Linear Time Series

Christian Francq

http://christian.francq140.free.fr/

CREST-ENSAE

Chapter 1: Introduction

Motivation: numerous data sets are time indexed. The analysis of such data requires specific tools and models.

Aim: presenting the univariate and multivariate linear time series models in discrete time.

References: Pamplemousse +

- Brockwell and Davis (1991) Time Series: Theory and Methods. Springer Verlag.
- Brockwell and Davis (2002) Introduction to Time Series and Forecasting. Springer Verlag.
- Shumway and Stoffer (2000) Time series analysis and its applications. New York: springer.
- Gouriéroux et Monfort (1995) Séries temporelles et modèles dynamiques. Economica.
- Hamilton (1994) Times Series Analysis. Princeton University Press.
- Box and Jenkins (1970) Time Series Analysis: Forecasting and Control. Holden-Day.

Plan of the course

1. Introduction (stationnarity, models, estimation)
2. ARMA models
3. Practical use of ARMA and SARIMA
4. Unit root tests
5. VAR and cointegration
6. Asymptotic properties of the OLS estimator and of the unit root tests

Plan of the chapter

Time series

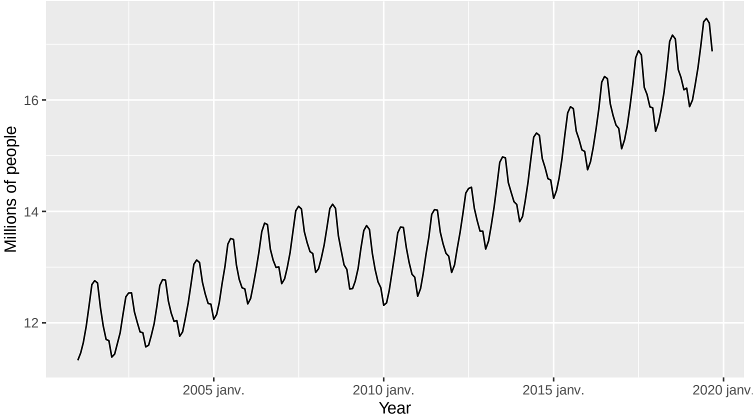
sequence (X_t) of observations of a same quantity indexed by a date $t \in \mathscr{T}$, where $\mathscr{T}$ is a discrete set of dates with a fixed interval between dates (one can assume $\mathscr{T} \subset \mathbb{Z}$).

> *Contrary to a sample, the order of the observations is important (arrow of time)*

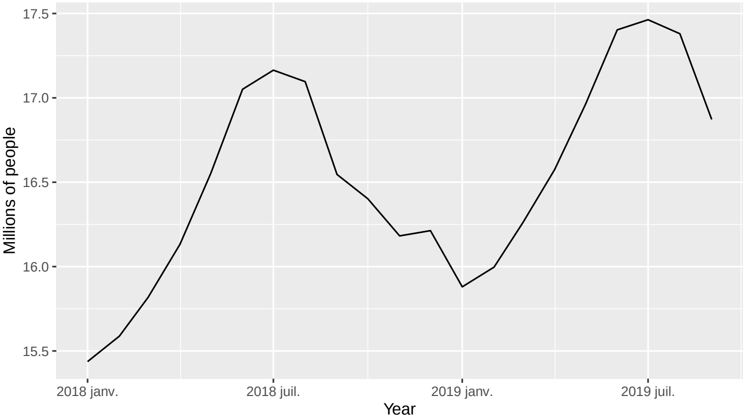
Examples:

- Monthly leisure and hospitality jobs;
- Black and white pepper price (bivariate time series);
- Daily returns of the CAC index;

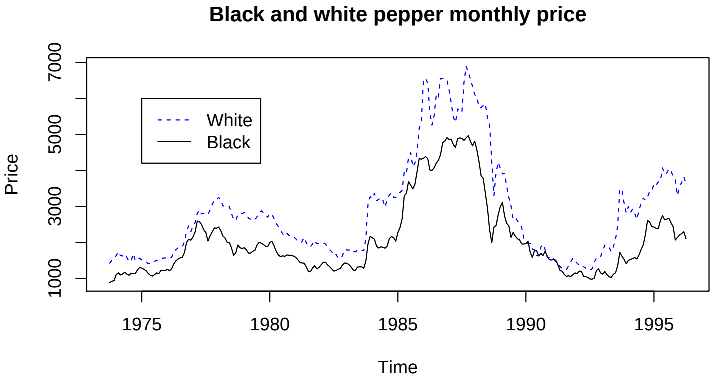
Monthly US leisure and hospitality jobs from January 2000 to September 2019



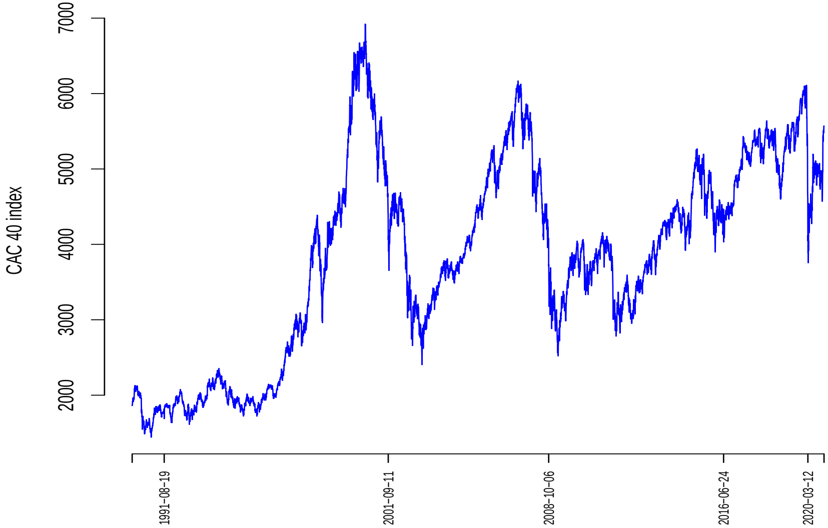
US leisure and hospitality jobs (zoom on the last values)



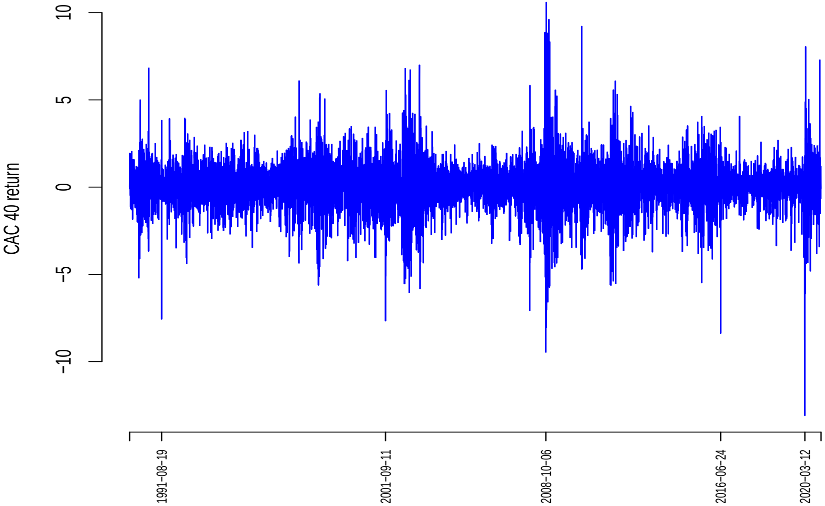
Bivariate series of black and white pepper prices



CAC daily index from 1992-01-03 to 2020-11-26



Log-returns $\epsilon_t = \log(p_t/p_{t-1}) \approx \frac{p_t - p_{t-1}}{p_{t-1}}$



Time series as realization of a stochastic process

In statistics the observations are supposed to be realizations of independent and identically distributed (iid) random variables or vectors. This is not a relevant framework in time series.

Let $X = (X_t)_{t\in\mathscr{T}}$ be a stochastic process in discrete time ($\mathscr{T} = \mathbb{N}^*$ or $\mathscr{T} = \mathbb{Z}$), that is a countable set of randon variables defined on the same probability space $(\Omega, \mathscr{A}, P)$, valued in $\mathbb{R}^d$.

A realization $X(\omega) = (x_t)$ is called a trajectory.

The observed time series $x_1, \ldots, x_t, \ldots, x_n$ (often denoted $X_1, \ldots, X_n$) is a part of a trajectory. This is less than a single observation!

When $d = 1$ we have a univariate time series. When $d > 1$ we have a multivariate time series.

Strict stationarity

A time series (X_t) is said to be strictly stationary if its marginal and joint probability distributions do not change when shifted in time.

Definition

The series (X_t), $X_t \in \mathbb{R}^d$, is strictly stationary if

$(X_1, X_2, \ldots, X_k)$ has the same distribution as $(X_{1+h}, X_{2+h}, \ldots, X_{k+h})$

for any h and any $k \geq 1$.

This concept can be difficult to manipulate

Second-order stationarity

Definition

Let (X_t) such that $E\|X_t\|^2 < \infty$. The mean function of (X_t) is

$$\mu_X(t) = E(X_t)$$

The autocovariance function of (X_t) is

$$\gamma_X(r,s) = \mathrm{Cov}(X_r, X_s)$$

Definition

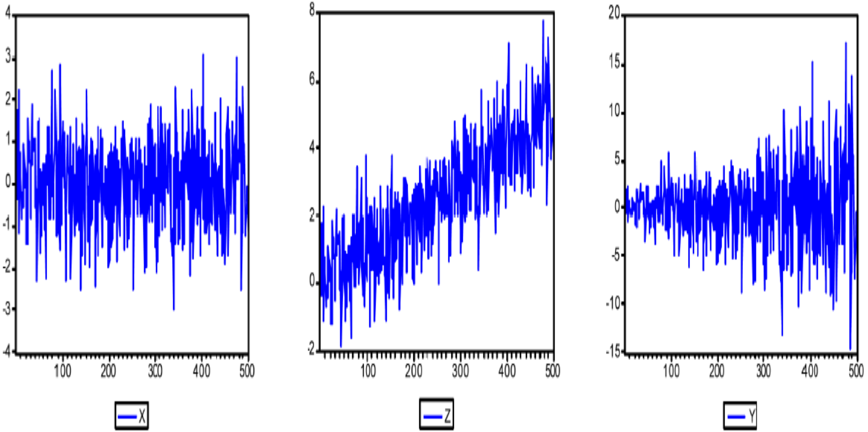
(X_t) is (weakly or second-order) stationary if

(i) $\mu_X(t)$ is independent of t, and

(ii) $\gamma_X(t, t+h)$ is independent of t, for any h.

Illustration on simulated series (see the previous graphs for real examples)

(X_t) is stationary, (Y_t) and (Z_t) are not



Autocovariance and autocorrelation functions

Definition (case $d = 1$)

Let (X_t) a univariate second-order stationary time series. The autocovariance function of (X_t) is

$$\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h}), \quad h = 0, \pm 1, \pm 2, \ldots$$

The autocorrelation function is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Cor}(X_t, X_{t+h}), \quad h = 0, \pm 1, \pm 2, \ldots$$

Remark: even functions

$$\gamma_X(h) = \gamma_X(-h), \quad \rho_X(h) = \rho_X(-h).$$

Interpretation

Autocovariance and autocorrelation functions, case $d \geq 1$

For a multivariate, second-order stationary process, the autocovariance of lag h is denoted

$$\begin{aligned}
\Gamma_X(h) &= \operatorname{Cov}(X_t, X_{t-h}) \\
&= EX_1 X'_{1-h} - EX_1 EX'_1 \\
&= \Gamma'_X(-h) = \left[\gamma_{ij}(h)\right]_{i,j=1,\dots,d},
\end{aligned}$$

and the autocorrelation of lag h is the matrix

$$\begin{aligned}
R_X(h) &= \left[\rho(X_{it}, X_{j,t-h}) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}\right]_{i,j=1,\dots,d} \\
&= \left(\operatorname{diag}\Gamma_X(0)\right)^{-1/2} \Gamma_X(h) \left(\operatorname{diag}\Gamma_X(0)\right)^{-1/2}.
\end{aligned}$$

Links between strict and 2nd-order stationarity

- (X_t) strictly stationary $+E\|X_1\|^2 < \infty \Rightarrow (X_t)$ 2nd-order stationary.
- (X_t) Gaussian* and 2nd-order stationary $\Rightarrow (X_t)$ strictly stationary.

*i.e. any linear combination of the X_t's is Gaussian

White noise

Definition

A (weak) white noise is a sequence (ϵ_t) of uncorrelated variables with zero mean, constant variance:

$$E(\epsilon_t) = 0, \quad \mathrm{Var}(\epsilon_t) = \sigma^2 (\Sigma \text{ if multivariate}), \quad \mathrm{Cov}(\epsilon_t, \epsilon_s) = 0,\ t \neq s$$

Notation: $(\epsilon_t) \sim \ WN\,(0, \sigma^2)$ if $d = 1$,
$(\epsilon_t) \sim \ WN\,(0, \Sigma)$ if $d > 1$.

Autocovariance function: when $d = 1$,

$$\gamma_\epsilon(h) = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \neq 0 \end{cases}, \qquad \text{when } d > 1, \quad \Gamma_\epsilon(h) = \begin{cases} \Sigma, & h = 0 \\ 0, & h \neq 0. \end{cases}$$

Strong or semi-strong white noise

Definition

A strong WN is a sequence (ϵ_t) of independent and identically distributed (iid) variables, with zero mean and finite variance σ^2 (iid $(0,\Sigma)$ in multivariate).

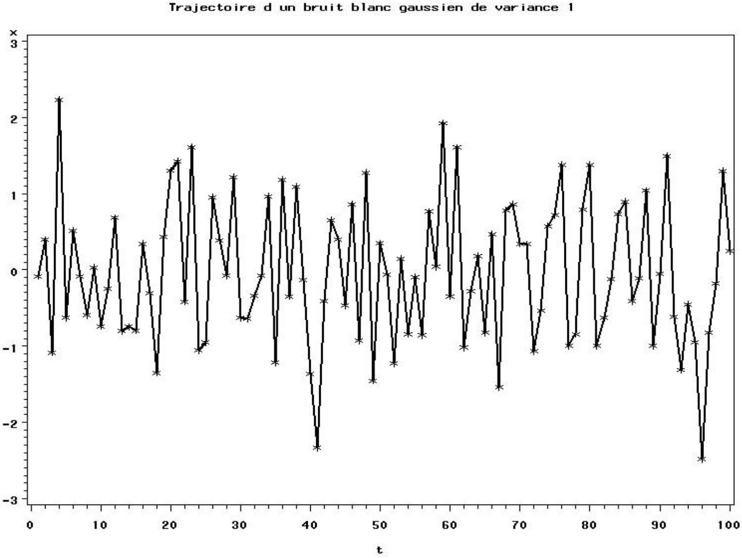
Notation: (ϵ_t) $iid\,(0,\sigma^2)$ or (ϵ_t) $iid\,(0,\Sigma)$

Definition

A semi-strong WN is a strictly and 2nd-order stationary sequence (ϵ_t), such that the optimal prediction of ϵ_t given its past is 0:

$$E(\epsilon_t|\epsilon_u, u<t)=0.$$

Notation: (ϵ_t) $MD\,(0,\sigma^2)$ (MD for martingale difference).


Trajectoire d un bruit blanc gaussien de variance 1
x
3
2
1
0
-1
-2
-3
0
10
20
30
40
50
60
70
80
90
100
t

Links between the different types of WN

$$\text{weak WN} \supset \text{semi-strong} \supset \text{strong} \supset \text{Gaussian WN}$$

Remark: A semi-strong WN (resp. weak WN) cannot be predicted (resp. linearly predicted) from its past.

Most time series models thus have the form

$$X_t = \varphi(X_{t-1}, X_{t-2}, \dots) + \epsilon_t.$$

In this course, the focus is on linear functions φ.

Theoretical predictions of X_t such that $EX_t^2 < \infty$ $(d = 1)$

- Conditional expectation:

$$E(X_t \mid X_u, u < t) = \arg \min_{Z \in \sigma(X_u, u<t)} E(X_t - Z)^2,$$

  i.e. the best approximation of X_t as a function of the past.

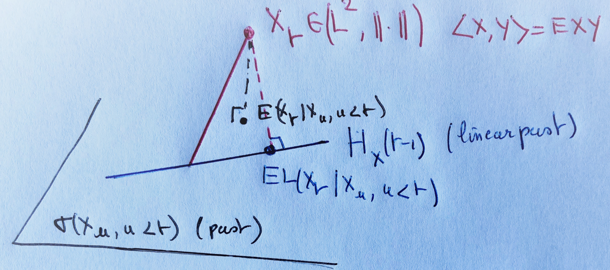
- Linear conditional expectation:

$$EL(X_t \mid X_u, u < t) = \arg \min_{Z \in \mathscr{H}_X(t-1)} E(X_t - Z)^2$$

  i.e. the best approximation of X_t as a linear function of the past.
  ($\mathscr{H}_X(t-1)$ Hilbert space generated by the linear combinations of $1, X_{t-1}, X_{t-2}, \dots$).

▸ Projection and conditional expectation

Conditional and linear conditional expectations



- $E(X_t|X_u, u<t) \in \sigma(X_u, u<t)$ and $X_t - E(X_t|X_u, u<t) \perp \sigma(X_u, u<t)$
- $EL(X_t|X_u, u<t) \in H_X(t-1)$ and $X_t - EL(X_t|X_u, u<t) \perp H_X(t-1)$
- $\|X_t - E(X_t|X_u, u<t)\| \le \|X_t - EL(X_t|X_u, u<t)\|$

Innovations and white noises

For (X_t) stationary (strict and 2nd-order), 2 concepts:

- Strong innovation:

$$\epsilon_t = X_t - E(X_t \mid X_u, u < t),$$

semi-strong WN "orthogonal" to any function of the past of X_t

$$E(\epsilon_t Z_{t-1}) = 0, \quad \forall Z_{t-1} \in \sigma(X_u, u < t).$$

- Linear innovation:

$$\epsilon_t = X_t - EL(X_t \mid X_{t-1}, X_{t-2}, \ldots),$$

weak WN "orthogonal" to any linear function of the past of X_t

$$E(\epsilon_t Z_{t-1}) = 0, \quad \forall Z_{t-1} \in \mathcal{H}_X(t-1).$$

▸ Proof

Moving Average of order 1: MA(1)

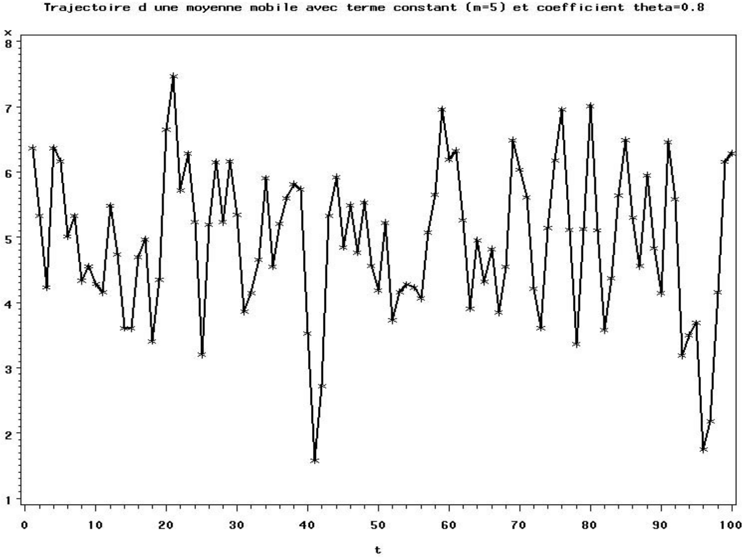
$$X_t = m + \epsilon_t + \theta\epsilon_{t-1}, \quad (\epsilon_t) \sim WN(0, \sigma^2), \quad m, \theta \in \mathbb{R}.$$

We have $E(X_t) = m$, $EX_t^2 = \sigma^2(1+\theta^2) < \infty$ and

$$\gamma_X(h) = \begin{cases} \sigma^2(1+\theta^2), & h = 0 \\ \sigma^2\theta, & h = \pm 1 \\ 0, & |h| > 1. \end{cases}$$

Thus (X_t) is second-order stationary.
Autocorrelation function:

$$\rho_X(h) = \begin{cases} 1, & h = 0 \\ \theta/(1+\theta^2), & h = \pm 1 \\ 0, & |h| > 1. \end{cases}$$


Trajectoire d une moyenne mobile avec terme constant (m=5) et coefficient theta=0.8
x
8
7
6
5
4
3
2
1
0
10
20
30
40
50
60
70
80
90
100
t

First-order autoregressive: AR(1)

Let $(\epsilon_t) \sim WN(0, \sigma^2)$ and $\phi \in \mathbb{R}$.

Is there a (stationary) process (X_t) satisfying the autoregressive equation

$$X_t = \phi X_{t-1} + \epsilon_t?$$

Starting from an initial value X_0, it is easy to construct a process - generally non stationary - satisfying the AR(1) equation for all $t \geq 1$.

Conditions are needed to ensure the existence of a stationary solution for all $t \in \mathbb{Z}$.

Causal stationary solution of $X_t = \phi X_{t-1} + \epsilon_t$

If $|\phi| < 1$ then
$X_t(N) := \sum_{i=0}^{N} \phi^i \epsilon_{t-i} = \epsilon_t + \phi(\epsilon_{t-1} + \cdots + \phi^{N-1}\epsilon_{t-1-(N-1)}$ satisfies

$$X_t(N) = \phi X_{t-1}(N-1) + \epsilon_t.$$

Moreover $X_t = \lim_{N\to\infty} X_t(N)$ exists with probability 1 because

$$E\sum_{i=0}^{\infty} |\phi|^i |\epsilon_{t-i}| = E|\epsilon_t| \frac{1}{1-|\phi|} < \infty.$$

Hence, with probability 1, the series

$$X_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$$

converges absolutely, and it is a solution to the AR(1) model.
We will see later that this solution is 2nd-order stationary (and also strictly if the WN is strong).

Stationary anticipative solution of $X_t = \phi X_{t-1} + \epsilon_t$

If $|\phi| > 1$ then

$$X_t = -\frac{1}{\phi}\epsilon_{t+1} + \frac{1}{\phi}X_{t+1} = -\sum_{i=1}^{\infty}\frac{1}{\phi^i}\epsilon_{t+i}$$

is called the anticipative (or non causal) solution of the AR(1) model. We will see later that this solution is 2nd-order stationary (and also strictly if the WN is strong).

No stationary solution

If $|\phi| = 1$ and $\sigma^2 > 0$ then, there exists no stationary solution.
Indeed, if (X_t) were a stationary solution,

$$\mathrm{Var}(X_t - \phi^N X_{t-N}) = 2\,\{\mathrm{Var}(X_t) \pm \mathrm{Cov}(X_t, X_{t-N})\}$$

would be bounded. But since $X_t - \phi^N X_{t-N} = \sum_{i=0}^{N-1} \phi^i \epsilon_{t-i}$, we have

$$\mathrm{Var}(X_t - \phi^N X_{t-N}) = \sum_{i=0}^{N-1} \phi^{2i} \mathrm{Var}\, \epsilon_{t-i} = N\sigma^2 \to \infty,$$

which leads to a contradiction.

Autocorrelations of the causal AR(1)

Let (X_t) be the solution of

$$X_t = \phi X_{t-1} + \epsilon_t, \quad (\epsilon_t) \sim WN(0, \sigma^2), \; |\phi| < 1$$

where ϵ_t is non correlated with the X_{t-i}, $i > 0$. We have $E(X_t) = 0$ and

$$\gamma_X(h) \quad = \quad \phi \gamma_X(h-1) + \mathrm{Cov}(\epsilon_t, X_{t-h}).$$

Thus

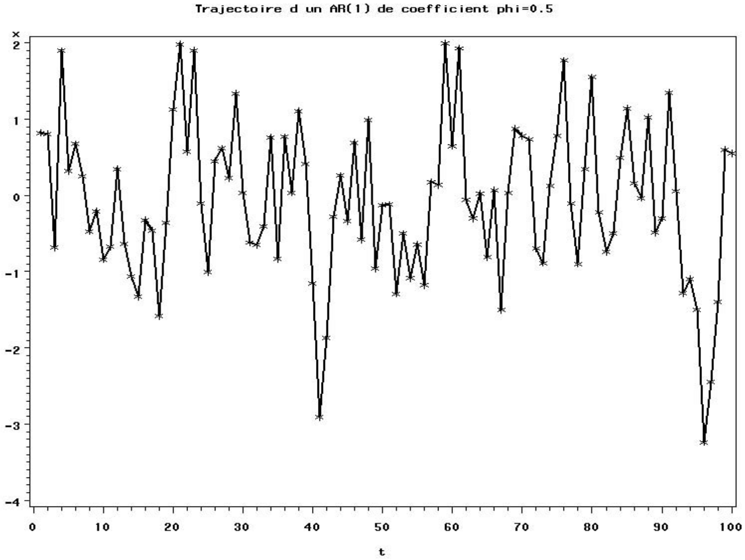
$$\gamma_X(h) \quad = \quad \phi \gamma_X(h-1) = \phi^h \gamma_X(0), \quad h > 0$$

and, using $\gamma_X(h) = \gamma_X(-h)$,

$$\gamma_X(0) \quad = \quad \phi \gamma_X(-1) + \sigma^2 = \phi^2 \gamma_X(0) + \sigma^2 = \frac{\sigma^2}{1-\phi^2}$$

Autocorrelation function:

$$\rho_X(h) = \phi^{|h|}.$$


Trajectoire d un AR(1) de coefficient phi=0.5
x
2
1
0
-1
-2
-3
-4
0
10
20
30
40
50
60
70
80
90
100
t

ARMA models

MA(1) and AR(1) are particular ARMA(p,q) models:

$$\begin{cases} X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \epsilon_t - \psi_1 \epsilon_{t-1} - \cdots - \psi_q \epsilon_{t-q} \\ (\epsilon_t) \text{ weak white noise} \end{cases}$$

which will be studied in the next chapter.

Models with trends

Many time series sample paths display a trend, which can be modelled by

$$X_t = m_t + Y_t$$

where m_t is a non constant function of time, called trend, and (Y_t) is a centred time series.
We have $E(X_t) = m_t$ thus (X_t) is not stationary.
Examples:
- linear trend: $m_t = a_0 + a_1 t,\ a_1 \neq 0$
- quadratic trend: $m_t = a_0 + a_1 t + a_2 t^2,\ a_2 \neq 0$

The coefficients a_0, a_1, a_2 can be estimated by least-squares:

$$\min_{a_0,a_1,a_2} \sum_{t=1}^{n} (x_t - m_t)^2 \quad \Longrightarrow \quad \hat{a}_0, \hat{a}_1, \hat{a}_2$$

Models with tend and seasonality

Many series also display seasonal features, generally linked to the seasons or the economic activity (Ex: decrease in consumption of certain goods in August etc)..
It is natural to complete the model with trend as

$$X_t = m_t + s_t + Y_t$$

where s_t is a periodic function of time, called seasonality:

$$s_1, \ldots, s_d \text{ and } s_t = s_{t-d} \text{ for } t > d.$$

Random walk

Let $(\epsilon_t) \sim WN(0, \sigma^2)$. The random walk is defined by

$$X_0 = 0, \quad X_t = X_{t-1} + \epsilon_t, \quad t > 0.$$

Thus

$$X_t = \epsilon_t + \cdots + \epsilon_1, \quad t > 0.$$

Thus $EX_t = 0$ but

$$EX_t^2 = E\epsilon_t^2 + \cdots + E\epsilon_1^2 = t\sigma^2.$$

Therefore the random walk is not stationary.

Random walk with trend

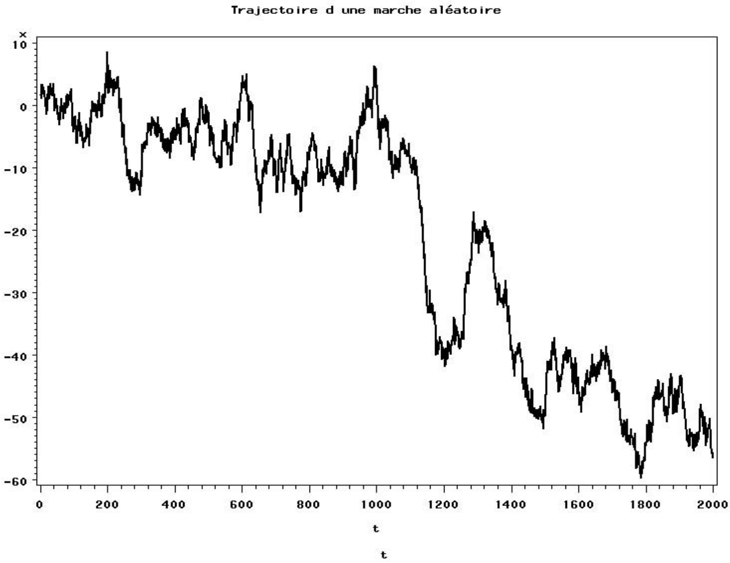
Let $(\epsilon_t) \sim WN(0, \sigma^2)$. The random walk with trend is defined by

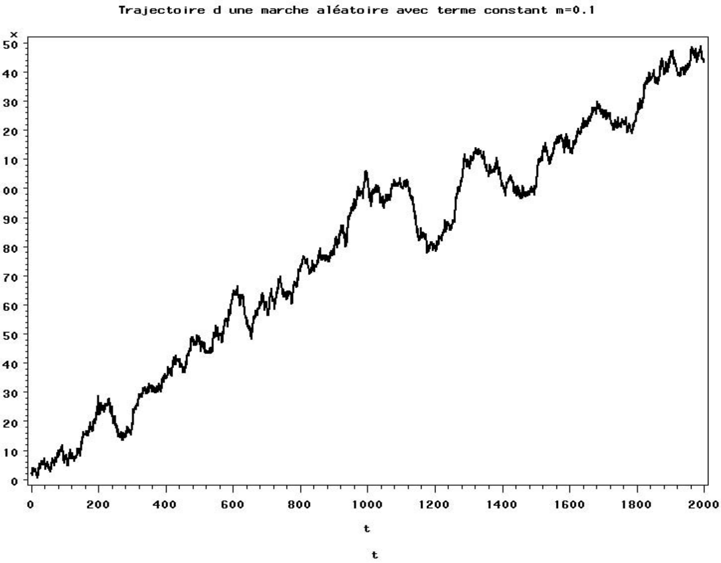
$$X_0 = a, \quad X_t = X_{t-1} + b + \epsilon_t, \quad t > 0.$$

Hence

$$X_t = \epsilon_t + \cdots + \epsilon_1 + bt + a, \quad t > 0.$$

The component $a + bt$ is the (linear) deterministic trend and $\epsilon_1 + \cdots + \epsilon_t$ is a stochastic trend.


Trajectoire d une marche aléatoire
x
10
0
-10
-20
-30
-40
-50
-60
0
200
400
600
800
1000
1200
1400
1600
1800
2000
t
t


Trajectoire d une marche aléatoire avec terme constant m=0.1
x
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000
t
t

The Box and Jenkins (1976) approach

Non stationary time series can **often** be made stationary by applying repetitively the difference operator

$$\begin{aligned} \nabla X_t &= X_t - X_{t-1}, \\ \nabla^2 X_t &= \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2}, \qquad \text{etc.} \end{aligned}$$

or the seasonal difference operator

$$\nabla_s X_t = X_t - X_{t-s}.$$

Examples:

$$\begin{aligned} X_t &= a_0 + a_1 t + Y_t \quad \Longrightarrow \nabla X_t = a_1 + Y_t - Y_{t-1} \\ X_t &= a_0 + a_1 t + a_2 t^2 + Y_t \quad \Longrightarrow \nabla^2 X_t = 2a_2 + \nabla^2 Y_t \\ X_t &= a_0 + a_1 t + s_t + Y_t \quad \Longrightarrow \nabla_s X_t = a_1 s + Y_t - Y_{t-s}. \end{aligned}$$

Vector Autoregressive (VAR)

Let the bivariate VAR(1) $X_t = (X_{1t}, X_{2t})'$ satisfying

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} X_{t-1} + \epsilon_t,$$

where ϵ_t is a bivariate WN. If $|a| < 1$ and $|c| < 1$ then

$$\begin{aligned} X_t &= \sum_{i\geq 0} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}^i \epsilon_{t-i} = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} + \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \begin{pmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{pmatrix} \\ &\quad + \begin{pmatrix} a^2 & ab + bc \\ 0 & c^2 \end{pmatrix} \begin{pmatrix} \epsilon_{1t-2} \\ \epsilon_{2t-2} \end{pmatrix} + \cdots \end{aligned}$$

This MA(∞) is useful to analyze the effects of shocks, or impulse responses: an impulse of +1 on ϵ_{1t-i} induces a variation of a^i on X_{1t} and has no effect on X_{2t}.

Even if we don't care, the time series structure matters

A spurious regression provides misleading statistical evidence of a relationship between non-stationary variables.

Example of spurious regression

US Export Index (Y), 1960-1990, annual data, on Australian males' life expectancy (X)

$$\widehat{Y} = \underset{(-16.70)}{-2943.} + \underset{(17.76)}{45.7974}X, \text{with } R^2 = .916.$$

Explanation: the observations (Y_t, X_t) $t = 1, \ldots, n$ are not iid, even not stationary. What do we estimate ? What is the limit of

$$\hat{\beta} = \frac{\frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X}_n)(Y_t - \overline{Y}_n)}{\frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X}_n)^2}?$$

(see for instance this paper for a review)

Estimation of moments

The theoretical moments can be estimated from the observations $X_1, \ldots, X_n$. A natural estimator for the expectation EX_1 is the empirical mean

$$\overline{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t.$$

As the observations are not iid, the usual Law of Large Numbers (LLN) does not apply. Instead, we will rely on the concept of ergodicity.

A stationary sequence is called ergodic if it satisfies the strong LLN.

Two types of means

A stationary sequence (X_t) defined on $(\Omega, \mathscr{A}, P)$ is ergodic if the mean of its values over a trajectory is a.s. equal to the mean value of the variable X_t over all trajectories.

Ergodicity

Let (X_t) strictly stationary, valued in $\mathbb{R}$ and such that $E|X_1| < \infty$. It is called ergodic if there exists Ω_0 such that $P(\Omega_0) = 1$ and

$$\forall \omega_0 \in \Omega_0, \quad \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} X_t(\omega_0) = \int_\Omega X_1(\omega)\, dP(\omega).$$

Ergodic stationary process

A stationary sequence is ergodic if it satisfies the strong LLN (even if the usual conditions are not satisfyed).

Définition

A strictly stationary process $(Z_t)_{t\in\mathbb{Z}}$, valued in $\mathbb{R}^d$, is called ergodic if for any integer k, and any Borel set B of $\mathbb{R}^{dk}$,

$$n^{-1}\sum_{t=1}^{n} I_B(Z_t, Z_{t+1}, \ldots, Z_{t+k-1}) \to P\{(Z_1, \ldots, Z_{1+k}) \in B\}\ a.s.$$

The concept of ergodicity is much more general. It can be extended to non stationary sequences (see *e.g.* Billingsley (1995) "Probability and Measure", Wiley, New York.)

Ergodic Theorem

Any fixed transformation of a stationary ergodic sequence is also stationary and ergodic.

Theorem

If $(Z_t)_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic sequence and if $(Y_t)_{t\in\mathbb{Z}}$ is defined by

$$Y_t = f(Z_t, Z_{t-1}, \ldots),$$

or even more generally by

$$Y_t = f(Z_t, Z_{t-1}, \ldots; Z_{t+1}, Z_{t+2}, \ldots),$$

where f is a measurable function, from $\mathbb{R}^\infty$ to $\mathbb{R}^d$, then $(Y_t)_{t\in\mathbb{Z}}$ a strictly stationary and ergodic sequence.

Examples and counter-examples of ergodic processes

- A strong WN (ϵ_t) is stationary and ergodic.
- A MA(q)

$$X_t = \sum_{i=0}^{q} c_i \epsilon_{t-i}$$

  or the *causal* solution of an AR(1)

$$X_t = aX_{t-1} + \epsilon_t, \qquad |a| < 1,$$

  are stationary and ergodic.
- The process defined by

$$X_t = X, \quad \forall t,$$

  where X is a non-degenerate r.v. is stationary but is not ergodic.

Weak ergodicity for the mean

If (X_t) is a univariate, 2nd-order stationary process, whose autocovariance function satisfies $\gamma(h) \to 0$ when $h \to \infty$, then

$$\overline{X}_n := \frac{1}{n} \sum_{t=1}^{n} X_t \xrightarrow{L^2} EX_1 \qquad \text{as } n \to \infty.$$

Proof: Using the Cesàro lemma

$$E\left(\overline{X}_n - EX_1\right)^2 = \frac{1}{n^2} \sum_{t,s=1}^{n} \mathrm{Cov}(X_t, X_s)$$

$$= \quad \frac{1}{n} \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) \gamma_X(h) \leq \frac{1}{n} \sum_{|h|<n} \left|\gamma(h)\right| \to 0.$$

Empirical autocorrelations

Definition (univariate case)

The empirical autocovariance function is, for $|h| < n$,

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=|h|+1}^{n} (X_t - \overline{X}_n)(X_{t-|h|} - \overline{X}_n), \quad \overline{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t.$$

The empirical autocorrelation function is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad |h| < n.$$

In the multivariate case, the estimator of $\Gamma(h) = \mathrm{Cov}(X_t, X_{t-h})$ is

$$\hat{\Gamma}(h) = \frac{1}{n} \sum_{t=h+1}^{n} (X_t - \overline{X}_n)(X_{t-h} - \overline{X}_n)', \quad 0 \leq h \leq n-1.$$

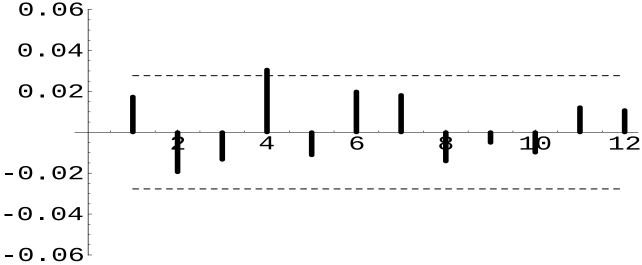
Empirical autocorrelations of a WN

If the observations are the realizations of a strong WN,

- $\overline{X}_n \to 0$, $\hat{\gamma}(0) \to \sigma^2$ (LGN) and, if $h \neq 0$, $\hat{\gamma}(h) \to 0$ p.s. .
- $\sqrt{n}\overline{X}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$,
  for $h \neq 0$, $\sqrt{n}\hat{\gamma}(h) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^4)$
  for $h \neq 0$, $\sqrt{n}\hat{\rho}(h) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ ("almost" usual CLT + Slutsky).

Empirical autocorrelogram of a WN

For a strong WN, $\hat{\rho}(h)$ belongs to the interval $\pm 1.96/\sqrt{n}$ in dotted lines with a probability ≃95%.



Empirical autocorrelations of a strong WN, for $n = 5000$.

Empirical autocorrelations of a stationary process

For a stationary (strict and 2nd-order) and ergodic process,

$$\overline{X}_n \to EX_1, \qquad \hat{\gamma}(0) \to \mathrm{V}(X_1)$$

and, if $h \neq 0$,

$$\hat{\gamma}(h) \to \gamma(h) \text{ a.s. (ergodic theorem).}$$

For ARMA-type processes (see chapter 7 in Brockwell Davis)

$$\sqrt{n}\{\hat{\rho}(h) - \rho(h)\} \xrightarrow{\mathscr{L}} \mathscr{N}(0, \sigma_h^2)$$

where σ_h^2 only depends on $\rho(\cdot)$ (Bartlett's formula).

End of Chapter 1 ☺ !

Interpretation of a correlation coefficient

Let X and Y be real random variable with non zero variances.

- $|\rho(X,Y)| \le 1$ by Cauchy-Schwarz's inequality $E|XY| \le \sqrt{EX^2 EY^2}$; Proof
- If X and Y are independent then $\rho(X,Y) = 0$, but but the converse is wrong; Example
- We have

$$\rho = \pm 1 \quad \text{iff} \quad Y = aX + b, \quad \text{sign}(a) = \text{sign}(\rho).$$

Proof

Return

Proof of Cauchy-Schwarz's inequality

For all a we have

$$E(|Y|-a|X|)^2 = EY^2 + a^2 EX^2 - 2aE|XY| \geq 0.$$

Taking $a = E|XY|/EX^2$, we obtain

$$EY^2 - \frac{(E|XY|)^2}{EX^2} \geq 0.$$

Return

Uncorrelatedness does not entail independence

Let (X, Y) such that

$X \backslash Y$	-1	0	1
-1	1/4	0	1/4
1	0	1/2	0

We have $\rho(X, Y) = 0$, but X and Y are not independent.

Return

Proof that $\rho = \pm 1$ iff $Y = aX + b$

If $Y = aX + b$ with $a \neq 0$, then $\mathrm{Var}\, Y = a^2 \mathrm{Var} X$ and $\mathrm{Cov}(X, Y) = a \mathrm{Var} X$, thus we have $\rho = a/|a|$.

Conversely, if $\rho^2 = 1$, then

$$\begin{aligned} & E\left\{Y - EY - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}X}(X - E(X)\right\}^2 \\ = \quad & \mathrm{Var}\, Y + \frac{\mathrm{Cov}^2(X,Y)}{\mathrm{Var}X} - 2\frac{\mathrm{Cov}^2(X,Y)}{\mathrm{Var}X} = 0. \end{aligned}$$

Return

Projection and conditional expectation

The space $L^2 = L^2(\Omega, \mathscr{A}, P)$ of the square integrable random variables is a Hilbert space associated with the inner product

$$< X, Y >= EXY.^{\dagger}$$

For $(X_t) \in L^2$, let $\sigma(X_u, u < t)$ the past of X_t and $\mathscr{H}_X(t-1)$ the linear past of X_t, that is the closed subspaces of L^2 generated by all the square integrable measurable functions of $\{X_u, u < t\}$ and the linear combinations of $1, X_{t-1}, X_{t-2}, \ldots$, respectively.
The projection of $X \in L^2$ on a closed subspace $\mathscr{M}$ of the Hilbert space L^2 is the conditional expectation $E(X \mid \mathscr{M})$, characterized by

$$1) E(X \mid \mathscr{M}) \in \mathscr{M}, \quad 2) X - E(X \mid \mathscr{M}) \perp Y, \quad \forall Y \in \mathscr{M}.$$

Return

† assuming that 2 random variables are equal when they are almost surely equal (so that $\|X\| := \sqrt{< X, X >} = 0$ iff $X = 0$).

Proof that the innovations are white noises

The strict stationarity of $E(X_t \mid X_u, u < t) = \varphi(X_u, u < t)$ and $\epsilon_t = X_t - E(X_t \mid X_u, u < t)$ is a consequence of the ergodic theorem given below. Note that $E(X_t \mid X_u, u < t)$ belongs to L^2, and thus is also second order stationary. We have
$E\epsilon_t = E\{E(\epsilon_t \mid X_u, u < t)\} = E\{E(X_t \mid X_u, u < t) - E(X_t \mid X_u, u < t)\} = 0.$
Note that $\sigma(\epsilon_u, u < t) \subset \sigma(X_u, u < t)$. By the tower property, we thus have

$$E(\epsilon_t \mid \epsilon_u, u < t) = E\{E(\epsilon_t \mid X_u, u < t) \mid \epsilon_u, u < t\} = 0.$$

We have shown that the strong innovation is a semi-strong white noise. The other result is obtained similarly. return

Représentation causale faible

Soit $|\phi| > \mathbf{1}$ et

$$X_t = -\sum_{i=1}^{\infty} \frac{1}{\phi^i} \epsilon_{t+i}$$

la solution non causale de $X_t = \phi X_{t-1} + \epsilon_t$. Posons $\epsilon_t^* = X_t - \frac{1}{\phi} X_{t-1}$. Puisque $\gamma_X(h+1) = \phi \gamma_X(h)$ pour $h \geq 0$, on a

$$\begin{aligned} \mathrm{cov}(\epsilon_t^*, \epsilon_{t-h}^*) &= \mathrm{cov}\left(X_t - \frac{1}{\phi} X_{t-1}, X_{t-h} - \frac{1}{\phi} X_{t-h-1} \right) \\ &= \gamma_X(h) - \frac{1}{\phi} \gamma_X(h-1) - \frac{1}{\phi} \gamma_X(h+1) + \frac{1}{\phi^2} \gamma_X(h) = 0 \end{aligned}$$

pour $h \geq 1$, donc ϵ_t^* est un bruit blanc (faible), ce qui montre que l'hypothèse $|\phi| \leq 1$ n'est pas très restrictive pour l'AR(1) faible.

Retour