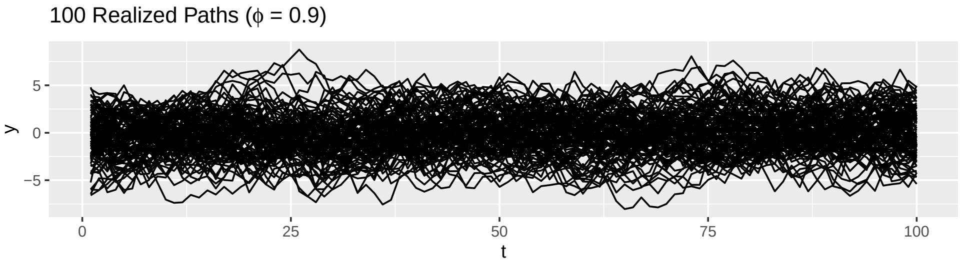
$$(x_1, ..., x_h) \overset{d}{=} (x_{t+1}, ..., x_{t+h}), \quad \forall \, (t, h) \in \mathbb{N} \times \mathbb{N}.$$

The idea: A time series $\{x_t\}_{t \in \mathbb{Z}}$ is stationary if its *unconditional* distribution is invariant in time.

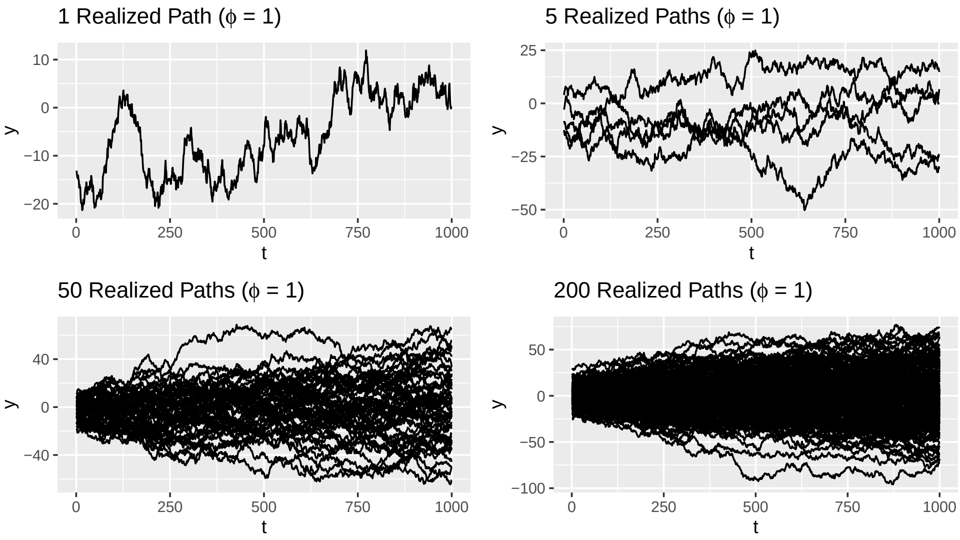
Introductory econometrics AR(1): IF $|\phi| < 1$ THEN we have a time-varying *conditional distribution* but time-invariant *unconditional distribution*!

Linear AR(1) model with $|\phi| < 1$





Linear AR(1) model with $\phi = 1$



Linear AR(1) model

Theorem (Stationarity of the AR(1) model)

Let $\{x_t\}_{t\in\mathbb{Z}}$ be a time-series generated by the linear Gaussian AR(1) model

$$x_t = \phi x_{t-1} + \varepsilon_t \qquad \forall\ t \in \mathbb{Z}$$

with $|\phi| < 1$ and innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ that are NID$(0, \sigma^2_\varepsilon)$. Then $\{x_t\}_{t\in\mathbb{Z}}$ is strictly stationary.

Note: Gaussian AR(1) is stationary if it does not exhibit 'too much' temporal dependence.

Important: stationarity property of the time-series allows us to make use of LLNs and CLTs!

Deriving the ML estimator

Introductory econometrics: You learned about the famous *maximum likelihood* (ML) estimator first introduced by Gauss in 1809.

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} f(x_1, ..., x_T; \boldsymbol{\theta}).$$

- Parameter vector $\boldsymbol{\theta} = (\phi, \sigma_\varepsilon^2)$,
- $f(x_1, ..., x_T; \boldsymbol{\theta})$ is the *joint pdf* of $(x_1, ..., x_T)$, which we call the *likelihood function*.

$$\text{Note: } f(x_1, ..., x_T; \boldsymbol{\theta}) = f(x_1; \boldsymbol{\theta}) \times \prod_{t=2}^{T} f(x_t | x_{t-1}, ...x_1; \boldsymbol{\theta}).$$

Deriving the ML estimator

Important: Re-writing the likelihood as a product of conditional densities is very useful!

Indeed: we saw that in the linear Gaussian AR(1) model:

$$x_t | x_{t-1} \ \sim \ N(\phi x_{t-1}, \sigma_\varepsilon^2).$$

Hence: $f(x_t | x_{t-1}, ...; \boldsymbol{\theta})$ is given by the well known formula

$$f(x_t | x_{t-1}, ... x_1; \boldsymbol{\theta}) \ = \ f(x_t | x_{t-1}; \boldsymbol{\theta}) \ = \ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[- \frac{(x_t - \phi x_{t-1})^2}{2\sigma^2} \Big].$$

As a result: in terms of the log-likelihood we have:

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} \sum_{t=2}^{T} - \log \sqrt{2\pi\sigma^2} - \frac{(x_t - \phi x_{t-1})^2}{2\sigma^2}.$$

Deriving the ML estimator

Fortunately: finding the ML estimator is easy! Just take the derivative of the likelihood function and set it to zero!

Example: Suppose $\sigma_\varepsilon^2 = 1$. Then $\boldsymbol{\theta} = \phi$ and we have

$$\frac{\partial \log f(x_1, ..., x_T; \phi)}{\partial \phi} = \sum_{t=2}^{T} (x_t - \phi x_{t-1}) x_{t-1}$$

and since, $\hat{\phi}_T$ sets the derivative to zero

$$\frac{\partial \log f(x_1, ..., x_T; \hat{\phi}_T)}{\partial \phi} = 0 \;\Leftrightarrow\; \sum_{t=2}^{T} (x_t - \hat{\phi}_T x_{t-1}) x_{t-1} = 0,$$

$$\text{we conclude that:} \quad \hat{\phi}_T = \frac{\sum_{t=2}^{T} x_t x_{t-1}}{\sum_{t=2}^{T} x_{t-1}^2}.$$

Homework: Verify that this is also the OLS estimator.

Properties of the ML estimator: Consistency

Consistency of MLE: Follows by appealing to an LLN.

Assumption (Correct specification)

The Gaussian linear AR(1) model $x_t = \phi x_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim NID(0, \sigma_\varepsilon^2)$ *is correctly specified.*

Note: We can substitute x_t by $\phi x_{t-1} + \varepsilon_t$

$$\hat{\phi}_T = \frac{\sum_{t=2}^T x_t x_{t-1}}{\sum_{t=2}^T x_{t-1}^2} = \frac{\sum_{t=2}^T (\phi x_{t-1} + \varepsilon_t) x_{t-1}}{\sum_{t=2}^T x_{t-1}^2} = \phi + \frac{\sum_{t=2}^T \varepsilon_t x_{t-1}}{\sum_{t=2}^T x_{t-1}^2}.$$

Note: We can re-write the remainder as

$$\frac{\sum_{t=2}^T \varepsilon_t x_{t-1}}{\sum_{t=2}^T x_{t-1}^2} = \frac{\frac{1}{T}\sum_{t=2}^T \varepsilon_t x_{t-1}}{\frac{1}{T}\sum_{t=2}^T x_{t-1}^2}.$$

Properties of the ML estimator: Consistency

Finally: we obtain consistency by applying an LLN:

$$\frac{\frac{1}{T}\sum_{t=2}^{T}\varepsilon_t x_{t-1}}{\frac{1}{T}\sum_{t=2}^{T}x_{t-1}^2} \xrightarrow{p} \frac{\mathbb{E}(\varepsilon_t x_{t-1})}{\mathbb{E}(x_t^2)} = \frac{0}{\mathbb{E}(x_t^2)} = 0$$

Problem: $\{x_t\}$ is not iid. $\{\varepsilon_t x_{t-1}\}$ and $\{x_t^2\}$ are also not iid.

Solution: Apply the LLN for stationary ergodic data!

Theorem (Birkhoff-Khinchin Theorem)

Let $\{z_t\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic random sequence with finite first moment $\mathbb{E}|z_1| < \infty$. Then we have

$$\frac{1}{T}\sum_{t=2}^{T} z_t \xrightarrow{p} \mathbb{E}(z_t) \quad as \quad T \to \infty.$$

Properties of the ML estimator: Consistency

Applying the LLN: we obtain the consistency

$$\hat{\phi}_T \;=\; \phi + \frac{\sum_{t=2}^{T} \varepsilon_t x_{t-1}}{\sum_{t=2}^{T} x_{t-1}^2} \xrightarrow{p} \phi + 0 \;=\; \phi \quad \text{as } T \to \infty.$$

Theorem (Consistency of ML estimator)

Let $\{x_t\}_{t\in\mathbb{Z}}$ be generated by a strictly stationary ($|\phi| < 1$), correctly-specified linear Gaussian AR(1) model

$$x_t = \phi x_{t-1} + \varepsilon_t \quad \{\varepsilon_t\}_{t\in\mathbb{Z}} \sim NID(0, \sigma_\varepsilon^2),$$

with $\mathbb{E}(\varepsilon_t | x_{t-1})$ and $\sigma_\varepsilon^2 < \infty$. Then, $\hat{\phi}_T \xrightarrow{p} \phi$ as $T \to \infty$.

Note: all stationarity and moment conditions are satisfied! Why?

Properties of the ML estimator: Asymptotic Normality

Important: It is very easy to establish the asymptotic normality of the ML estimator!

1. (Correct-specification) Substitute x_t by $\phi x_{t-1} + \varepsilon_t$ and re-write $\sqrt{T}(\hat{\phi}_T - \phi)$ as

$$\sqrt{T}(\hat{\phi}_T - \phi) \;=\; \sqrt{T}\Big(\phi + \frac{\sum_{t=2}^{T} x_{t-1}\varepsilon_t}{\sum_{t=2}^{T} x_{t-1}^2} - \phi\Big) \;=\; \sqrt{T}\frac{\sum_{t=2}^{T} x_{t-1}\varepsilon_t}{\sum_{t=2}^{T} x_{t-1}^2}.$$

2. Re-write remainder term as

$$\sqrt{T}(\hat{\phi}_T - \phi) \;=\; \sqrt{T}\frac{\frac{1}{T}\sum_{t=2}^{T} x_{t-1}\varepsilon_t}{\frac{1}{T}\sum_{t=2}^{T} x_{t-1}^2} \;=\; \frac{\sqrt{T}\Big(\frac{1}{T}\sum_{t=2}^{T} x_{t-1}\varepsilon_t - \mathbb{E}(x_{t-1}\varepsilon_t)\Big)}{\frac{1}{T}\sum_{t=2}^{T} x_{t-1}^2}.$$

3. Apply a CLT to the numerator and an LLN to the denominator!

Properties of the ML estimator: Asymptotic Normality

Problem: we cannot make use of CLT for iid data.
Solution: Apply the CLT for stationary and ergodic martingale difference sequences (proved by Billingsley in 1961)

Definition (Martingale difference sequence)

A sequence $\{z_t\}_{t\in\mathbb{Z}}$ is said to be a martingale difference sequence if and only if $\mathbb{E}(z_t|z_{t-1}, z_{t-2}, ...) = 0 \ \forall \ t \in \mathbb{Z}$.

Theorem (Billingsley's Central Limit Theorem)

Let $\{z_t\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic martingale difference sequence of random variables with $\mathbb{E}(z_1) = \mu$, $\mathbb{V}\mathrm{ar}(z_1) = \sigma^2 < \infty$. Then

$$\sqrt{T}\Big(\frac{1}{T}\sum_{t=2}^{T} z_t - \mu\Big) \xrightarrow{d} N(0, \sigma^2) \quad as \quad T \to \infty.$$

Properties of the ML estimator: Asymptotic Normality

Applying the CLT and LLN:

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=2}^{T} x_{t-1}\varepsilon_t - \mathbb{E}(x_{t-1}\varepsilon_t)\right) \xrightarrow{d} N\Big(0, \sigma_\varepsilon^2 \mathbb{E}(x_{t-1}^2)\Big) \text{ as } T \to \infty,$$

$$\frac{1}{T}\sum_{t=2}^{T} x_{t-1}^2 \xrightarrow{p} \mathbb{E}(x_{t-1}^2) \text{ as } T \to \infty.$$

Note: The term $\mathbb{E}(x_{t-1}\varepsilon_t)$ is zero

$$\begin{aligned}
\mathbb{E}\big(x_{t-1}\varepsilon_t|\mathcal{F}_{t-1}\big) &= \mathbb{E}\Big(\mathbb{E}\big(x_{t-1}\varepsilon_t|x_{t-1},\mathcal{F}_{t-1}\big)\Big|\mathcal{F}_{t-1}\Big)\\
&= \mathbb{E}\Big(x_{t-1}\mathbb{E}\big(\varepsilon_t|x_{t-1},\mathcal{F}_{t-1}\big)\Big|\mathcal{F}_{t-1}\Big)\\
&= \mathbb{E}\Big(x_{t-1}\mathbb{E}(\varepsilon_t)\Big|\mathcal{F}_{t-1}\Big) = \mathbb{E}\big(x_{t-1}\cdot 0|\mathcal{F}_{t-1}\big) = 0.
\end{aligned}$$

Properties of the ML estimator: Asymptotic Normality

Finally: by Slutsky's theorem we obtain

$$\sqrt{T}(\hat{\phi}_T - \phi) \;=\; \frac{\frac{1}{\sqrt{T}}\sum_{t=2}^{T} x_{t-1}\varepsilon_t}{\frac{1}{T}\sum_{t=2}^{T} x_{t-1}^2} \quad \xrightarrow{d} \quad \frac{N\big(0, \sigma_\varepsilon^2 \mathbb{E}(x_{t-1}^2)\big)}{\mathbb{E}(x_{t-1}^2)}$$

Theorem (Asymptotic Normality of ML estimator)

Let $\{x_t\}_{t\in\mathbb{Z}}$ be generated by a strictly stationary ($|\phi| < 1$), correctly-specified, linear Gaussian AR(1) model

$$x_t = \phi x_{t-1} + \varepsilon_t \quad \forall\, t \in \mathbb{Z}$$

with $\mathbb{E}(\varepsilon_t|x_{t-1}) = 0$ and $\mathbb{V}\mathrm{ar}(\varepsilon) = \sigma_\varepsilon^2 < \infty$ Then,

$$\sqrt{T}(\hat{\phi}_T - \phi) \;\xrightarrow{d}\; N\Big(0\;,\;\sigma^2\big[\mathbb{E}(x_{t-1}^2)\big]^{-1}\Big) \quad \textit{as} \;\; T \to \infty.$$

Note: all stationarity and moment conditions are satisfied!

The challenges ahead...

Intractable estimators

Establishing stationarity

Incorrect specification

The questions we would like to ask!

Complications of general models

VERY IMPORTANT: results obtained until now were only possible because the models we considered were sufficiently simple and correctly specified.

1. the linearity of the models was crucial for obtaining an **analytical expression** for the estimator;
2. the linearity of the model was important in establishing the strict **stationarity** of the time-series;
3. the correct specification assumption was essential for writing the estimator in deviations from the **true parameter**.

Problem: Things become much more complicated if we consider general nonlinear models and allow for the possibility of mis-specification!

The challenges ahead...

Intractable estimators

Establishing stationarity

Incorrect specification

The questions we would like to ask!

Difficulties with nonlinear regression

Nonlinearities in Economics: Economic theory tells us that many economic variables are nonlinearly related.

Example: *AK production function* with *additive* errors where output y_t is a function of capital x_t according to

$$y_t = \alpha x_t^{\beta} + \varepsilon_t$$

where $\alpha > 0$ and $\beta \in (0,1)$ are unknown parameters.

LS estimator: suppose that $\alpha = 1$. Then

$$\hat{\beta}_T = \arg\min_{\beta} \sum_{t=1}^{T} \left(y_t - x_t^{\beta}\right)^2.$$

OK: Until now everything is fine!

Difficulties with nonlinear regression

Problem: Can you derive the expression of $\hat{\beta}_T$?

Introductory approach: set the derivative to zero!

$$\frac{\partial \sum_{t=1}^{T}(y_t - x_t^{\beta})^2}{\partial \beta} \overset{s}{=} 0 \quad \Leftrightarrow \quad -2\sum_{t=1}^{T}(y_t - x_t^{\hat{\beta}})x_t^{\hat{\beta}}\log(x_t) = 0.$$

Problem: Can you solve this equation for $\hat{\beta}_T$?

Answer: It will be very difficult and many times impossible!

Question: Without an expression for the estimator, how can we study its properties?

Practice: It is easy to obtain estimates! Just minimize the least-squares function numerically!

Difficulties with nonlinear regression

Example: Consider the simple *exponential growth* model with additive errors:

$$y_t = \alpha e^{\beta t} + \epsilon_t.$$

These models are typically used to model *population* size over time.

LS estimator: suppose that $\alpha = 1$. Then,

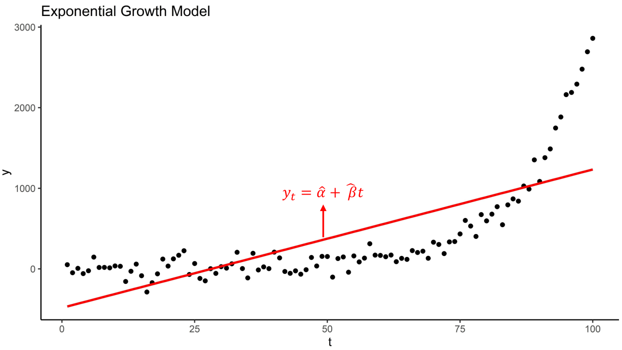
$$\hat{\beta}_T = \arg\min_{\beta} \sum_{t=1}^{T} \left(y_t - e^{\beta t} \right)^2 .$$

Introductory approach: Setting the derivative to zero is again difficult:

$$\frac{\partial \sum_{t=1}^{T} \left(y_t - e^{\beta t} \right)^2}{\partial \beta} \stackrel{s}{=} 0 \quad \Leftrightarrow \quad -2 \sum_{t=1}^{T} t e^{\hat{\beta} t} \left(y_t - e^{\hat{\beta} t} \right) = 0.$$

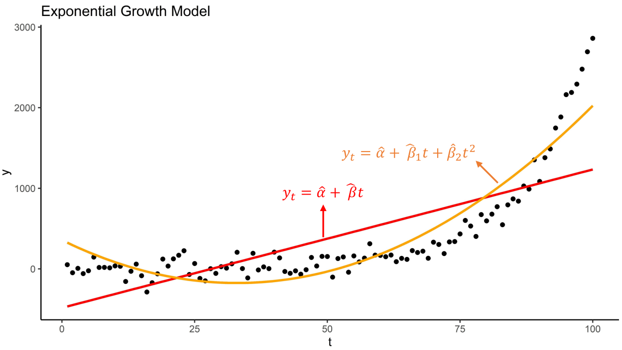
Difficulties with nonlinear regression

Question: Can we not just use a linear or quadratic trend instead?



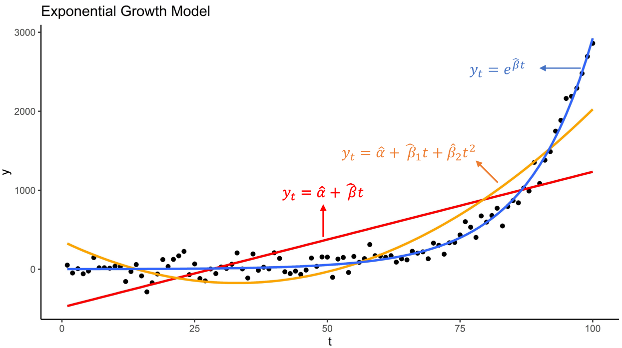
Difficulties with nonlinear regression

Question: Can we not just use a linear or quadratic trend instead?



Difficulties with nonlinear regression

Question: Can we not just use a linear or quadratic trend instead?



Answer: Sometimes... Officially, we would need infinitely many terms according to the Taylor series representation: $e^{\beta t} = \sum_{k=0}^{\infty} \frac{\beta t^k}{k!} = \sum_{k=0}^{\infty} \beta_k t^k$, where $\beta_k = \frac{\beta}{k!}$.

The challenges ahead...

Intractable estimators

Establishing stationarity

Incorrect specification

The questions we would like to ask!

Difficulties with nonlinear models: stationarity

Problem: it is difficult to find stationarity conditions for time-series with nonlinear dynamics.

Example: Linear AR(1) model

$$x_t = \phi x_{t-1} + \varepsilon_t \quad \forall\, t \in \mathbb{Z}.$$

Introductory econometrics:

1. AR(1) model is stable if $|\phi| < 1$
2. $\{x_t\}_{t \in \mathbb{Z}}$ is strictly stationary if $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is mean-zero iid sequence.

Recall: IF $|\phi| < 1$, THEN the *conditional distribution* is time-varying, but *unconditional distribution* is not!

Difficulties with nonlinear models: stationarity

Example: Nonlinear AR(1)

$$x_t = g(x_{t-1}; \beta) + \varepsilon_t \quad \forall\, t \in \mathbb{Z}$$

Question: under what conditions is $\{x_t\}_{t\in\mathbb{Z}}$ stationary?

Problem: without knowing the stochastic properties of $\{x_t\}_{t\in\mathbb{Z}}$ we cannot apply LLNs and CLTs.

Difficulties with nonlinear models: stationarity

Example: GARCH model

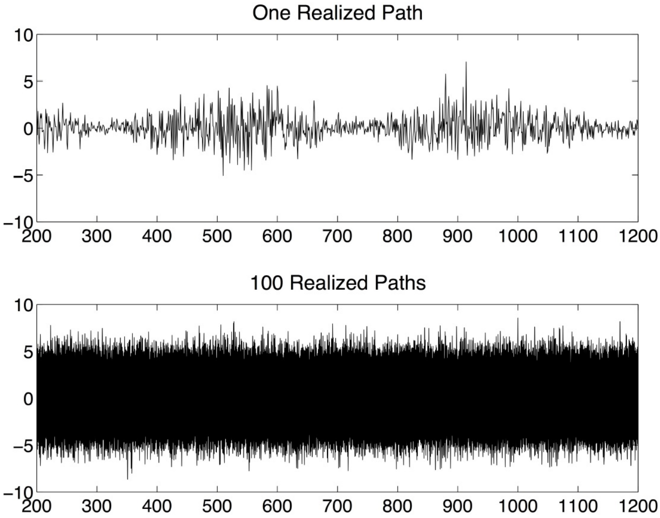
$$y_t = \sigma_t \varepsilon_t \ , \quad \varepsilon_t \sim N(0,1) \ ,$$

$$\sigma_{t+1}^2 = \omega + \alpha y_t^2 + \beta \sigma_t^2 \ \forall \ t \in \mathbb{Z}.$$

Question: for which values of ω, α and β is $\{y_t\}_{t\in\mathbb{Z}}$ stationary?

Important: We know that $\{y_t\}_{t\in\mathbb{Z}}$ has a time-varying *conditional variance*. However, its *unconditional variance* may be invariant!

Difficulties with nonlinear models: stationarity



The challenges ahead...

Intractable estimators

Establishing stationarity

Incorrect specification

The questions we would like to ask!

Difficulties with incorrect specification

"All models are wrong, some are useful." - George Box

Problem: Correct specification assumption is very restrictive!

1. No one really believes that such simple models are an exact description of the relation between economic variables.
2. DGP is most likely an immensely complex process involving potentially infinitely many variables.
3. Econometric models are simplifications of reality.

Problem: Even if the estimator is tractable and the properties of the model are known, we cannot analyze the properties of estimators in the usual way if the model is mis-specified!

Difficulties with mis-specified regression models

Example: OLS in linear regression

$$\hat{\beta}_T = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2}. \tag{1}$$

Important: we used the correct-specification assumption to substitute y_t by $\beta x_t + \varepsilon_t$ and obtain

$$\hat{\beta}_T = \beta + \frac{\sum_{t=1}^{T} \varepsilon_t x_t}{\sum_{t=1}^{T} x_t^2}.$$

Problem: If we make no specification assumption we do not know how y_t is related to x_t.

In practice: We can still estimate the model using the expression in (1)!

Difficulties with mis-specified regression models

If the data is well behaved: the estimator will still converge to some point!

Example: if $\{y_t\}_{t\in\mathbb{N}}$ and $\{x_t\}_{t\in\mathbb{N}}$ is iid with $\mathbb{E}|y_t x_t| < \infty$ and $\mathbb{E}|x_t|^2 < \infty$, we can apply LLNs to obtain

$$\hat{\beta}_T = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2} \xrightarrow{p} \frac{\mathbb{E}(y_t x_t)}{\mathbb{E}(x_t^2)} \quad \text{as} \;\; T \to \infty.$$

Question: what is this limit point? How can we interpret it? Does it have any meaning?

The challenges ahead...

Intractable estimators

Establishing stationarity

Incorrect specification

The questions we would like to ask!

The questions we would like to answer

1. What are the stochastic properties of nonlinear dynamic models (e.g. stationarity and bounded moments)?
2. What are the properties of estimators that are analytically intractable (e.g. consistency and asymptotic normality)?
3. What are the properties of estimators when the model is mis-specified and how can we interpret parameter estimates?
4. How can we use (possibly incorrect) complex models for predictive and structural analysis?