Introduction to Financial Econometrics

Chapter 5: Introduction to Time Series Models

Christophe Hurlin

Univ Orléans

May, 2019

1. Introduction

Definition (time series)

A **time series** is a set of observations $x_1, \ldots, x_T$, each one being recorded at a specific time t.

Notes

- The time elapsed between two observations is assumed to be constant (e.g., daily data, weekly data, annual data).
- Sampling frequency matters (cf. Chapter 2).

1. Introduction

Definition (data generating process)

The **data generating process (DGP)** underlying the realizations $\{x_t\}$ is a (real-valued) discrete time stochastic process, denoted $\{X_t\}$.

Notes

1. The DGP is the "true" model that has generated the dataset $x_1, \ldots, x_T$.
2. In reality we can only observe the time series at a **finite number of times**, and the sequence of random variables $(X_1, , \ldots, X_T)$ is a T-dimensional random vector.
3. However, it is convenient to allow the number of observations to be **infinite**. In that case $\{X_t, t \in \mathbb{Z}\}$ is called a discrete time **stochastic process**.

Remarks

Consider a stochastic process $\{X_t, t \in T\}$, where T is called the index set.

Example

Some examples of index sets are $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, $\mathbb{N} = \{0, 1, 2, \ldots\}$,etc.

- The stochastic process $\{X_t, t \in T\}$ is often called a **time series**.

time series = the set of observations $x_1, \ldots, x_T$

time series = the stochastic process $\{X_t, t \in \mathbb{Z}\}$

Definition (time series model)

A **time series model** for the observed data $\{x_t\}$ is a specification of the joint distribution (or possibly only the means and covariances) of a sequence of random variables $\{X_t, t \in \mathbb{Z}\}$ of which $\{x_t, t \in \mathbb{Z}\}$ is postulated to be a realization.

Notes

- The general idea of time series econometrics consists in specifying a time series model which is close as possible to the DGP.
- The time series model is likely to be different from the DGP: there is a **model risk**.

1. Introduction

Example (time series model)

An example of model is the **autoregressive (AR)** process of order p, defined as

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \varepsilon_t$$

where ε_t is an innovation process. This model specifies the conditional mean of $\{X_t\}$ with

$$\mathbb{E}\left(X_t | \underline{X_{t-1}}\right) = \alpha_0 + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p}$$

where $\underline{X_{t-1}} = \{X_{t-1}, X_{t-2}, \ldots\}$ denotes the past values of the process $\{X_t\}$

1. Introduction

How to specify a "good" time series model?

1. Study some **statistical properties** of the observed data $\{x_t\}$, for instance the **stationarity**, the patterns of the autocorrelation function (**ACF**) or the partial autocorrelation function (**PACF**), etc.
2. Compare these properties to the "theoretical" properties of some **typical time series models**, e.g. AR, MA, ARIMA, SARIMA, ARFIMA, GARCH, etc.
3. Choose the most appropriate model and estimate its parameters (generally by ML).
4. Use this model for forecasting.

1. Introduction

The outline of this chapter is the following:

- **Section 2:** Stationarity
- **Section 3:** Wold decomposition and prediction
- **Section 4:** The Box-Jenkins modeling approach
- **Section 5:** Univariate time series models

1. Introduction

References (theoretical)

- Davidson, J. (2000), Econometric Theory, Blackwell Publishers.
- Greene W. (2007), Econometric Analysis, sixth edition, Pearson - Prentice Hill.
- Hamilton, James D. (1994), Time Series Analysis, New Jersey: Princeton University Press (**main reference**).
- Lütkepohl, H. (2005), New Introduction to Multiple Time Series Analysis, Springer.

References (applied)

- Cryer, J.D. and Chan, K.-S. (2008), Time series Analysis with applications in R, Springer.
- Enders, W. (2003), Applied Econometric Time Series, Wiley.
- Shumway, D.H. and Stoffer, D.S. (2006), Time Series Analysis and its applications with R Examples, Springer.

Section 2

Stationarity

2. Stationarity

Objectives

1. To define the strict **stationarity**
2. To define the weak (second-order) **stationarity**
3. To define the concept of **strict white noise** or **IID noise**
4. To define the concept of (uncorrelated) **white noise**
5. To define the concept of **martingale difference**

Stationarity

Loosely speaking, a stochastic process is **stationary**, if its statistical properties do not change with time.

There exist two definitions of the stationarity:

1. **The strict stationarity**
2. **The weak or second order stationarity**

Stationarity

Loosely speaking, a stochastic process is **stationary**, if its statistical properties do not change with time.

There exist two definitions of the stationarity:

1. **The strict stationarity**

2. **The weak or second order stationarity**

2. Stationarity

Let $\{X_t, \ t \in \mathbb{Z}\}$ be a stochastic process and let $F_X\left(x_{t_1+\tau}, \ldots, x_{t_k+\tau}\right)$ represent the cdf of the unconditional joint distribution of $\{X_t\}$ at times $t_1 + \tau, \ldots, t_k + \tau$.

Definition (strict stationarity)

The process $\{X_t, \ t \in \mathbb{Z}\}$ is said to be **strictly stationary** if, for all k and τ, and for all $t_1, \ldots, t_k$,

$$F_X\left(x_{t_1+\tau}, \ldots, x_{t_k+\tau}\right) = F_X\left(x_{t_1}, \ldots, x_{t_k}\right)$$

Interpretation: The unconditional joint probability distribution does not change when shifted in time.

Stationarity

Loosely speaking, a stochastic process is **stationary**, if its statistical properties do not change with time.

There exist two definitions of the stationarity:

1. **The strict stationarity**
2. **The weak or second order stationarity**

2. Stationarity

Definition (weak or second-order stationarity)

The time series $\{X_t, \ t \in \mathbb{Z}\}$ is said to be **(weakly) stationary** if:

- $\forall t \in \mathbb{Z}, \ \mathbb{E}\left(X_t^2\right) < \infty$
- $\forall t \in \mathbb{Z}, \mathbb{E}\left(X_t\right) = \mu$
- $\forall \left(t, h\right) \in \mathbb{Z}^2, \mathbb{C}ov\left(X_t, X_{t-h}\right) = \gamma\left(h\right)$, does not depend on t.

2. Stationarity

Remarks

1. By default, we consider the second-order or weakly stationarity, i.e. we assume that the two first moments of $\{X_t,\ t \in \mathbb{Z}\}$ are constant over time.

$$\mathbb{E}(X_t) = \mu \quad \mathbb{C}ov(X_t, X_{t-h}) = \gamma(h) \qquad \forall t \in \mathbb{Z}$$

2. The condition $\mathbb{C}ov(X_t, X_{t-h}) = \gamma(h)$ implies that the variance of $\{X_t,\ t \in \mathbb{Z}\}$ is constant over time

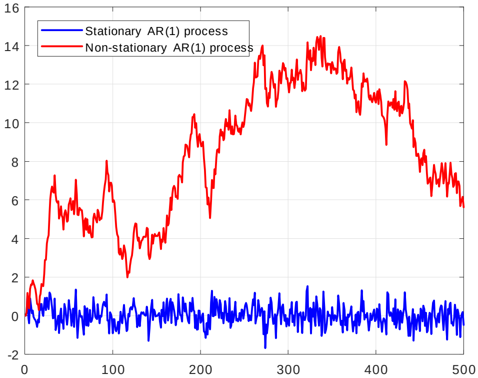
$$\mathbb{V}(X_t) = \mathbb{C}ov(X_t, X_t) = \gamma(0) \qquad \forall t \in \mathbb{Z}$$

3. The condition on $\mathbb{C}ov(X_t, X_{t-h})$ can be interpreted as the "*covariance does not change when shifted in time*".

$$\mathbb{C}ov(X_r, X_s) = \mathbb{C}ov(X_{r+t}, X_{s+t}) \quad \forall (t, r, s) \in \mathbb{Z}^3$$

2. Stationarity

Figure: Simulation of stationary and non-stationary AR(1) processes



2. Stationarity

Stylized Fact 1: Stationarity (reminder Chapter 1)

Fact (stationarity)

In general, the prices are **non-stationary** *whereas the returns are* **stationary**.

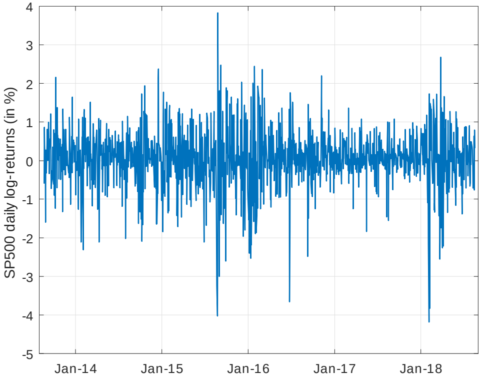
- The prices of an asset recorded over times are often not stationary due to the increase of productivity, the financial crisis, etc.
- However the returns, typically fluctuates around a constant level, suggesting a constant mean over time.

2. Stationarity

Figure: Daily closing prices for the S&P500 index are non stationary

2. Stationarity

Figure: Daily returns for the S&P500 index are stationary



2. Stationarity

Two particular stationary processes are the:

1. **The white-noise processes**
2. **The martingale difference**

Two particular stationary processes are the:

1. **The white-noise processes**
2. **The martingale difference**

2. Stationarity

Definition (strict white noise)

A process $\{\varepsilon_t, t \in \mathbb{Z}\}$ is said to be a **strict white noise** or **IID noise**, written

$$\varepsilon_t \sim \text{i.i.d.} \left(0, \sigma^2\right) \quad \text{or} \quad \varepsilon_t \sim \text{IID} \left(0, \sigma^2\right)$$

if the random variables ε_t are independent and identically distributed with $\mathbb{E}\left(\varepsilon_t\right) = 0$ and $\mathbb{V}\left(\varepsilon_t\right) = \sigma^2$, $\forall t \in \mathbb{Z}$.

Note: In signal processing, white noise is a random signal having equal intensity at different frequencies. White noise draws its name from white light.

Remarks

1. A strict white noise contains no trend or seasonal components and that there is **no dependence** (linear or nonlinear) between observations.

$$\varepsilon_t \sim \text{i.i.d.} \left(0, \sigma^2\right) \implies \varepsilon_t \text{ is independent from } \varepsilon_{t-s} \quad \forall s \in \mathbb{Z}$$

2. Sequence $\{\varepsilon_t\}$ is called a **purely random process**, **IID noise** or simply **strict white noise**.

Definition (white noise)

A process $\{\varepsilon_t, t \in \mathbb{Z}\}$ is said to be an (uncorrelated) **white noise**, written

$$\varepsilon_t \sim \text{WN}\left(0, \sigma^2\right)$$

if the random variables ε_t and ε_s are uncorrelated for $t \neq s$, with $\mathbb{E}\left(\varepsilon_t\right) = 0$ and $\mathbb{V}\left(\varepsilon_t\right) = \sigma^2$, $\forall t \in \mathbb{Z}$.

Note: By definition

$$\varepsilon_t \sim \text{IID}\left(0, \sigma^2\right) \implies \varepsilon_t \sim \text{WN}\left(0, \sigma^2\right)$$

but the reverse is not true.

2. Stationarity

Definition (Gaussian white noise)

A process $\{\varepsilon_t, t \in \mathbb{Z}\}$ is said to be a Gaussian **white noise**, written

$$\varepsilon_t \sim \text{i.i.d.}\ \mathcal{N}\left(0, \sigma^2\right) \quad \text{or} \quad \varepsilon_t \sim \text{IID}\ \mathcal{N}\left(0, \sigma^2\right) \quad \text{or} \quad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

if the random variables ε_t have a normal distribution with $\mathbb{E}\left(\varepsilon_t\right) = 0$ and $\mathbb{V}\left(\varepsilon_t\right) = \sigma^2$, $\forall t \in \mathbb{Z}$.

Note: For a normal distribution, the zero correlation implies independence so that Gaussian white noise is also a strict white noise.

White-noise and stationarity

By definition, a white noise (strict or weak) is a **stationary process** since

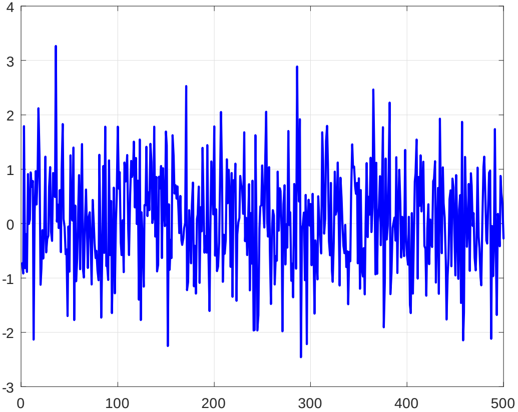
$$\mathbb{E}\left(\varepsilon_t^2\right) = \sigma^2 < \infty$$

$$\mathbb{E}\left(\varepsilon_t\right) = 0, \forall t \in \mathbb{Z}$$

$$\mathbb{C}ov\left(\varepsilon_t, \varepsilon_{t-h}\right) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{otherwise} \end{cases}, \text{ does not depend on } t$$

2. Stationarity

Figure: Simulation of a Gaussian white noise with $\sigma^2 = 1$.



Two particular stationary processes are the:

1. **The white-noise processes**
2. **The martingale difference**

2. Stationarity

Definition (martingale)

A process $\{X_t, t \in \mathbb{Z}\}$ is called a **martingale** if

$$\mathbb{E}\left(X_{t+1} | \underline{X_t}\right) = X_t$$

where $\underline{X_t} = \{X_t, X_{t-1}, \ldots\}$ is the information set available to time t including X_t.

Note: If X_t represents an asset's price at date t, then the martingale process implies that tomorrow's price is expected to be equal to today's price, given the information set containing price history of the asset.

Definition (martingale difference)

A process $\{Y_t, t \in \mathbb{Z}\}$, defined as the first difference of a martingale X_t is called a **martingale difference**, with

$$Y_t = X_t - X_{t-1}$$

$$\mathbb{E}\left(Y_{t+1} \middle| \underline{Y_t}\right) = \mathbb{E}\left(Y_{t+1} \middle| \underline{X_t}\right) = \mathbb{E}\left(X_{t+1} - X_t \middle| \underline{X_t}\right) = 0$$

Notes

- The martingale difference process says that conditional on the asset's price history, the asset's expected price changes are zero.
- In this sense, information $\underline{X_t}$ contained in past prices is instantly and fully reflected in the asset's current price and hence useless in predicting rates of return.

Remarks

- A martingale difference is similar to a (uncorrelated) **white noise** except that it needs not have constant conditional variance and that its conditional mean is zero.
- (Uncorrelated) white noise and martingale differences have **constant mean** and **zero autocorrelations**. Note that definitions do not specify the nonlinear properties of such sequences.
- A martingale difference with the conditional mean equal to zero and a constant variance

$$\mathbb{E}\left(Y_{t+1} \,|\, \underline{Y_t} \right) = 0 \quad \mathbb{V}\left(Y_{t+1} \right) = \sigma^2$$

is called a **homoscedastic martingale difference**.

Summary

Name	Notation	Properties
IID noise	$\varepsilon_t \sim \text{IID}\left(0, \sigma^2\right)$	No dependencies (linear or nonlinear) with past/future values. Constant variance and mean
White noise	$\varepsilon_t \sim \text{WN}\left(0, \sigma^2\right)$	No correlation with past/future values Constant variance and mean
Gaussian WN	$\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right)$	No dependencies (linear or nonlinear) with past/future values. ε_t has a normal distribution
Martingale diff.	$\mathbb{E}\left(\varepsilon_t \middle\vert \underline{\varepsilon_{t-1}}\right) = 0$	No correlation with past values Conditional mean equal to 0, no constraint on the conditional variance

Key Concepts

1. Strict stationarity
2. (Weak) stationarity
3. IID noise or strict white noise
4. Uncorrelated white noise or white noise
5. Gaussian white noise
6. Martingale and martingale difference

Section 3

Wold Decomposition and Prediction

3. Wold decomposition and prediction

Objectives

1. To define the **Wold decomposition**
2. To introduce the notion of **optimal forecast**
3. To define the **innovation** process
4. To introduce the **lag operator**
5. To define the **lag polynomials** process

3. Wold decomposition and prediction

Theorem (Wold decomposition)

Any (weak) stationary time series $\{X_t, \; t \in \mathbb{Z}\}$ *can be represented as a* **Wold decomposition**, *given by*

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \mu = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots + \mu$$

where the parameters ψ_j *satisfy* $\psi_0 = 1$, $\psi_j \in \mathbb{R}, \forall j \in \mathbb{N}^*$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ *,* $\varepsilon_t \sim WN(0, \sigma^2)$ *is a white noise process and* $\mu = \mathbb{E}(X_t)$ *denotes the mean of* X_t.

Wold, H. (1938), A Study in the Analysis of Stationary Time Series. Almqvist and Wiksell.

Wold, H. (1954) A Study in the Analysis of Stationary Time Series, Second revised edition.

Remarks

1. This representation only exploits the covariance stationary property: neither a distributional assumption, nor the independence of the error terms are required.

2. The Wold representation can also be written as

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \mu_t$$

where μ_t denotes the **deterministic linear component** such that $cov(\mu_t, \varepsilon_s) = 0$, $\forall \, (t, s) \in \mathbb{Z}^2$.

3. The condition $\psi_0 = 1$ is a **normalization** of the variance of the white noise process.

3. Wold decomposition and prediction

Example (normalization of the variance of the white noise)

Let us consider the following process

$$X_t = \mu + \sum_{j=0}^{\infty} \widetilde{\psi}_j v_{t-j} = \mu + \frac{1}{2} v_t + \left(\frac{1}{2}\right)^2 v_{t-1} + \left(\frac{1}{2}\right)^3 v_{t-2} + \ldots$$

with $v_t \sim \text{WN}\left(0, \sigma^2\right)$ and $\sigma_v^2 = 1$. It is possible to **normalize** the variance of the white noise process such that the first parameter ψ_0 is equal to one. Define ε_t such that

$$\varepsilon_t = \frac{1}{2} v_t \sim \text{WN}\left(0, \sigma_\varepsilon^2\right)$$

with $\sigma_\varepsilon^2 = 1/4$. The process $\{X_t, \ t \in \mathbb{Z}\}$ can be rewritten as

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \mu + \varepsilon_t + \frac{1}{2} \varepsilon_{t-1} + \left(\frac{1}{2}\right)^2 \varepsilon_{t-2} + \left(\frac{1}{2}\right)^3 \varepsilon_{t-3} + \ldots.$$

Forecasts

- Suppose we are interested in forecasting the value of Y_{t+1} based on a set of variables X_t observed at date t.
- For instance, we might want to forecast Y_{t+1} based on its m most recent values. In this case, X_t would consist in a constant plus $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-m}$.
- Let $\widehat{Y}_{t+1|t}$ denote a **forecast** of Y_{t+1} based on X_t.
- To evaluate the usefulness of this forecast we need a to specify a **loss function**.

3. Wold decomposition and prediction

Definition (mean squared error and optimal forecast)

The **mean squared error (MSE)** associated to the forecast $\widehat{Y}_{t+1|t}$ is a quadratic loss function defined as

$$MSE\left(\widehat{Y}_{t+1|t}\right) = \mathbb{E}\left(\left(Y_{t+1} - \widehat{Y}_{t+1|t}\right)^2\right)$$

The **optimal forecast** with the smallest MSE is the expectation of Y_{t+1} conditional on X_t

$$\widehat{Y}^*_{t+1|t} = \mathbb{E}\left(Y_{t+1} \middle| X_t \right)$$

Prediction and Wold decomposition

Any (weak) stationary time series $\{X_t,\ t \in \mathbb{Z}\}$ can be represented in the form:

$$X_{t+1} = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t+1-j} = \mu + \varepsilon_{t+1} + \psi_1 \varepsilon_t + \psi_2 \varepsilon_{t-1} + \psi_3 \varepsilon_{t-2} \ldots$$

Let $\widehat{X}_{t+1|t}$ denote a forecast of X_{t+1} based on the past values $\underline{X}_t = \{X_t, X_{t-1}, \ldots\}$

$$\widehat{X}_{t+1|t} = \mathbb{E}\left(X_{t+1} \middle| \underline{X}_t \right) = \mathbb{E}\left(X_{t+1} \middle| \underline{\varepsilon}_t \right)$$

since X_t depends on the current and past values of ε_t.

$$\begin{aligned} \widehat{X}_{t+1|t} &= \mathbb{E}\left(X_{t+1} \middle| \underline{\varepsilon}_t \right) \\ &= \mu + \sum_{j=0}^{\infty} \psi_j \mathbb{E}\left(\varepsilon_{t+1-j} \middle| \underline{\varepsilon}_t \right) \\ &= \mu + \psi_1 \varepsilon_t + \psi_2 \varepsilon_{t-1} + \psi_3 \varepsilon_{t-2} + \ldots \end{aligned}$$

3. Wold decomposition and prediction

Definition (Wold decomposition and optimal forecast)

The **optimal forecast** $\widehat{X}_{t+1|t}$ of X_{t+1} based on the Wold decomposition is given by

$$\widehat{X}_{t+1|t} = \mathbb{E}\left(X_{t+1} | \underline{X}_t\right) = \mu + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t+1-j}$$

The corresponding **forecast error** is defined by

$$X_{t+1} - \widehat{X}_{t+1|t} = \varepsilon_{t+1}$$

Notes:

1. ε_{t+1} is a (weak) white noise process. Say differently, ε_{t+1} is the new information that appears at time $t+1$ and that was not predictable at time t.

2. Note that $\mathbb{E}\left(\varepsilon_{t+1}\right) = 0$ and $\mathbb{E}\left(\varepsilon_{t+1} Y_{t-k}\right) = \mathbb{C}ov\left(\varepsilon_{t+1}, Y_{t-k}\right) = 0$ for $k \geq 0$. This is like an "exogeneity assumption".

Definition (innovation process)

The **innovation process** of $\{X_t, \ t \in \mathbb{Z}\}$ is defined to be

$$\varepsilon_t = X_t - \mathbb{E}\left(X_t | \, \underline{X}_{t-1} \right)$$

where the optimal forecast of X_t given the available information at time $t-1$ denoted $\underline{X}_{t-1} = \{X_{t-1}, X_{t-2}, \ldots\}$ is defined to be

$$\widehat{X}_{t|t-1} = \mathbb{E}\left(X_t | \, \underline{X}_{t-1} \right)$$

Example (approximation of the Wold decomposition)

Consider the annual US GDP growth rate Y_t for the period 1961-2017 (source: World Bank national accounts data), and generate a Gaussian white noise $\varepsilon_t \sim$i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$ with $\sigma^2 = 1$. **Question:** (1) estimate the parameters of the following model (without normalization on ψ_0)

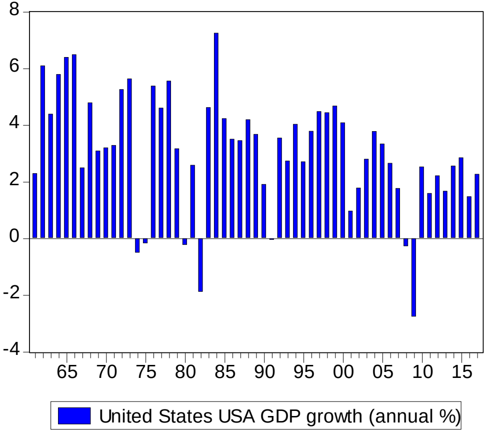
$$Y_t = \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \ldots + \psi_{20} \varepsilon_{t-20} + v_t$$

where v_t is an error term, and (2) evaluate the goodness of fit.

Note: the data are available within the file `GDP_growth-rate.xlsx`.

3. Wold decomposition and prediction

Figure: US GDP annual growth rate in percentage (1961-2017)



3. Wold decomposition and prediction

Figure: Approximation of the Wold decomposition for the US GDP annual growth rate (1961-2017)

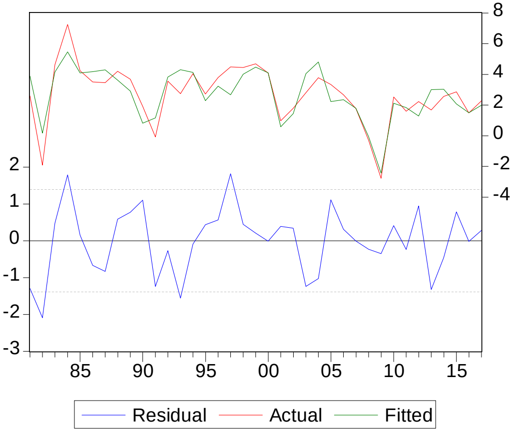
Dependent Variable: Y_US
Method: Least Squares
Date: 10/20/18 Time: 22:44
Sample(adjusted): 1981 2017
Included observations: 37 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.19792	3.512660	4.041928	0.0011
EPS	-1.784704	1.198886	-1.488635	0.1573
EPS(-1)	-1.276554	1.102529	-1.157842	0.2650
EPS(-2)	0.522025	0.989626	0.527497	0.6056
EPS(-3)	-2.596075	1.041371	-2.492940	0.0248
EPS(-4)	-4.133671	0.923146	-4.477808	0.0004
EPS(-5)	-1.362799	0.965588	-1.411367	0.1785
EPS(-6)	-1.297035	0.978888	-1.325008	0.2050
EPS(-7)	-1.208553	0.952680	-1.268583	0.2239
EPS(-8)	0.642747	0.923126	0.696271	0.4969
EPS(-9)	0.993863	0.883733	1.124620	0.2784
EPS(-10)	-2.322782	0.899470	-2.582390	0.0208
EPS(-11)	-2.025511	0.916455	-2.210158	0.0431
EPS(-12)	-1.505990	0.959527	-1.569513	0.1374
EPS(-13)	-0.566906	0.992408	-0.571243	0.5763
EPS(-14)	-1.470803	0.997424	-1.474602	0.1610
EPS(-15)	-0.629718	1.013401	-0.621391	0.5437
EPS(-16)	-0.833692	0.979533	-0.851112	0.4081
EPS(-17)	-0.329598	1.005074	-0.327934	0.7475
EPS(-18)	-0.707648	0.962812	-0.734981	0.4737
EPS(-19)	-1.676324	1.132913	-1.479658	0.1597
EPS(-20)	-0.351043	1.025700	-0.342247	0.7369

R-squared	0.770004	Mean dependent var	2.686138
Adjusted R-squared	0.448009	S.D. dependent var	1.872178
S.E. of regression	1.390955	Akaike info criterion	3.784179
Sum squared resid	29.02133	Schwarz criterion	4.742022
Log likelihood	-48.00731	F-statistic	2.391356
Durbin-Watson stat	1.625226	Prob(F-statistic)	0.044078

3. Wold decomposition and prediction

Figure: Approximation of the Wold decomposition for the US GDP annual growth rate (1961-2017)

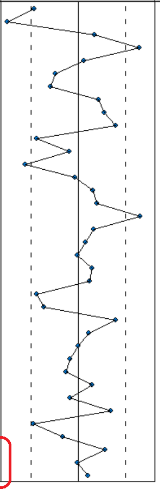


3. Wold decomposition and prediction

Figure: Approximation of the Wold decomposition for the US GDP annual growth rate (1961-2017)

obs	Actual	Fitted	Residual
1981	2.59447	3.88703	-1.29256
1982	-1.91089	0.17755	-2.08844
1983	4.63246	4.15525	0.47721
1984	7.25909	5.46867	1.79042
1985	4.23874	4.08603	0.15271
1986	3.51161	4.17991	-0.66830
1987	3.46175	4.29188	-0.83014
1988	4.20397	3.62095	0.58302
1989	3.68052	2.90423	0.77629
1990	1.91937	0.81815	1.10122
1991	-0.07408	1.16336	-1.23744
1992	3.55540	3.82530	-0.26990
1993	2.74586	4.30358	-1.55773
1994	4.03764	4.12938	-0.09174
1995	2.71898	2.28331	0.43566
1996	3.79588	3.23298	0.56290
1997	4.48703	2.66744	1.81958
1998	4.44991	4.00240	0.44751
1999	4.68520	4.47635	0.20885
2000	4.09218	4.10189	-0.00971
2001	0.97598	0.58285	0.39313
2002	1.78613	1.44278	0.34335
2003	2.80678	4.04878	-1.24201
2004	3.78574	4.81123	-1.02549
2005	3.34522	2.23457	1.11065
2006	2.66663	2.35347	0.31315
2007	1.77857	1.78748	-0.00891
2008	-0.29162	-0.06265	-0.22897
2009	-2.77553	-2.42467	-0.35086
2010	2.53192	2.12115	0.41077
2011	1.60145	1.83462	-0.23316
2012	2.22403	1.28115	0.94288
2013	1.67733	3.00141	-1.32408
2014	2.56919	3.02584	-0.45665
2015	2.86159	2.07791	0.78368
2016	1.48528	1.50271	-0.01743
2017	2.27334	1.99281	0.28053

Residual Plot



It is possible to rewrite the Wold decomposition by introducing a lag operator.

Definition (lag operator)

Consider a time series process $\{X_t, \ t \in \mathbb{Z}\}$, the **lag operator** (or backshift operator), denoted L or B, is defined by

$$LX_t = X_{t-1} \ \ \forall \, t \in \mathbb{Z}$$

Property 1. *$L^j X_t = X_{t-j}$, $\forall\, j \in \mathbb{Z}$, and in particular $L^0 X_t = X_t$.*

Property 2. *If $X_t = c$, $\forall\, t \in \mathbb{Z}$ with $c \in \mathbb{R}$, $L^j X_t = L^j c = c$, $\forall\, j \in \mathbb{Z}$.*

Property 3. $L^i \left(L^j X_t \right) = L^{i+j} X_t = X_{t-i-j} \;\; \forall\, (i, j) \in \mathbb{Z}^2$.

Property 4. $L^{-i} X_t = X_{t+i} \;\; \forall\, i \in \mathbb{Z}$.

Property 5. $\left(L^i + L^j \right) X_t = L^i X_t + L^j X_t = X_{t-i} + X_{t-j} \;\; \forall\, (i, j) \in \mathbb{Z}^2$.

Property 6. *If* $|a| < 1$ then

$$(1 - aL)^{-1} X_t = \sum_{j=0}^{\infty} a^j L^j X_t = \sum_{j=0}^{\infty} a^j X_{t-j} = X_t + aX_{t-1} + a^2 X_{t-2} + \ldots$$

3. Wold decomposition and prediction

Definition (lag polynomial)

A polynomial of lag operators is called a **lag polynomial**.

Example (lag polynomial)

Consider the lag polynomial given by

$$\Theta(L) = 1 - 2L + 3L^2$$

and a time series $\{X_t, \ t \in \mathbb{Z}\}$. Then,

$$\Theta(L) X_t = \left(1 - 2L + 3L^2\right) X_t = X_t - 2X_{t-1} + 3X_{t-2}$$

Example (lag polynomial)

Consider the lag (infinite) **polynomial** given by

$$\Psi\left(L\right)=\sum_{j=0}^{\infty}\psi_{j}L^{j}=\psi_{0}+\psi_{1}L+\psi_{2}L^{2}+\ldots$$

and a time series $\{X_t,\ t\in\mathbb{Z}\}$. Then,

$$\begin{aligned}\Psi\left(L\right)X_{t} &= \sum_{j=0}^{\infty}\psi_{j}L^{j}X_{t}\\ &= \sum_{j=0}^{\infty}\psi_{j}X_{t-j}\\ &= \psi_{0}X_{t}+\psi_{1}X_{t-1}+\psi_{2}X_{t-2}+\ldots\end{aligned}$$

3. Wold decomposition and prediction

Theorem (Wold decomposition)

Any stationary time series $\{X_t, \ t \in \mathbb{Z}\}$ can be represented as a **Wold decomposition**, *given by*

$$X_t = \Psi(L)\,\varepsilon_t + \mu$$

where the **lag (infinite) polynomial** $\Psi(L)$ *is defined as* $\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$*, with* $\psi_0 = 1$, $\psi_j \in \mathbb{R}, \forall j \in \mathbb{N}^*$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, $\varepsilon_t \sim WN(0, \sigma^2)$ *is a white noise process and* $\mu = \mathbb{E}(X_t)$ *denotes the mean of* X_t.

Key Concepts

1. Wold decomposition
2. Optimal forecast
3. Loss function
4. Innovation process
5. Lag operator
6. Lag polynomial

Section 4

The Box-Jenkins Modeling Approach

4. The Box-Jenkins modeling approach

Objectives

1. To introduce the **Box-Jenkins** modeling approach
2. To introduce the principle of **parsimony**
3. To define the **autocorrelation function**
4. To define the **partial autocorrelation function**

The Box-Jenkins modeling approach

1. The Wold decomposition is useful for theoretical reasons. However, in practice, applications of models with an infinite number of parameters are hardly useful.
2. Many forecasters are persuaded of the benefits of **parsimony**, or using as few parameters as possible.
3. Although complicated parameters can track the data very well over the historical period for which parameters are estimated, they often perform poorly when used for **out-of-sample** forecasts.
4. **Box and Jenkins (1976)** recommend the use of univariate time series models with a "small" number of parameters.

Box, G.E and G.M. Jenkins, *Time Series Analysis, Forecasting and Control*, Wiley, 1976.

4. The Box-Jenkins modeling approach

The approach of Box and Jenkins (1976) can be broken into five steps:

Step 1: Transform the data, if necessary, so that the assumption of (weak) stationarity is a reasonable one.

Step 2: Use some **identification tools** (autocorrelation function, partial autocorrelation function, etc.) in order to compare some properties of the data to the "theoretical" properties of some **times series models** (AR, MA, ARMA, etc.), and choose a model.

Step 3: Estimate the parameters of the model.

Step 4: Perform diagnostic analysis to confirm that the model is indeed consistent with the observed features of the data.

Step 5: Use the estimated model to produce the forecasts.

4. The Box-Jenkins modeling approach

There are two main identification tools for the times series models

1. **The autocorrelation function (ACF)**
2. **The partial autocorrelation function (PACF)**

4. The Box-Jenkins modeling approach

There are two main identification tools for the times series models

1. **The autocorrelation function (ACF)**
2. **The partial autocorrelation function (PACF)**

Definition (autocorrelation function)

Let $\{X_t, t \in \mathbb{Z}\}$ be a stationary time series, its **autocorrelation function** is defined as

$$\rho(h) \equiv \mathbb{C}orr(X_t, X_{t-h}) = \frac{\mathbb{C}ov(X_t, X_{t-h})}{\mathbb{V}(X_t)} = \frac{\gamma(h)}{\gamma(0)} \quad \forall h \in \mathbb{Z}$$

where $\gamma(h) \equiv \mathbb{C}ov(X_t, X_{t-h})$ is the autocovariance function.

Properties:

- $\rho(h) = \rho(-h), \quad \forall h \in \mathbb{Z}$
- $\rho(0) = 1.$
- The range of $\rho(h)$ is $[-1, 1]$.

Definition (sample autocorrelation)

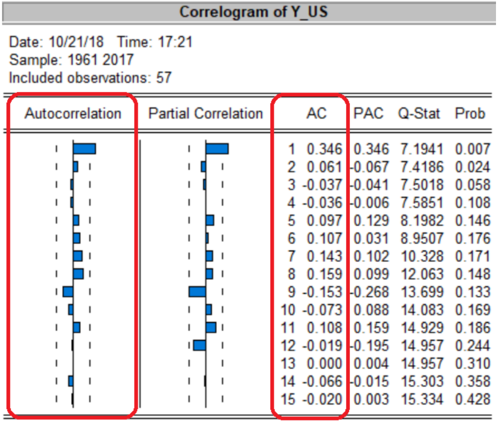
The **(sample) autocorrelation function (ACF)**, denoted $\widehat{\rho}(h)$, of a stationary process $\{X_t, t \in \mathbb{Z}\}$ is a consistent estimator of $\rho(h)$ defined as

$$\widehat{\rho}(h) = corr(X_t, X_{t-h}) = \frac{\sum_{t=h+1}^{T} (x_t - \widehat{\mu})(x_{t-h} - \widehat{\mu})}{\sum_{t=1}^{T} (x_t - \widehat{\mu})^2}$$

where $\widehat{\mu} = T^{-1} \sum_{t=1}^{T} x_t$ is the sample mean of $\{x_1, \ldots, x_T\}$.

Note: in general the ACF refers to the **(sample)** autocorrelation function.

Figure: ACF of the US GDP annual growth rate (1961-2017)


Correlogram of Y_US

Date: 10/21/18 Time: 17:21
Sample: 1961 2017
Included observations: 57

	AC	PAC	Q-Stat	Prob
1	0.346	0.346	7.1941	0.007
2	0.061	-0.067	7.4186	0.024
3	-0.037	-0.041	7.5018	0.058
4	-0.036	-0.006	7.5851	0.108
5	0.097	0.129	8.1982	0.146
6	0.107	0.031	8.9507	0.176
7	0.143	0.102	10.328	0.171
8	0.159	0.099	12.063	0.148
9	-0.153	-0.268	13.699	0.133
10	-0.073	0.088	14.083	0.169
11	0.108	0.159	14.929	0.186
12	-0.019	-0.195	14.957	0.244
13	0.000	0.004	14.957	0.310
14	-0.066	-0.015	15.303	0.358
15	-0.020	0.003	15.334	0.428

4. The Box-Jenkins modeling approach

There are two main identification tools for the times series models

1. **The autocorrelation function (ACF)**
2. **The partial autocorrelation function (PACF)**

Definition (partial autocorrelation function)

The **partial autocorrelation function** $\alpha(h)$ of a stationary time series $\{X_t, t \in \mathbb{Z}\}$ is defined as the correlation between X_t and X_{t-h}, conditional to the intervening observations $x_{t-1}, \ldots, x_{t-h+1}$.

$$\alpha(h) \equiv \mathbb{C}orr\left(X_t, X_{t-h} \middle| X_{t-1}, \ldots, X_{t-h+1}\right) \quad \forall h \in \mathbb{Z}$$

Definition (partial autocorrelation function)

The **partial autocorrelation function** $\alpha(h)$ corresponds to the sequence of the hth autoregressive coefficients a_{hh} obtained in the multiple linear regression model:

$$X_t = c + a_{h1}X_{t-1} + a_{h2}X_{t-2} + \ldots + a_{hh}X_{t-h} + v_t$$

Then, we have

$$\alpha(h) = a_{hh}$$

Properties:

- $\alpha(0) = 1.$
- $\alpha(1) = \rho(1).$

Definition (sample partial autocorrelation function)

The **(sample) partial autocorrelation function (PACF)**, denoted $\widehat{\alpha}(h)$, corresponds to the sequence of the hth autoregressive **estimated** coefficients a_{hh} obtained by OLS in the regression

$$X_t = c + a_{h1} X_{t-1} + a_{h2} X_{t-2} + \ldots + a_{hh} X_{t-h} + v_t$$

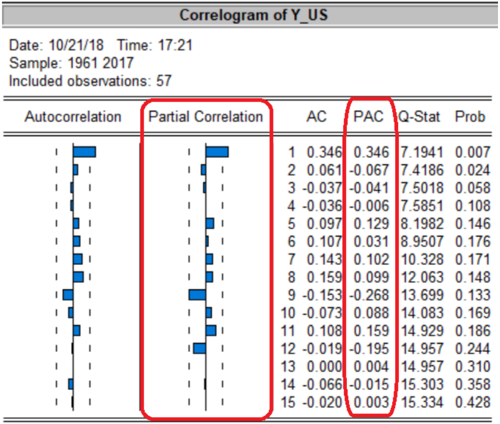
Then, we have

$$\widehat{\alpha}(h) = \widehat{a}_{hh}$$

Note: in general the PACF refers to the **(sample)** partial autocorrelation function.

4. The Box-Jenkins modeling approach

Figure: PACF of the US GDP annual growth rate (1961-2017)



Correlogram of Y_US

Date: 10/21/18 Time: 17:21
Sample: 1961 2017
Included observations: 57

	AC	PAC	Q-Stat	Prob
1	0.346	0.346	7.1941	0.007
2	0.061	-0.067	7.4186	0.024
3	-0.037	-0.041	7.5018	0.058
4	-0.036	-0.006	7.5851	0.108
5	0.097	0.129	8.1982	0.146
6	0.107	0.031	8.9507	0.176
7	0.143	0.102	10.328	0.171
8	0.159	0.099	12.063	0.148
9	-0.153	-0.268	13.699	0.133
10	-0.073	0.088	14.083	0.169
11	0.108	0.159	14.929	0.186
12	-0.019	-0.195	14.957	0.244
13	0.000	0.004	14.957	0.310
14	-0.066	-0.015	15.303	0.358
15	-0.020	0.003	15.334	0.428

Figure: PACF of the US GDP annual growth rate (1961-2017)

Correlogram of Y_US

Date: 10/21/18 Time: 17:40
Sample: 1961 2017
Included observations: 57

	AC	PAC	Q-Stat	Prob
1	0.346	0.346	7.1941	0.007
2	0.061	-0.067	7.4186	0.024
3	-0.037	-0.041	7.5018	0.058
4	-0.036	-0.006	7.5851	0.108
5	0.097	0.129	8.1982	0.146
6	0.107	0.031	8.9507	0.176
7	0.143	0.102	10.328	0.171
8	0.159	0.099	12.063	0.148
9	-0.153	-0.268	13.699	0.133
10	-0.073	0.088	14.083	0.169
11	0.108	0.159	14.929	0.186
12	-0.019	-0.195	14.957	0.244
13	0.000	0.004	14.957	0.310
14	-0.066	-0.015	15.303	0.358
15	-0.020	0.003	15.334	0.428

Dependent Variable: Y_US
Method: Least Squares
Date: 10/21/18 Time: 17:44
Sample(adjusted): 1964 2017
Included observations: 54 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.170981	0.638049	3.402533	0.0013
Y_US(-1)	0.377050	0.141300	2.668433	0.0102
Y_US(-2)	-0.053536	0.148142	-0.361383	0.7193
Y_US(-3)	-0.040946	0.138166	-0.296353	0.7682

R-squared	0.132344	Mean dependent var	3.030896
Adjusted R-squared	0.080285	S.D. dependent var	2.061144
S.E. of regression	1.976674	Akaike info criterion	4.271895
Sum squared resid	195.3620	Schwarz criterion	4.419227
Log likelihood	-111.3412	F-statistic	2.542181
Durbin-Watson stat	1.975338	Prob(F-statistic)	0.066744

Key Concepts

1. The Box-Jenkins modeling approach
2. The parsimony principle
3. The autocorrelation function
4. The partial autocorrelation function

Section 5

Univariate Time Series Models

Objectives

1. To define the **moving average (MA)** process
2. To define the **autoregressive (AR)** process
3. To define the **autoregressive moving average (ARMA)** process
4. To introduce the **invertibility** and **stationarity** conditions
5. To **identify** the AR, MA and ARMA processes

Some time series models are particularly useful for empirical applications

1. The moving average (**MA**) model
2. The autoregressive (**AR**) model
3. The mixed autoregressive moving average (**ARMA**) model

Definition (moving average - MA process)

The process $\{X_t,\ t \in \mathbb{Z}\}$ is said to be a **moving average** of order q or a **MA**(q) **process**, if

$$X_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

where $\theta_1, \ldots, \theta_q$ are constants and $\varepsilon_t \sim \text{WN}\left(0, \sigma^2\right)$ is a white noise process.

Note: by definition, $c = \mathbb{E}\left(X_t\right)$.

Example (MA processes)

Denote by ε_t a white noise process with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{V}(\varepsilon_t) = \sigma^2$. The process $\{X_t, \ t \in \mathbb{Z}\}$ is a **MA(1) process** with a null mean if

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

The process $\{Z_t, \ t \in \mathbb{Z}\}$ is a **MA(3) process** if

$$Z_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3}$$

Notice that $\{Y_t, \ t \in \mathbb{Z}\}$ is also a **MA(3) process** defined as

$$Y_t = c + \varepsilon_t + \theta_3 \varepsilon_{t-3}$$

even if the parameters θ_1 and θ_2 are null.

The MA(q) process can be rewritten using a lag polynomial.

Definition (moving average - MA process)

The process $\{X_t,\ t \in \mathbb{Z}\}$ is said to be a **MA(q) process**, if

$$X_t = c + \Theta(L)\,\varepsilon_t$$

where the lag polynomial $\Theta(L)$ is defined by $\Theta(L) = \sum_{j=0}^{q} \theta_j L^j$ with $\forall j < q$, $\theta_j \in \mathbb{R}$, $\theta_0 = 1$, $\theta_q \in \mathbb{R}^*$, and $\varepsilon_t \sim \text{WN}\left(0, \sigma^2\right)$ is a white noise process.

Example (MA processes)

Consider the following MA processes

$$X_t = \varepsilon_t + 0.5\varepsilon_{t-1}$$

$$Z_t = 1 - 0.8\varepsilon_{t-3} + 1.2\varepsilon_{t-2} + \varepsilon_t$$

$$Y_t = -0.5 + \varepsilon_t - 0.6\varepsilon_{t-3}$$

Question: write the lag order polynomials associated to these MA processes.

Solution

$$X_t = \Theta(L)\,\varepsilon_t \quad \text{with } \Theta(L) = 1 + 0.5L$$

$$Z_t = c + \Theta(L)\,\varepsilon_t \quad \text{with } \Theta(L) = 1 + 1.2L^2 - 0.8L^3 \quad \text{and } c = 1$$

$$Y_t = c + \Theta(L)\,\varepsilon_t \quad \text{with } \Theta(L) = 1 - 0.6L^3 \quad \text{and } c = -0.5$$

Wold decomposition

The Wold decomposition is an MA process with an infinite order, denoted MA(∞).

$$X_t = \Psi(L)\,\varepsilon_t + \mu$$

with

$$\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$$

$$\psi_0 = 1 \qquad \sum_{j=0}^{\infty} \psi_j^2 < \infty$$

$$\varepsilon_t \sim \text{WN}\left(0, \sigma^2\right)$$

Stationarity of a MA(q) process

1. A MA(q) process is the weighted sum of q lagged values of a white noise, which is a stationary process.
2. By definition, a MA(q) process is **always stationary** whatever the values of the parameters $\theta_1, \ldots, \theta_q$.

Example (simulation of MA processes)

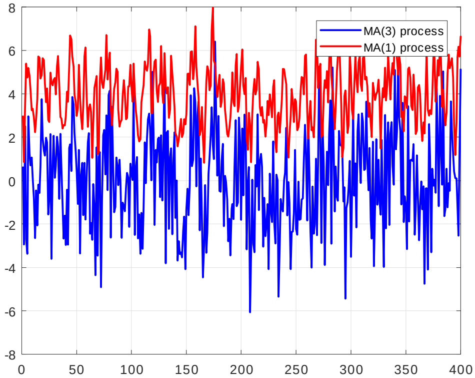
Consider a Gaussian white noise $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and

$$X_t = \varepsilon_t + 2\varepsilon_{t-3}$$

$$Y_t = 4 + \varepsilon_t - 0.8\varepsilon_{t-1}$$

where X_t is a MA(3) process with $\mathbb{E}(X_t) = 0$, and Y_t is a MA(1) process with $\mathbb{E}(Y_t) = 4$. **Question:** simulate a sample of size $T = 500$ of the two MA processes.

Figure: Simulation of MA(1) and MA(3) processes



Definition (AutoRegressive - AR process)

The process $\{X_t,\ t \in \mathbb{Z}\}$ is said to be an autoregressive process of order p, or briefly an **AR**(p) **process**, if

$$X_t = c + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t$$

where $\phi_1, \ldots, \phi_p$ are constants and $\varepsilon_t \sim \mathrm{WN}\left(0, \sigma^2\right)$ is a white noise process.

Property: The mean of the process X_t is given by

$$\mathbb{E}\left(X_t\right) = \frac{c}{1 - \phi_1 - \ldots - \phi_p}$$

Example (AR processes)

Denote by ε_t a white noise process with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{V}(\varepsilon_t) = \sigma^2$. The process $\{X_t, \, t \in \mathbb{Z}\}$ is a **AR(1) process** with a null mean if

$$X_t = \phi_1 X_{t-1} + \varepsilon_t$$

The process $\{Z_t, \, t \in \mathbb{Z}\}$ is a **AR(3) process** if

$$Z_t = c + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + \varepsilon_t$$

Notice that $\{Y_t, \, t \in \mathbb{Z}\}$ is also a **AR(3) process** defined as

$$Y_t = c + \phi_3 Y_{t-3} + \varepsilon_t$$

even if the parameters ϕ_1 and ϕ_2 are null.

The AR(p) process can be rewritten using a lag polynomial.

Definition (AutoRegressive - AR process)

The process $\{X_t,\ t \in \mathbb{Z}\}$ is said to be a **AR(p) process**, if

$$\Phi(L) X_t = c + \varepsilon_t$$

where the lag polynomial $\Phi(L)$ is defined by $\Phi(L) = \sum_{j=0}^{p} \phi_j L^j$ with $\forall j < p$, $\phi_j \in \mathbb{R}$, $\phi_0 = 1$, $\phi_p \in \mathbb{R}^*$, and $\varepsilon_t \sim \mathrm{WN}\left(0, \sigma^2\right)$ is a white noise process.

Property: The mean of the process X_t is

$$\mathbb{E}(X_t) = \frac{c}{1 - \phi_1 - \ldots - \phi_p} = c\Phi(1)^{-1}$$

Example (AR processes)

Consider the following AR processes

$$X_t = 0.5X_{t-1} + \varepsilon_t$$

$$Z_t = 1 - 0.8Z_{t-1} + 1.2Z_{t-2} + \varepsilon_t$$

$$Y_t = -0.5 - 0.6Y_{t-2} + \varepsilon_t$$

Question: write the lag order polynomials associated to these AR processes.

Solution

$$\Phi(L) X_t = \varepsilon_t \quad \text{with } \Phi(L) = 1 - 0.5L$$

$$\Phi(L) Z_t = c + \varepsilon_t \quad \text{with } \Phi(L) = 1 + 0.8L - 1.2L^2 \quad \text{and } c = 1$$

$$\Phi(L) Y_t = c + \varepsilon_t \quad \text{with } \Phi(L) = 1 + 0.6L^2 \quad \text{and } c = -0.5$$

Stationarity of a AR(p) process

1. An AR(p) process may be stationary or not, depending on the values of the parameters $\phi_1, \ldots, \phi_q$.

Example (simulation of a AR processes)

Consider a Gaussian white noise $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, 0.25\right)$ and

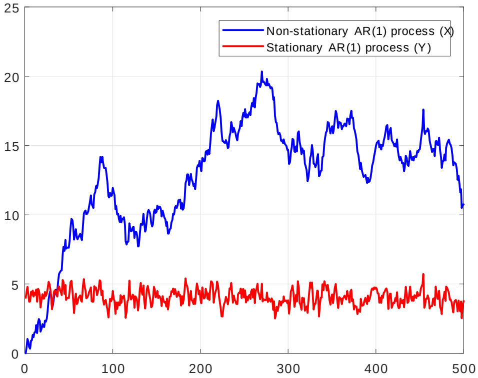
$$X_t = X_{t-1} + \varepsilon_t$$

$$Y_t = 2 + 0.5 X_{t-1} + \varepsilon_t$$

where X_t is a non-stationary AR(1) process with $\mathbb{E}\left(X_t\right) = 0$, and Y_t is a stationary AR(1) process with $\mathbb{E}\left(Y_t\right) = 2/\left(1 - 0.5\right) = 4$. **Question:** simulate a sample of size $T = 500$ of the two AR processes.

5. Univariate time series models

Figure: Simulation of two AR(1) processes



Definition (ARMA process)

The process $\{X_t,\ t \in \mathbb{Z}\}$ is said to be a **ARMA**(p, q) process, if

$$\Phi(L) X_t = c + \Theta(L) \varepsilon_t$$

where the lag polynomials $\Phi(L)$ and $\Theta(L)$ are defined by $\Phi(L) = \sum_{j=0}^{p} \phi_j L^j$ with $\forall j < p$, $\phi_j \in \mathbb{R}$, $\phi_0 = 1$, $\phi_p \in \mathbb{R}^*$, $\Theta(L) = \sum_{j=0}^{q} \theta_j L^j$ with $\forall j < q$, $\theta_j \in \mathbb{R}$, $\theta_0 = 1$, $\theta_q \in \mathbb{R}^*$, and $\varepsilon_t \sim \text{WN}\left(0, \sigma^2\right)$ is a white noise process.

Property: The mean of the process X_t is given by

$$\mathbb{E}(X_t) = \frac{c}{1 - \phi_1 - \ldots - \phi_p} = c\Phi(1)^{-1}$$

Example (ARMA processes)

Denote by ε_t a white noise process with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{V}(\varepsilon_t) = \sigma^2$. The process $\{X_t,\ t \in \mathbb{Z}\}$ is a **ARMA(1,3) process** with a null mean if

$$X_t = \phi X_{t-1} + \theta_3 \varepsilon_{t-3} + \theta_2 \varepsilon_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

The process $\{Z_t,\ t \in \mathbb{Z}\}$ is a **ARMA(2,1) process** if

$$Z_t = c + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

with

$$\mathbb{E}(Z_t) = \frac{c}{1 - \phi_1 - \phi_2}$$

Two properties of the AR / MA / ARMA processes are generally considered

1. The **invertibility condition** means that the process can be inverted. For instance, an AR(p) model can be alternatively represented as an MA(∞), an MA(q) process can be represented as an AR(∞), etc.

2. The **stationarity condition** guarantees that the process is (weakly) stationary.

Theorem (invertibility and stationarity of the MA processes)

A process $\{X_t,\ t \in \mathbb{Z}\}$ *with an* $MA(q)$ *representation is always* **stationary**. *It is also* **invertible** *if all the roots* λ_j *of the polynomial* $\Theta(L)$ *are all outside the unit circle, i.e. their modulus is larger than 1.*

$$\Theta(\lambda_j) = \sum_{i=0}^{q} \theta_i \lambda_j^i = \prod_{j=1}^{q} \left(1 - \frac{1}{\lambda_j} L\right) = 0 \quad \textit{with} \ \left|\lambda_j\right| > 1,\ \forall j$$

Example (invertibility and stationarity of the MA processes)

Consider the MA(1) process given by

$$X_t = \varepsilon_t - 0.5\varepsilon_{t-1} = \Theta\left(L\right)\varepsilon_t$$

with $\Theta\left(L\right) = 1 - 0.5L$. The root of the polynomial is equal to 2.

$$\Theta\left(\lambda\right) = 1 - 0.5\lambda = 0 \Leftrightarrow \lambda = 2$$

The root is outside the unit circle: the MA process is stationary (by definition) and invertible. This last property means that X_t can be expressed as an AR(∞)

$$\Theta\left(L\right)^{-1} X_t = \varepsilon_t$$

$$\Theta\left(L\right)^{-1} = \left(1 - 0.5L\right)^{-1} = \sum_{j=0}^{\infty} 0.5^j L^j$$

or equivalently

$$X_t + 0.5X_{t-1} + 0.5^2 X_{t-2} + \ldots = \varepsilon_t$$

Theorem (invertibility and stationarity of the AR processes)

A process $\{X_t, t \in \mathbb{Z}\}$ *with an* $AR(p)$ *representation is always* **invertible**. *It is also* **stationary** *if all the roots* λ_j $\forall j \leq p$ *of the polynomial* $\Phi(L)$ *are all outside the unit circle, i.e. their modulus is larger than 1.*

$$\Phi(\lambda_j) = \sum_{i=0}^{p} \phi_i \lambda_j^i = \prod_{j=1}^{p} \left(1 - \frac{1}{\lambda_j} L\right) = 0 \quad \textit{with } \left|\lambda_j\right| > 1, \forall j$$

Example (invertibility and stationarity of the AR processes)

Consider the two AR(2) processes given by

$$X_t = 2 - 0.5X_{t-1} + 0.2X_{t-3} + \varepsilon_t$$

$$Y_t = 0.2Y_{t-1} + 1.5Y_{t-2} + \varepsilon_t$$

where ε_t is a white noise. **Question:** Check if the processes X_t and Y_t are stationary.

Solution: the two corresponding lag polynomials are given by

$$\Phi(L) X_t = 2 + \varepsilon_t \text{ with } \Phi(\lambda) = 1 + 0.5\lambda - 0.2\lambda^3 = 0$$

$$\Longleftrightarrow \lambda_1 = 2.18 \text{ and } \lambda_j = -1.09 \pm 1.04i \quad j = 2, 3$$

$|\lambda_1| = 2.18 \quad |\lambda_2| = |\lambda_3| = 1.50$ All the roots are outside the unit circle: X_t **is stationary**

$$\Phi(L) Y_t = \varepsilon_t \text{ with } \Phi(\lambda) = 1 - 0.2\lambda - 1.5\lambda^2 = 0 \Longleftrightarrow \lambda_1 = -0.88 \text{ and } \lambda_2 = 0.75$$

$|\lambda_1| = 0.88$ and $|\lambda_2| = 0.75 \quad X_t$ **is non-stationary**

Theorem (invertibility of an ARMA process)

A process $\{X_t, \, t \in \mathbb{Z}\}$ with an ARMA(p, q) representation is **invertible** *if the roots of its MA polynomial $\Theta(L)$ are all outside the unit circle.*

Theorem (stationarity of an ARMA process)

A process $\{X_t, \, t \in \mathbb{Z}\}$ with an ARMA(p, q) representation is **stationary** *if the roots of its AR polynomial $\Phi(L)$ are all outside the unit circle.*

5. Univariate time series models

Figure: Example of ARMA estimation

Dependent Variable: Y_US
Method: Least Squares
Date: 10/21/18 Time: 21:20
Sample(adjusted): 1963 2017
Included observations: 55 after adjusting endpoints
Convergence achieved after 29 iterations
Backcast: 1961 1962

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.669006	1.332900	2.002406	0.0507
AR(1)	0.512444	0.198203	2.585452	0.0127
AR(2)	0.429447	0.239839	1.790560	0.0794
MA(1)	-0.421034	0.215269	-1.955852	0.0561
MA(2)	-0.824697	0.239246	-3.447073	0.0012

R-squared	0.356008	Mean dependent var	3.055789
Adjusted R-squared	0.304489	S.D. dependent var	2.050298
S.E. of regression	1.709894	Akaike info criterion	3.997248
Sum squared resid	146.1869	Schwarz criterion	4.179733
Log likelihood	-104.9243	F-statistic	6.910176
Durbin-Watson stat	2.003130	Prob(F-statistic)	0.000165

Inverted AR Roots	.96	-.45
Inverted MA Roots	1.14	-.72

Estimated MA process is noninvertible

How to choose a specification?

=> steps 2 and 3 of the Box-Jenkins approach

Step 2: Use some **identification tools** (autocorrelation function, partial autocorrelation function, etc.) in order to compare some properties of the data to the "theoretical" properties of some **times series models** (AR, MA, ARMA, etc.), and choose a model.

Step 3: Estimate the parameters of the model by **Maximum Likelihood** (ML).

Lemma (identification for MA process)

The autocorrelation function (ACF) of a **MA**(q) **process** *is zero at lag* $q+1$ *and greater.*

$$\rho\left(h\right)=0 \quad \textit{for } h>q$$

Note: Therefore, we determine the appropriate maximum lag order by examining the (sample) autocorrelation function to see where it becomes insignificantly different from zero for all lags beyond a certain lag, which is designated as the maximum lag q.

5. Univariate time series models

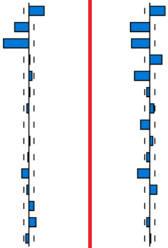
Figure: ACF of a simulated MA(1) process

Date: 10/21/18 Time: 21:38
Sample: 1 500
Included observations: 499

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.399	0.399	79.867	0.000
		2	-0.013	-0.204	79.946	0.000
		3	-0.001	0.107	79.947	0.000
		4	0.012	-0.041	80.022	0.000
		5	0.037	0.061	80.720	0.000
		6	0.007	-0.043	80.744	0.000
		7	-0.030	-0.009	81.199	0.000
		8	-0.016	-0.001	81.335	0.000
		9	-0.022	-0.027	81.589	0.000
		10	-0.026	-0.008	81.947	0.000
		11	-0.083	-0.090	85.435	0.000
		12	-0.058	0.024	87.134	0.000
		13	0.065	0.081	89.305	0.000
		14	0.051	-0.024	90.655	0.000
		15	-0.052	-0.060	92.061	0.000

Figure: ACF of a simulated MA(3) process

Date: 10/21/18 Time: 21:38
Sample: 1 500
Included observations: 497



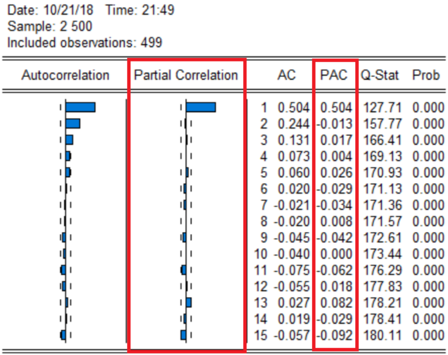
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.264	0.264	34.915	0.000
		2	-0.241	-0.334	64.014	0.000
		3	-0.428	-0.308	156.02	0.000
		4	0.024	0.212	156.32	0.000
		5	0.042	-0.259	157.21	0.000
		6	0.013	-0.047	157.29	0.000
		7	-0.039	0.079	158.05	0.000
		8	0.004	-0.153	158.06	0.000
		9	0.023	0.053	158.33	0.000
		10	-0.019	-0.043	158.52	0.000
		11	-0.117	-0.208	165.49	0.000
		12	-0.031	0.122	165.98	0.000
		13	0.096	0.000	170.71	0.000
		14	0.129	-0.049	179.29	0.000
		15	-0.049	0.011	180.55	0.000

Lemma (identification for AR process)

The partial autocorrelation function (PACF) of an **AR**(p) **process** *is zero at lag* $p+1$ *and greater.*

$$\alpha(p) = \phi_p \quad \alpha(h) = 0 \;\; \text{for } h > p$$

Figure: PACF of a simulated AR(1) process


Date: 10/21/18 Time: 21:49
Sample: 2 500
Included observations: 499

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.504	0.504	127.71	0.000
		2	0.244	-0.013	157.77	0.000
		3	0.131	0.017	166.41	0.000
		4	0.073	0.004	169.13	0.000
		5	0.060	0.026	170.93	0.000
		6	0.020	-0.029	171.13	0.000
		7	-0.021	-0.034	171.36	0.000
		8	-0.020	0.008	171.57	0.000
		9	-0.045	-0.042	172.61	0.000
		10	-0.040	0.000	173.44	0.000
		11	-0.075	-0.062	176.29	0.000
		12	-0.055	0.018	177.83	0.000
		13	0.027	0.082	178.21	0.000
		14	0.019	-0.029	178.41	0.000
		15	-0.057	-0.092	180.11	0.000

Figure: PACF of a simulated AR(3) process

Date: 10/21/18 Time: 21:51
Sample: 2 500
Included observations: 499

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.463	-0.463	107.81	0.000
		2	0.118	-0.124	114.78	0.000
		3	0.162	0.215	127.99	0.000
		4	-0.196	-0.021	147.47	0.000
		5	0.147	0.025	158.33	0.000
		6	-0.026	0.029	158.69	0.000
		7	-0.066	-0.036	160.87	0.000
		8	0.092	0.007	165.14	0.000
		9	-0.098	-0.050	170.08	0.000
		10	0.050	-0.001	171.33	0.000
		11	-0.008	-0.005	171.36	0.000
		12	-0.088	-0.083	175.38	0.000
		13	0.084	-0.011	178.99	0.000
		14	0.010	0.091	179.04	0.000
		15	-0.083	-0.026	182.61	0.000

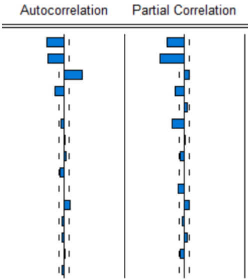
Lemma

In the case of a mixed **ARMA**(p, q) *with* $p \neq 0$ *and* $q \neq 0$*, neither the theoretical autocorrelation function not the theoretical partial autocorrelation function have any abrupt cutoffs.*

Note: Thus, there is little that can be inferred from ACF and PACF beyond the fact that neither a pure MA model nor a pure AR model would be inappropriate.

Figure: PACF of a simulated ARMA(3,2) process

Date: 10/21/18 Time: 22:09
Sample: 3 500
Included observations: 498



	AC	PAC	Q-Stat	Prob
1	-0.300	-0.300	45.198	0.000
2	-0.283	-0.411	85.542	0.000
3	0.307	0.084	132.92	0.000
4	-0.151	-0.150	144.39	0.000
5	-0.004	0.046	144.40	0.000
6	-0.042	-0.201	145.31	0.000
7	0.021	0.015	145.53	0.000
8	0.039	-0.072	146.30	0.000
9	-0.063	0.007	148.33	0.000
10	-0.007	-0.103	148.36	0.000
11	0.101	0.096	153.60	0.000
12	-0.038	-0.025	154.34	0.000
13	-0.032	0.055	154.86	0.000
14	0.020	-0.068	155.07	0.000
15	-0.025	0.002	155.39	0.000

Identification of the lag orders for an ARMA process

1. Brockwell and Davis (2009) recommend using **Akaike information criterion** (AIC) for finding p and q.

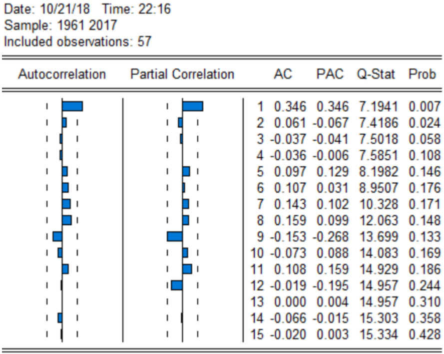
$$AIC\left(p,q\right) = 2\left(p+q+1\right) - 2\ln \widehat{L}_N\left(p,q\right)$$

   where $\ln \widehat{L}_N\left(p,q\right)$ is the log-likelihood of the sample associated to an ARMA(p,q) model.

2. ARMA models in general can be, after choosing p and q, fitted by least squares regression to find the values of the parameters which minimize the error term. It is generally considered good practice to find the smallest values of p and q which provide an acceptable fit to the data.

Brockwell, P. J.and Davis, R. A. (2009). *Time Series: Theory and Methods* (2nd ed.). New York: Springer. p. 273.

Figure: ACF and PACF of the US GDP annual growth rate (1961-2017)



Date: 10/21/18 Time: 22:16
Sample: 1961 2017
Included observations: 57

	AC	PAC	Q-Stat	Prob
1	0.346	0.346	7.1941	0.007
2	0.061	-0.067	7.4186	0.024
3	-0.037	-0.041	7.5018	0.058
4	-0.036	-0.006	7.5851	0.108
5	0.097	0.129	8.1982	0.146
6	0.107	0.031	8.9507	0.176
7	0.143	0.102	10.328	0.171
8	0.159	0.099	12.063	0.148
9	-0.153	-0.268	13.699	0.133
10	-0.073	0.088	14.083	0.169
11	0.108	0.159	14.929	0.186
12	-0.019	-0.195	14.957	0.244
13	0.000	0.004	14.957	0.310
14	-0.066	-0.015	15.303	0.358
15	-0.020	0.003	15.334	0.428

5. Univariate time series models

Figure: ARMA(3,6) for the US GDP annual growth rate (1961-2017)

Dependent Variable: Y_US
Method: Least Squares
Date: 10/21/18 Time: 22:35
Sample(adjusted): 1964 2017
Included observations: 54 after adjusting endpoints
Convergence achieved after 21 iterations
Backcast: 1958 1963

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.523759	0.463474	5.445311	0.0000
AR(1)	0.608617	0.168313	3.615992	0.0007
AR(3)	0.214054	0.141912	1.508352	0.1383
MA(1)	-0.257688	0.109938	-2.343933	0.0235
MA(2)	-0.176264	0.153662	-1.147089	0.2573
MA(3)	-0.601445	0.151934	-3.958598	0.0003
MA(5)	0.460263	0.146128	3.149724	0.0029
MA(6)	-0.252384	0.151611	-1.664678	0.1028

R-squared	0.380509	Mean dependent var	3.030896
Adjusted R-squared	0.286238	S.D. dependent var	2.061144
S.E. of regression	1.741346	Akaike info criterion	4.083147
Sum squared resid	139.4851	Schwarz criterion	4.377811
Log likelihood	-102.2450	F-statistic	4.036353
Durbin-Watson stat	1.937763	Prob(F-statistic)	0.001600

Inverted AR Roots	.88	-.14 -.47i	-.14+.47i	
Inverted MA Roots	.91	.48 -.34i	.48+.34i	-.40 -.90i
	-.40+.90i	-.81		

Figure: In sample fit for the US GDP annual growth rate (1961-2017)

Figure: In sample fit for the US GDP annual growth rate (1961-2017)

2013	1.67733	1.74097	-0.06364
2014	2.56919	2.50757	0.06163
2015	2.86159	3.65593	-0.79435
2016	1.48528	1.36497	0.12031
2017	2.27334	2.64773	-0.37439

Key Concepts

1. Moving average (MA) process
2. Autoregressive (AR) process
3. Autoregressive moving average (ARMA) process
4. Invertibility and stationarity conditions
5. ACF and PACF of the AR and MA processes
6. Identification of the ARMA processes

End of Chapter 5

Christophe Hurlin