### Time Series Modeling and Analysis

#### Dionissios T. Hristopulos

School of Electrical and Computer Engineering Technical University of Crete, Chania, Greece

#### October 13, 2023





European Social Fund

Operational Programme Human Resources Development, Education and Lifelong Learning

Co-financed by Greece and the European Union





(GSLAB)

# Acknowledgment of Support

"Support for the Internationalization of Higher Education, School of Electrical and Computer Engineering, Technical University of Crete" (MIS 5150766), under the call for proposals "Support of the Internationalization of Higher Education - Technical University of Crete" (EDULLL 153).

The project is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning 2014-2020.



European Union European Social Fund

Operational Programme Human Resources Development, Education and Lifelong Learning

Co-financed by Greece and the European Union



### Goals of Time Series Analysis

- Are there trends in the data?
- Are there seasonal variations, periodicity ?
- Are there remaining temporal correlations (correlated residuals) ?
- Can the data be used to build time series models suitable for forecasting future values of the process ?
- How to estimate the parameters of a time series model?
- What is the accuracy and the reliability of forecasting models?

# **Overview** I



#### Introduction to R

- Data Importing in R
- Plotting Time Series



#### Review of fundamental concepts

- Stochastic processes (discrete and continuous)
- Expectation and autocorrelation functions
- Stationarity
- Frequency-domain analysis

#### Time series models

- Autoregressive models
- Moving Average Models
- Box-Jenkins (ARIMA) Approach
- SARIMA Models

# **Overview II**



Regarding Non-Stationarity

- Stationarity tests
- Heteroskedasticity





**Parameter Estimation** 

- Method of moments
- Method of least squares
- Maximum likelihood estimation method

# **Overview III**

#### 7

#### Forecasting methodologies

- Stationary time series
- Forecasting Non-Stationary Time Series
- Holt-Winters Model

#### Model Assessment

- Optimal model selection
- Cross validation analysis
- Analysis of Residuals

### Introduction to R

### Introduction to R

- R is a free updated version of S-Plus (developed by Bell Labs)
- R is distributed by CRAN ("Comprehensive R Archive Network")
- Linux, MACOS X and Windows versions can be downloaded and installed from the CRAN web-site.
- R has more flexible and powerful features for data handling, processing and graphical presentation than software such as Microsoft Excel.
- R uses a command-line interface (hence steeper learning curve)
- RSTUDIO is an Integrated Development Environment for R, which can be downloaded here.

### Introduction to R

- Loading data in R is not trivial. Different file types require different R functions.
- For an introduction to data importing in R see here.
- It is useful to know R since it is widely used in data analysis.
- You can do the time series analysis in Matlab® using the econometrics toolbox.
- You can also play with the Matlab® time series analysis and forecast TSAF package which offers a GUI environment.
- TSAF requires the econometrics toolbox.

# Time Series Analysis in R

### Functions available in the base P package

acf	Sample autocorrelation function. User can specify maximum lags, or a vector
	of required lags.
pacf	Sample partial autocorrelation function
arima	Fits a SARIMA model of order $(p, d, q) \times (P, D, Q)$ , with period <i>s</i> . Fitting method can be chosen from: ML Maximum Likelihood; CSS Minimising conditional sum of squares; CSS-ML Conditional sum of squares to find starting values, followed by ML.
predict	Predicts $n$ steps ahead from any fitted model, incl. time series fitted using arima
arima.sim	Simulates an ARIMA model of specified length, order $(p, d, q)$ , and specified innovations variance
tsdiag	3 standard diagnostic charts for a fitted ARIMA model: (1) plot of residuals from the model; (2) Sample ACF of residuals; (3) Ljung-Box portmanteau statistic for specified maximum number of lags
spectrum	Spectral density using one of two methods: (1) "periodogram" – based on FFT, optionally smoothed; (2) "autoregressive" – based on the spectral density of fitted AR model.

# Time Series Analysis in R

Specialized Time Series Analysis Packages								
ZOO	Functions for both regularly- and irregularly-spaced time series.							
tseries	Contains fitting and prediction for specialised time series models, e.g. GARCH (Generalised AutoRegressive Conditional Heteroscedastic) model							
cts	Continuous-time AutoRegressive models.							
dse	Dynamic Systems Estimation—tools for multivariate, time-invariant models in- cluding state-space representations.							
dlm	Bayesian and likelihood analysis of Dynamic Linear Models.							
sspir	Tools for the specification of formulas that can be used to define and fit state- space models.							
astsa	Time series analysis package used in the book of Shumway & Stoffer.							

To install packages use: install.packages("astsa")

To use the package it should be loaded: require ("astsa")

To update the packages use: update.packages(ask=FALSE)

### Time Series Datasets in the R Environment

- Several datasets are available for time series analysis with base R
- Such datasets may have been formulated into a ts (time series) class object
- These data can be recalled simply by inputting the dataset name at the command prompt.
- Example: AirPassengers

```
> AirPassengers
```

	Jan	Feb	Mar	Apr	Мау	Jun	Jul	Aug	Sep	0ct	Nov	Dec
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	209	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306
1957	315	301	356	348	355	422	465	467	404	347	305	336
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	508	461	390	432
>												

### Time Series Datasets in the R Environment

#### Example: sunspot.month Monthly Sunspot data from 1749 to 1997

> sunspot.month										
	Jan	Feb	Mar	Apr	мау	Jun	Jul	Aug	Sep	Oct
1749	58.0	62.6	70.0	55.7	85.0	83.5	94.8	66.3	75.9	75.5
1750	73.3	75.9	89.2	88.3	90.0	100.0	85.4	103.0	91.2	65.7
1751	70.0	43.5	45.3	56.4	60.7	50.7	66.3	59.8	23.5	23.2
1752	35.0	50.0	71.0	59.3	59.7	39.6	78.4	29.3	27.1	46.6
1753	44.0	32.0	45.7	38.0	36.0	31.7	22.0	39.0	28.0	25.0
1754	0.0	3.0	1.7	13.7	20.7	26.7	18.8	12.3	8.2	24.1
1755	10.2	11.2	6.8	6.5	0.0	0.0	8.6	3.2	17.8	23.7
1756	12.5	7.1	5.4	9.4	12.5	12.9	3.6	6.4	11.8	14.3
1757	14.1	21.2	26.2	30.0	38.1	12.8	25.0	51.3	39.7	32.5
1758	37.6	52.0	49.0	72.3	46.4	45.0	44.0	38.7	62.5	37.7
1759	48.3	44.0	46.8	47.0	49.0	50.0	51.0	71.3	77.2	59.7
1760	67.3	59.5	74.7	58.3	72.0	48.3	66.0	75.6	61.3	50.6
1761	70.0	91.0	80.7	71.7	107.2	99.3	94.1	91.1	100.7	88.7
1762	43.8	72.8	45.7	60.2	39.9	77.1	33.8	67.7	68.5	69.3
1763	56.5	31.9	34.2	32.9	32.7	35.8	54.2	26.5	68.1	46.3
1764	59.7	59.7	40.2	34.4	44.3	30.0	30.0	30.0	28.2	28.0
1765	24.0	26.0	25.0	22.0	20.2	20.0	27.0	29.7	16.0	14.0
1766	12.0	11.0	36.6	6.0	26.8	3.0	3.3	4.0	4.3	5.0
1767	27.4	30.0	43.0	32.9	29.8	33.3	21.9	40.8	42.7	44.1
1768	53.5	66.1	46.3	42.7	77.7	77.4	52.6	66.8	74.8	77.8
1769	73.9	64.2	64.3	96.7	73.6	94.4	118.6	120.3	148.8	158.2
1770	104.0	142.5	80.1	51.0	70.1	83.3	109.8	126.3	104.4	103.6
1771	36.0	46.2	46.7	64.9	152.7	119.5	67.7	58.5	101.4	90.0
1772	100.9	90.8	31.1	92.2	38.0	57.0	77.3	56.2	50.5	78.6
1773	54 6	29.0	51 2	32 9	41 1	28.4	27 7	12 7	29.3	26 3
								-		

- Let us consider the .csv file Recife.csv which contains monthly temperature values from Recife (Brazil).
- The data are organized in a single column.
- To read the data we use (modify the path accordingly):

```
recife=ts(read.csv("D:\\Data\\Recife_v.csv",header=FALSE),
    start=1953,frequency=12)
```

• To view the data we call the variable name defined above:

```
recife
```

```
> recife
```

```
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

1953 26.8 27.2 27.1 26.3 25.4 23.9 23.8 23.6 25.3 25.8 26.4 26.9

1954 27.1 27.5 27.4 26.4 24.8 24.3 23.4 23.4 24.6 25.4 25.6 26.7

1955 26.9 26.3 25.7 25.7 24.8 24.0 23.4 23.5 24.8 25.6 26.2 26.5

1956 26.8 26.9 26.7 26.1 26.2 24.7 23.9 23.7 24.7 25.8 26.1 26.5

1957 26.3 27.1 27.4 26.4 25.5 24.9 24.2 24.6 25.5 25.9 26.4 26.9

1958 27.1 27.1 27.4 26.4 25.5 24.7 24.3 24.4 24.8 26.2 26.3 27.0

1959 27.1 27.5 26.2 28.2 27.1 25.4 25.6 24.5 24.7 26.0 26.5 26.8

1960 26.3 26.7 26.6 25.8 25.2 25.1 23.3 23.8 25.2 25.5 26.4 26.7

1961 27.0 27.4 27.0 26.3 25.9 24.6 24.1 24.3 25.2 26.3 26.4 26.7

> I
```

To plot the data we use the command:

The resulting plot is:



Recife, Brazil Temperature Data

```
Year
```

Plotting CO<sub>2</sub> atmospheric concentration data in parts per million (ppm)<sup>a</sup>

> plot(co2, ylab = expression("Atmospheric concentration of CO"[2]), las = 1) > title(main = "CO2 data set")

• What does "las=1" do? It provides control over the label orientation:

Ias=0 produces labels placed parallel to their axes

- Ias=1 produces labels horizontally oriented
- Ias=2 produces labels which are at right-angles to the axis
- Ias=3 produces labels which are vertically oriented.

<sup>&</sup>lt;sup>a</sup>Keeling, C. D. and Whorf, T. P., Scripps Institution of Oceanography (SIO), University of California, La Jolla, California USA.



### Review of fundamental concepts

- A stochastic process is a random function that evolves in time ⇒ multiple realizations (states).
- We can consider the stochastic process as a sequence of random variables that depends on time.





### Mathematical definition

A scalar stochastic process  $X(t; \vartheta)$  is a collection of random variables  $X(t; \vartheta)$ ;  $t \in T$  defined in a probability space  $(\Omega, \mathcal{F}, P)$ , and which depends on a set of time instants T.

### Reminder: how to define the probability space

- Sample Space  $\Omega$ : Set of all possible states.
- Event Space  $\mathcal{F}$ : Set of events. Each event is a set of states of  $\Omega$ .
- Probability function P: Assigns to each event in the space  $\mathcal{F}$  a probability, i.e. a number between 0 and 1.

The expectation  $\mathbb{E}[\cdot]$  denotes an average over the entire sample space. For example,  $\mathbb{E}[X] = \int_{-\infty}^{\infty} dx \, x \, f_X(x)$ .

### What is the set T

- The set T can be one of the following:  $\mathbb{R}^1, \mathbb{R}^+ = [0, \infty),$  $\mathbb{Z} = \{\dots, -1, 0, 1, 2, \dots\}, \mathbb{Z}^+ = \{0, 1, 2, \dots\}$
- For a given state θ the stochastic process X(t; θ)—for short X(t)—is a function of time.
- When referring to specific states (realizations) we will use lower case letters, i.e. *x*(*t*).
- When we focus on a specific time *t* the stochastic process is reduced to a random variable.
- The symbol *x*(*t*) can denote either the stochastic process or an implementation of it depending on the context.

#### What is ultimately a stochastic process

A stochastic process  $X(t; \vartheta)$  is a function of time whose values are random variables for each  $t \in \mathbb{R}$ 

# What is the connection between stochastic processes and time series?

- A stochastic process is a probabilistic model.
- The term time series is used in different ways:
  - For a temporal sequence of values  $(x_{t_1}, x_{t_2}, \dots, x_{t_n})$  indexed by a discrete set of times  $\{t_i\}_{i=1}^n$ .
  - **②** For a discrete sample (e.g., measurements) of a *continuous in time* function x(t).
  - Solution For a discrete sample of a *realization* x(t) of the stochastic process  $X(t; \vartheta)$ .
- Herein, time series will refer to a discrete sample,  $X_t$  of a stochastic process  $X(t; \vartheta)$ .

### Auto-covariance and auto-correlation functions

• The auto-covariance function is defined as follows:

$$C_X(t_1, t_2) = \mathbb{E}[X(t_1; \vartheta) X(t_2; \vartheta)] - \mathbb{E}[X(t_1; \vartheta)] \mathbb{E}[X(t_2; \vartheta)]$$
  
=  $\mathbb{E}[X(t_1; \vartheta) X(t_2; \vartheta)] - \mu(t_1) \mu(t_2)$   
=  $\mathbb{E}[\{X(t_1; \vartheta) - \mu(t_1)\} \{X(t_2; \vartheta) - \mu(t_2)\}].$  (1)

- The variance is given by the covariance evaluated at the same time:  $\sigma_X^2(t) = C_X(t, t)$ .
- The (auto-) correlation function is defined as follows:

$$p_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{\sigma_X(t_1) \, \sigma_X(t_2)}.$$
(2)

- It general,  $\rho_X(t_1, t_2) \in [-1, 1]$  (based on the Cauchy-Schwartz inequality).
- Admissible covariances are non-negative (positive) definite functions.

### Example: Number of solar spots



- Sunspots: Disturbances of the solar magnetic field
- Part of the solar activity cycle with a "period" of about 11 years
- Associated with geomagnetic storms



### What questions are we concerned about?

- What is the physical mechanism—the system—the process that creates sunspots?
- Is the time series a periodic system with noise admixture?
- Is the system stochastic?
- Or is the process chaotic? (non-linear dynamic system)?
- Can we predict the number of sunspots?



Examples of time series in Bonn dataset. Sets A and B: EEG of healthy volunteer; sets C and D: iEEG for a patient during a seizure-free interval; set E: iEEG for a patient during epileptic seizure. Domingues, O. *et al.* (2019). Multifractal Analysis for Cumulant-Based Epileptic Seizure Detection in Eeg Time Series. 143–146. doi: 10.1109/ISBI.2019.8759288.

### Objectives of time series analysis

- Understanding the time-dependent structure of observations in a single series and identifying patterns (univariate analysis)
- Filling gaps
- Forecasting the future
- Generation of realizations, exploration of possible scenarios (stochastic simulation)
- Determining causal links and flow of information between different variables (multivariate analysis)

### Difficulties encountered in time series analysis

- Limited duration (small time series length)
- Limited resolution (large time step/ low frequency sampling)
- Often a single realization is available
- Lack of stationarity
- Noise admixture
- Stochastic trends

### Some definitions

- The observed quantity is the variable X.
- Observations are usually made with a uniform time step (sampling step).
- The time series consists of a set of values  $\{x_1, x_2, \ldots, x_N\}$  at *N* different times  $t_1, t_2, \ldots, t_N$ .
- We will also refer to a time series as  $\{x_t\}_{t=1}^N$  and if necessary as the vector  $\mathbf{x} = (x_1, \dots, x_N)^\top$ .
- For multivariate time series we will denote the *i*-th component by the vector x<sub>i</sub> = (x<sub>i;1</sub>,...,x<sub>i;N</sub>)<sup>⊤</sup>, i = 1,...,M.
- At any given time t the value xt of the time series can be viewed as a realization of the random variable Xt(θ) or Xt for short.
- For regular sampling (without gaps) with sampling step  $\delta t$  it holds that  $t_n = n\delta t, n = 1, 2, \dots$

### White Noise

- In the strong sense: A stochastic process ε(t; θ) is considered white noise, if it is a sequence of *independent identically distributed* (i.i.d.) zero-mean random variables ε<sub>t</sub>(θ), i.e., ε<sub>t</sub>(θ) is independent of ε<sub>t'</sub>(θ) for all t ≠ t'.
- "Weak" sense: The values of  $\varepsilon_{t_1}(\vartheta)$  and  $\varepsilon_{t_2}(\vartheta)$  are uncorrelated for every  $t_1, t_2 \neq t_1$ .

Properties of discretely-sampled white time

$$\mathbb{E}[\varepsilon(t_n)] = 0, \ n = 1, 2, \dots$$

$$\mathbb{E}[\varepsilon(t_n)\,\varepsilon(t_m)] = \sigma^2 \delta_{n,m}$$

 $\delta_{n,m}$ : Kronecker delta

$$\delta_{n,m} = 1$$
, if  $n = m$ 

$$\delta_{n,m} = 0$$
, if  $n \neq m$ .

# Expectation and autocorrelation functions

**•** Expectation: 
$$\mathbb{E}[X_t] \triangleq \int_{-\infty}^{\infty} dx \, x \, f_X(x;t).$$

- ②  $\mu_t \triangleq \mathbb{E}[X_t]$  is the deterministic part which contains the trend and seasonal component of the time series.
- 3 Fluctuation:  $X'_t = X_t \mu_t$ .

(

(Auto-)covariance function:  $C_{t,t'} \triangleq \mathbb{E}[(X_t - \mu_t)(X_{t'} - \mu_{t'})].$ 

Solution Variance: 
$$\sigma_X^2(t) \triangleq \mathbb{E}[X_t^2] - \mathbb{E}^2[X_t] = \mathbb{E}[X_t'^2]$$

(Auto-)correlation function:  $\rho_{t,t'} \triangleq \frac{C_{t,t'}}{\sigma_t \sigma_{t'}}$ .

In time series analysis the following decomposition is often used:

$$X_t = T_t + S_t + X_t',\tag{3}$$

where  $T_t$  is the trend,  $S_t$  is the seasonal component, and  $X'_t$  is the stochastic component.

### Stationarity

- Stationarity implies that the statistical properties of a time series do not change over time.
- In the weak sense (also, second-order, wide-sense), stationarity requires that the following relations hold:
  - **()** The expectation,  $\mu_t \triangleq \mathbb{E}[X_t]$  is independent of time, i.e.,  $\mu_t = m$ .
  - 2 The centered auto-covariance function,  $C_{t,t'} \triangleq \mathbb{E}[(X_t \mu_t)(X_{t'} \mu_{t'})]$  is a function purely of the time difference (time lag), i.e.,  $C_{t,t'} = C_{t-t'} = C_{t'-t}$ .



Source: Our World in Data based on the Global Carbon Project (2022) OurWorldinData org/co2-and-other-greenhouse-gas-emissions/ • CC B\*

 Foesil emissions: Foesil emissions measure the quantity of carbon dixoide (CD), emitted from the burning of foesil fuels, and directly from industrial processes such as centeria and sited production. Foesil CD, includes emissions from cool, eli, gas, faring, centeri, steel, and other industrial processes. Foesil emissions during the charge, directoration, relation vegetation.



Source: NOA4 (National Oceanic and Atmospheric Administration) via the US EPA OurWorldthData.org/natural-daasters • CC E Note: A value between -2 and -3 indicates moderate drought, -3 to -4 is severe drought, and -4 or below indicates extreme drought.



### Time series models

### Fundamentals of Time Series Modeling



### Fundamentals of Time Series Modeling

# Practical Decomposition of Time Series



# Fundamentals of Time Series Modeling

### Methodological Steps of Time Series Analysis

#### Model Identification

- Time Series plot
- Check for the existence of trend and/or seasonality
- Check for sharp changes
- Check for possible outliers
- Trend and seasonal component removal to obtain stationary residuals
- Estimation
  - MME (method of moments)
  - MLE (maximum likelihood estimation)
- Diagnostic Checking
  - Normality of error terms
  - Independence of error terms
  - Constant error variance (Homoscedasticity)

#### Forecasting

- Exponential smoothing methods
- Minimum MSE (mean-square error) forecasting
# White Noise (WN) Process: The Basic Building Block for Linear Time Series

- WN time series:  $\{\epsilon_t\}_{t=1}^N$
- Zero mean:  $\mathbb{E}[\epsilon_t] = 0$
- Variance:  $\mathbb{E}[\epsilon_t^2] = \sigma_{\epsilon}^2$
- Delta correlation:  $\mathbb{E}[\epsilon_t \epsilon_{t'}] = \sigma_{\epsilon}^2 \delta_{t,t'}$
- Gaussian white noise (GWN):  $\epsilon_t \sim \mathcal{N}(0, \sigma_{\epsilon}^2)$
- 2 Independent white noise:  $f(\epsilon_t, \epsilon_{t'}) = f(\epsilon_t) f(\epsilon_{t'}).$

IWN is a stronger condition.

#### How to measure correlations

• Definition of auto-covariance function:  $c_{t,t+\tau} = \mathbb{E} \left[ (X_{t+\tau} - \mu_{t+\tau}) (X_t - \mu_t) \right]$ 

- Definition of auto-correlation function:  $\rho_{t,t+\tau} = \frac{c_{t,t+\tau}}{\sigma_t \sigma_{t+\tau}}$
- $\tau$ : Time lag.

#### The concept of stationarity

- The basic idea of stationarity is that the probability laws governing the process do not change with time.
- A process is second-order stationary or simply weakly stationary if (i) the mean and the variance are constant and (ii) the autocovariance function depends only on the time lag τ, not on the time t.

$$\mathbb{E}[X_t] = \mu, \quad \mathbb{V}\mathrm{ar}(X_t) = \sigma^2, \quad c_\tau = \mathbb{E}\left[ (X_{t+\tau} - \mu) \ (X_t - \mu) \right].$$





- Stationary processes: Fast decline of correlations with time lag.
- Non-stationary processes: The correlations decline very slowly.

#### The Meaning of Ergodicity

- Model parameters are not known *a priori*; they should be estimated from the sample data.
- A process is called ergodic with respect to some population parameter, if this parameter can be precisely estimated from a single sample (time series).
- $X_t$  is ergodic in the mean of  $\mu = \mathbb{E}[X_t] = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T x_t$ .
- A second-order stationary process is ergodic with respect to the mean and second-order moments, if the sum of the absolute autocovariance values is finite: ∑<sub>k=1</sub><sup>∞</sup> |ρ<sub>k</sub>| < ∞.</li>
- Gaussian processes are therefore ergodic if  $\sum_{k=1}^{\infty} |\rho_k| < \infty$ .

#### Different Modes for Time Series Dynamics

- Moving Average (MA) processes
- Autoregressive (AR) processes
- Autoregressive Moving Average (ARMA) processes
- Autoregressive Integrated Moving Average (ARIMA) processes
- Autoregressive conditional heteroscedastic (ARCH) processes and GARCH (nonlinear)
- Self-Exciting Threshold AutoRegressive (SETAR) models exhibit regime switching behavior (nonlinear)
- Vector Autoregressive (VAR) processes (multivariate)
- ARIMA with exogenous variables (ARIMAX) model (multivariate)

#### Autoregressive of Order *p*: AR(p)

Linear model with memory of p steps from the past

 $X_{t} = \phi_{1} X_{t-1} + \phi_{2} X_{t-2} + \dots \phi_{p} X_{t-p} + \epsilon_{t},$ 

- The time series  $\{\epsilon_t\}$  is white noise (typically Gaussian),  $\epsilon_t \sim \mathcal{N}(0, \sigma_{\epsilon}^2)$ .
- The process  $\{\epsilon_t\}$  is called innovation process.
- Important facts:
  - The value of the TS  $X_t$  is independent of future innovations.
  - The innovations  $\epsilon_t$  are independent of past values of the TS.
- Short form of AR equation:  $X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t$ .

#### The time shift operator B

- Time shift operator:  $BX_t = X_{t-1}$
- $B^n = B B^{n-1} \implies B^n X_t = X_{t-n}.$
- Short-hand expression of AR(p) model:  $\Phi(B)X_t = \epsilon_t$ .
- $\phi(B) = 1 \phi_1 B \phi_2 B^2 \dots \phi_p B^p$ .
- $\phi(B)$  is the characteristic polynomial.

#### Stationarity Condition for AR Models

• AR(p) equation based on time shift operator:  $\phi(B)X_t = \epsilon_t$ , where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots \phi_p B^p$$

is the characteristic polynomial.

• **Theorem:** The time series  $X_t$  which satisfies  $\phi(B)X_t = \epsilon_t$  is stationary if and only if the roots of the characteristic polynomial

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots \phi_p z^p$$
, where  $z \in \mathbb{C}$ 

are outside the unit circle in the complex plane.

• Corollary: The roots of the reciprocal characteristic polynomial

$$\phi'(z) = z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \dots - \phi_p$$

should respectively lie inside the unit circle.

See Box, Jenkins, Reinsel, Ljung, (2016). *Time Series Analysis: Forecasting and Control*, Wiley.

(GSLA

#### Autoregressive process of order 1, AR(1)

- Innovation equation:  $X_t = \phi X_{t-1} + \epsilon_t$ .
- Characteristic polynomial:  $\phi(z) = 1 \phi z$ , root:  $z = 1/\phi$ .
- Stationarity condition:  $|\phi| < 1$ .
- Expectation:  $\mathbb{E}[X_t] = \phi \mathbb{E}[X_{t-1}] + \mathbb{E}[\epsilon_t] \implies \mathbb{E}[X_t] = \phi \mathbb{E}[X_{t-1}]$ . Due to stationarity it holds that  $\mathbb{E}[X_t] = \mathbb{E}[X_{t-1}] \implies \mathbb{E}[X_t] = 0$ .
- Variance:  $X_t^2 = \phi^2 X_{t-1}^2 + \epsilon_t^2 + 2\phi X_{t-1} \epsilon_t$ .
- Since  $\mathbb{E}[X_{t-1} \epsilon_t] = 0$  and  $\mathbb{E}[X_t] = 0$  it follows that:

$$\begin{split} \mathbb{E}[X_t^2] = &\phi^2 \,\mathbb{E}[X_{t-1}^2] + \mathbb{E}[\epsilon_t^2] \implies \mathbb{V}\mathrm{ar}[X_t] = \phi^2 \,\mathbb{V}\mathrm{ar}[X_{t-1}] + \sigma_\epsilon^2 \\ \implies &\sigma_X^2(1-\phi^2) = \sigma_\epsilon^2 \implies \sigma_X^2 = \frac{\sigma_\epsilon^2}{1-\phi^2} \,. \end{split}$$

#### AR(1) Auto-covariance function

• Auto-covariance at lag 1:  $c_1 = \mathbb{E}[X_t X_{t-1}].$ 

 $c_1 = \mathbb{E}[(\phi X_{t-1} + \epsilon_t) X_{t-1}] = \phi \mathbb{E}[X_{t-1}^2] + \mathbb{E}[\epsilon_t X_{t-1}] = \phi \mathbb{E}[X_{t-1}^2] = \phi \sigma_X^2.$ 

• Auto-covariance at lag  $n \ge 1$ :  $c_n = \mathbb{E}[X_t X_{t-n}]$ 

$$\mathcal{L}_{n} = \mathbb{E}[(\phi X_{t-1} + \epsilon_{t}) X_{t-n}] = \mathbb{E}\left[\left(\phi^{n} X_{t-n} + \sum_{k=1}^{n} \phi^{k-1} \epsilon_{t-k+1}\right) X_{t-n}\right]$$
$$= \phi^{n} \mathbb{E}[X_{t-n}^{2}] + \sum_{k=1}^{n} \phi^{k-1} \mathbb{E}[\epsilon_{t-k+1} X_{t-n}] = \phi^{n} \sigma_{X}^{2}.$$

- In the above, we used repeatedly the Eq. AR(1):  $X_{t-k} = \phi X_{t-k-1} + \epsilon_{t-k}$ in order to express  $X_t$  as a sum of  $X_{t-n}$  and the innovations  $\epsilon_{t-k+1}$ , k = 1, ..., n.
- We also used the independence of  $X_{t-n}$  on future values,  $X_{t-n+k}$  (where k > 0), of the innovation process.

AR(1) Auto-covariance function

•  $c_{\tau} = \phi^{\tau} \sigma_X^2 \implies \rho_{\tau} = \frac{c_{\tau}}{\sigma_X^2} = \phi^{\tau}$  for  $\tau \ge 0$ . For  $\tau < 0$  by symmetry it holds that  $\rho_{\tau} = \phi^{-\tau}$ .

• The above implies exponential dependence of the ACF on the time lag:

$$\rho_{\tau} = \phi^{|\tau|} = \begin{cases} \exp(\tau \log \phi) = \exp(|\tau| \log |\phi|), & 0 < \phi < 1, \\ \\ (-1)^{\tau} \exp(|\tau| \log |\phi|), & -1 < \phi < 0, \end{cases}$$

- Since  $|\phi| < 1$  it holds that  $\log |\phi| < 0$ .
- Hence, if we define the correlation time  $\tau_c = -1/\log(|\phi|) > 0$  the ACF declines exponentially as

$$\rho_{\tau} = \left(\operatorname{sign}(\phi)\right)^{|\tau|} \exp\left(-|\tau|/\tau_c\right).$$









# Autocorrelation of AR(1) Model

$$X_t = 0.8X_{t-1} + \varepsilon_t$$



(GSLAB)

# Autocorrelation for AR(1)

 $X_t = -0.8X_{t-1} + \varepsilon_t$ 



(GSLAB)

#### Yule-Walker equations for general AR(p) Models

For stationary AR(p) models, the coefficients and the correlation functions are related via the system of Yule-Walker equations:

1	$ ho_1$	$ ho_2$		$\rho_{p-1}$		$\phi_1$		$\rho_1$	
$ ho_1$	1	$ ho_1$	• • •	$ ho_{p-2}$		$\phi_2$		$\rho_2$	
$\rho_2$	$ ho_1$	1		$\rho_{p-3}$		$\phi_3$	=	$ ho_3$	,
		• • •							
$\rho_{p-1}$	$ ho_{p-2}$	• • •	$ ho_1$	1	l	$\phi_p$		$ ho_p$	

If the correlations are estimated from the ACF function, the Yule-Walker system can be solved to obtain estimates of the model coefficients.

#### The Concept of Partial Autocorrelation Function (PACF)

- In an AR(1) time series the value  $X_t$  is based on the preceding value,  $X_{t-1}$ , and the innovation,  $\epsilon_t$ , which is independent of the past values  $\{X_{t'}, t' = 1, 2, \dots, t-1\}$  of the process.
- $X_{t+1}$  is correlated with  $X_t$  but also with  $X_{t-1}$  due to the correlation of the latter with  $X_t$ .
- Likewise,  $X_{t+1}$  is also correlated with  $X_{t-1}$  etc. Recall that  $\rho_{\tau} = \phi^{\tau}$ .
- How to test for direct dependence between  $X_{t+1}$  and  $X_{t-1}$ ,  $X_{t-2}$ ,...? For the AR(1) model we know that such dependence does not exist.
- In more general cases, the answer to this question is provided by the partial autocorrelation function:

• 
$$\pi_{\tau} = \mathbb{C}$$
orr  $[X_{t+\tau}, X_t \mid X_{t+1} = x_{t+1}, X_{t+2} = x_{t+2}, \dots, X_{t+\tau-1} = x_{t+\tau-1}]$ 

#### Calculation of PACF

- The ACF values are linked to the PACF values, e.g.,  $\pi_1 = \rho_1$ ,  $\pi_2 = (\rho_2 \rho_1^2) / (1 \rho_1^2)$ . However, these relations are quite complex for larger lags.
- In AR(p) models,  $X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t$ , the following two relations hold: (R1)  $\pi_p = \phi_p$  and (R2)  $\pi_\tau = 0$  for  $\tau > p$ .
- (R1) holds because  $\phi_p$  completely determines the dependence of  $X_t$  on  $X_{t-p}$  when the intermediate values are fixed.
- (R2) holds because  $X_t$  does not depend directly on  $X_{t-\tau}$  for  $\tau > p$ .
- These properties are used in the R programming environment for PACF calculation:
  - Settimates  $\hat{\pi}_k$ , k = 1, ..., p are obtained by fitting models AR(k), k = 1, ..., p to the time series.
  - 2 Then  $\hat{\pi}_k = \hat{\phi}_{k,k}$  where  $\hat{\phi}_{k,k}$  is the estimate for the k-th lag coefficient  $\tau = k$  in the model AR(k).

#### PACF Estimation using AR(k) Models

Fit the data to AR(1) model and use Yule-Walker equations

$$k = 1, \ \phi_{1,1} = \rho_1 = \pi_1$$

Fit the data to AR(k) model, and use Yule-Walker equations

$$k = 2, \ \phi_{2,2} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}.$$

$$\phi_{k,k} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-4} & \rho_3 \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-4} & \rho_{k-3} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-4} & \rho_{k-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{vmatrix}}$$

#### What happens if the AR(p) model contains a constant term?

- Consider the innovation equation:  $X_t = \phi_0 + \phi X_{t-1} + \epsilon_t$ .
- Then,  $\mathbb{E}[X_t] = \phi_0 + \phi \mathbb{E}[X_{t-1}] \Rightarrow \mu = \phi_0 + \phi \mu$ , where  $\mu = \mathbb{E}[X_t]$ .
- Hence,

$$\mu = \frac{\phi_0}{1 - \phi}.$$

- Therefore, it is easy to construct an AR(1) model with a finite expectation.
- The same approach (i.e., using a constant offset), can be used for general AR(p) models (p > 1) as well.

$$\mu = \frac{\phi_0}{1 - \sum_{i=1}^p \phi_i}.$$

#### Moving Average of Order One MA(1)

- Innovation equation:  $X_t = \epsilon_t \theta \epsilon_{t-1}, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_{\epsilon}^2).$
- Equivalent form:  $X_t = \theta(B) \epsilon_t$ , where  $B[\epsilon_t] = \epsilon_{t-1}$  and  $\theta(B) = 1 \theta B$ .
- Characteristic polynomial:  $\theta(z) = 1 \theta z$  with root  $z = 1/\theta$ .
- Invertibility condition:  $|\theta| < 1 \implies \epsilon_t = \theta^{-1}(B)[X_t].$
- An MA model is invertible if it is equivalent to a converging infinite order AR model. Convergence means that AR coefficients decrease to zero as the time lags increase.
- Expectation:  $\mathbb{E}[X_t] = \mathbb{E}[\epsilon_t] \theta \mathbb{E}[\epsilon_{t-1}] \implies \mathbb{E}[X_t] = 0.$
- Variance:  $X_t^2 = \epsilon_t^2 + \theta^2 \epsilon_{t-1}^2 2\theta \epsilon_{t-1} \epsilon_t$ . This implies

$$\mathbb{E}[X_t^2] = \mathbb{E}[\epsilon_t^2] + \theta^2 \mathbb{E}[\epsilon_{t-1}^2] - 2\theta \mathbb{E}[\epsilon_{t-1}\overline{\epsilon_t}]^{\bullet} \stackrel{0}{\Longrightarrow} \sigma_X^2 = (1+\theta^2) \sigma_{\epsilon}^2.$$

#### MA(1) Autocorrelation Function

• Covariance at lag one:  $c_1 = \mathbb{E}[X_t X_{t-1}].$ 

$$c_{1} = \mathbb{E}[(\epsilon_{t} - \theta \epsilon_{t-1}) (\epsilon_{t-1} - \theta \epsilon_{t-2})] = -\theta \mathbb{E}[\epsilon_{t-1}^{2}] + \mathbb{E}[\theta^{2} \epsilon_{t-1} \epsilon_{t-2}]^{+0}$$
$$- \mathbb{E}[\theta \epsilon_{t} \epsilon_{t-2}]^{+0} + \mathbb{E}[\epsilon_{t} \epsilon_{t-1}]^{+0} = -\theta \sigma_{\epsilon}^{2}.$$

- Hence  $\rho_1 = \frac{-\theta}{1+\theta^2}$ .
- Covariance at lag n > 1:  $c_n = \mathbb{E}[X_t X_{t-n}]$

$$c_{n} = \mathbb{E}[(\epsilon_{t} - \theta \epsilon_{t-1}) (\epsilon_{t-n} - \theta \epsilon_{t-n-1})] = -\theta \mathbb{E}[\epsilon_{t-n} \epsilon_{t-1}]^{0}$$
$$+ \mathbb{E}[\theta^{2} \epsilon_{t-1} \epsilon_{t-n-1}]^{0} - \mathbb{E}[\theta \epsilon_{t} \epsilon_{t-n-1}]^{0} + \mathbb{E}[\epsilon_{t} \epsilon_{t-n}]^{0} = 0.$$

- Thus in the MA(1) model auto-covariance and auto-correlation vanish for τ > 1.
- This property is generalized to MA(q) models.

Ω

#### Higher-Order Moving Average Models MA(q)

Innovation equation:  $X_t = \epsilon_t - \sum_{i=1}^q \theta_i \epsilon_{t-i}$ .

Properties of MA(q) model:

$$\sigma_X^2 = \sigma_\epsilon^2 \left( 1 + \theta_1^2 + \dots + \theta_q^2 \right),$$

$$c_\tau = \sigma_\epsilon^2 \left( -\theta_\tau + \sum_{j=1}^{q-|\tau|} \theta_j \theta_{j+|\tau|} \right), \ \tau = 1, \dots, q_q$$

$$c_\tau = 0, \ \tau > q$$

#### Covariance function of MA(1) model

- Consider an MA(1) model with  $\theta$ . Then  $\rho_1 = \frac{-\theta}{1+\theta^2}$ .
- The same ACF value is obtained if  $\theta \to 1/\theta$ .
- However, between the two models only that with  $|\theta| < 1$  is invertible.



#### Example:

The time series  $X_t = \epsilon_t - 0.4\epsilon_{t-1}$  and  $X_t = \epsilon_t - 2.5\epsilon_{t-1}$  share the same ACF:  $\rho_1 = -1/2.9$  and  $\rho_n = 0, n \ge 2$ .

However, only  $X_t = \epsilon_t - 0.4\epsilon_{t-1}$  is invertible.

#### PACF of MA(1) Model

• 
$$\pi_1 = \rho_1 = \frac{-\theta}{1+\theta^2}$$
  
•  $\pi_2 = -\frac{\rho_1^2}{1-\rho_1^2} = \frac{-\theta^2}{1+\theta^2+\theta^4}$   
•  $\pi_3 = \frac{\rho_1^3}{1-2\rho_1^2} = \frac{-\theta^3}{1+\theta^2+\theta^4+\theta^6}$   
•  $\pi_\tau = -\frac{\theta^\tau (1-\theta^2)}{1-\theta^{2(\tau+1)}}$ 



#### The basic steps of the the Box-Jenkins methodology

**ARMA** model :  $X_t = \phi_0 + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots + \theta_q \epsilon_{t-q}$ .

- The Box-Jenkins approach assumes that the series is stationary. The series is differenced (if necessary) to achieve stationarity.
- Seasonal differences can be applied to remove cyclical behavior.
- Identification of a tentative ARMA or ARIMA model for the stationary component.
- Estimation of the model (maximum likelihood is the preferred method).
- Diagnostic checking of the residuals (goal: normally distributed, uncorrelated, constant mean and variance)
- If the model is found inadequate (residual checks fail) return to step 2.
- The optimal ARMA (ARIMA) model is used for forecasting and control.

# Estimating ARMA(p,q) orders

To determine p and q use the following table.

	MA(q)	AR(p)	ARMA(p,q)
ACF	Cuts off after $q$	Tails off	Tails off
PACF	Tails off	Cuts off after p	Tails off

**Note:** Usually  $p + q \le 4$ .

#### Using the ACF for Model Identification

ACF shape	Suggested Model
Exponential decay to zero	AR(p) model. PACF plot indi-
	cates correct <i>p</i>
Decay to zero with alternating	AR(p) model. PACF plot indi-
$\pm$ values	cates correct p
Mostly zero values with a few	MA(q) model; q = maximum lag
spikes	with $ACF \neq 0$
Decline starting after a few lags	ARMA model
All ACF values zero or close to	Random data
zero	
High values occurring at fixed	Seasonal AR model
intervals	
Very slow decay	Non-stationary series

Source: Wikipedia

#### The Box-Jenkins Approach

#### **Data Preparation:**

- 1) Plot data and examine for stationarity
- 2) Examine ACF for stationarity
- 3) If not stationary, take first differences
- 4) If variance appears non-constant, take logarithm before first differencing
- 5) Examine the ACF after transformations to determine if the series is stationary

#### **Model Identification and Estimation:**

- 1) Examine the ACF and PACF's to get some ideas about ARIMA(p,d,q) models to estimate.
- 2) Estimate these models
- Examine the parameter estimates, the SBC statistic and test of white noise for the residuals.

#### Forecasting:

- 1) Use the best model to construct forecasts
- 2) Graph your forecasts against actual values
- 3) Calculate the Mean Squared Error for the forecasts

#### Estimating ARIMA(p,d,q) orders

- The order d in ARIMA(p,d,q) stands for the number of times the series has been differenced to achieve stationarity.
- The order p in ARIMA(p,d,q) measures the order of the autoregressive component.
- If the time-series has an autoregressive order of 1, called AR(1), then only the first partial autocorrelation coefficient should be significant.
- If it has an AR(2) structure, then only the first and second partial autocorrelation PACF coefficients should be significant. (Note, that they could be positive and/or negative; what matters is the statistical significance.)
- Generally, the partial autocorrelation function PACF will have significant correlations up to lag p, and will quickly drop to near zero values beyond lag p.

#### Estimating ARIMA(p,d,q) orders

- The order q measures the order of the moving average component.
- To get an idea of what orders to consider, we examine the autocorrelation function (ACF).
- If the time-series is a moving average of order 1, called an MA(1), only one ACF coefficient at lag 1 is significant. An MA(1) process has a memory of only one step.
- If the time-series is an MA(2), only two ACF coefficients, at lags 1 and 2 are significant. An MA(2) process has a memory of only two steps.
- Generally, for a time-series that is a an MA(q), the ACF has significant correlations up to lag q, and quickly drops to near zero values after lag q.

#### Selection of the Optimal Model

- Schwartz Bayesian Criterion (SBC): SBC =  $-2 \log L + k \log N$
- k = p + q + 1 is the number of model parameters estimated
- N is the sample size
- *L* is the likelihood of the model which essentially depends on the sum of the squared residuals
- The model with the lowest SBC measure is considered "optimal"
- SBC can be positive or negative.
- Other selection criteria (e.g., AIC and modifications) are possible
- Important: A model's goodness of fit to the data should not be solely judged by the selection criterion: it should also be checked that the residuals are consistent with the white noise model.

- ARIMA models involve an autoregressive term and a moving average term.
- The integration element, "I", refers to the differencing operation used to remove potential trends.
- ARIMA models cannot handle data with seasonal behavior.
- However, ARIMA can be used with seasonal data if (i) the seasonal component is modeled and removed (parametric approach) or (ii) if the seasonal component is removed non-parametrically (e.g, by differencing).
- SARIMA models incorporate by construction the ability to handle seasonal data.

- SARIMA stands for Seasonal Autoregressive Integrated Moving Average model or Seasonal ARIMA.
- **Definition of SARIMA model** (Box & Jenkins, 1970):

 $\Phi_P(B^s) \phi_p(B) \nabla^D_s \nabla^d X_t = c_0 + \Theta_Q(B^s) \theta_q(B) \epsilon_t$ 

- $\bigcirc$   $c_0$ : Offset (constant term).
- S: Order (period) of the seasonal component.
- Order of ordinary (short-term) differencing.
- Ordinary differencing:  $\nabla^d = (1 B)^d$ .
- D: Order of seasonal differencing.
- Seasonal differencing:  $\nabla_s^D = (1 B^s)^D$ .

#### $\Phi_P(B^s) \, \phi_p(B) \, \nabla^D_s \, \nabla^d X_t = c_0 + \Theta_Q(B^s) \, \theta_q(B) \, \epsilon_t$

• The SARIMA model can be viewed as an  $ARIMA(p, d, q) \times (P, D, Q)_s$ .

(2)  $\phi_p(B)$  is the ordinary (short-term) autoregressive operator (characteristic polynomial):

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p.$$

•  $\Phi_P(B^s)$  is the seasonal autoregressive operator:

$$\Phi_P(B) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}.$$

•  $\theta_q(B)$  is the ordinary moving average operator:

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q.$$

•  $\Theta_Q(B^s)$  is the seasonal moving average operator:

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_Q B^{Qs}$$

#### An example of SARIMA model

- Consider the model  $ARIMA(1,1,1) \times (0,1,1)_{12}$
- The general equation is

 $\Phi_P(B^s) \phi_p(B) \nabla^D_s \nabla^d X_t = c_0 + \Theta_Q(B^s) \theta_q(B) \epsilon_t$ 

• Setting s = 12, p = 1, d = 1, q = 1, P = 0, D = 1, Q = 1 we obtain

$$(1 - \phi B) (1 - B^{12})(1 - B)X_t = (1 + \Theta B^{12})(1 + \theta B)\epsilon_t.$$

• Expanding both sides of the equation we get

$$\left[1 - (\phi + 1)B + \phi B^2 - B^{12} + (\phi + 1)B^{13} - \phi B^{14}\right]X_t = \left(1 + \theta B + \Theta B^{12} + \theta \Theta B^{13}\right)\epsilon_t$$

#### or, equivalently

$$X_t - (\phi + 1)X_{t-1} + \phi X_{t-2} - X_{t-12} + (\phi + 1)X_{t-13} - \phi X_{t-14}$$
  
=  $\epsilon_t + \theta \epsilon_{t-1} + \Theta \epsilon_{t-12} + \theta \Theta \epsilon_{t-13}$ .
## **Regarding Non-Stationarity**

## Brief Introduction to Non-Stationarity

#### Some types of Non-Stationarity

- *Empirical advice:* The values of stationary series tend to return to a long-term mean and the variance is stable (independent of time).
- Non-stationarity can arise due to trends, cyclical behavior and random walks or combinations of the above.
- Random walk process:  $X_t = X_{t-1} + \epsilon_t$ .
- Random walk with drift:  $X_t = c_0 + X_{t-1} + \epsilon_t$ .
- The random walk process has a characteristic polynomial with a unit root.
- Note that if  $X_t$  is random walk, then the difference  $Y_t = X_t X_{t-1}$  is stationary.
- The random walk is a discrete version of the continuous-time Wiener process  $W(t; \vartheta) = W(0; \vartheta) + \epsilon(t; \vartheta), W(t + t'; \vartheta) W(t; \vartheta) \sim \mathcal{N}(0, \sigma^2 t').$
- Wiener process properties:  $\mathbb{E}[W(t; \vartheta)] = 0$ , variance  $\mathbb{V}ar[W(t; \vartheta)] = \sigma^2 t$ , and covariance  $C(t_1, t_2) = \sigma^2 t_{\min}$  where  $t_{\min} = \min(t_1, t_2)$ .

#### Augmented Dickey-Fuller (ADF) Test

- ADF tests for the presence of a unit root in an AR model of the data.
- First, the simple Dickey-Fuller test assumes an AR(1) model  $X_t = \phi X_{t-1} + \epsilon_t$ .
- This can be written as  $\Delta X_t = (\phi 1)X_{t-1} + \epsilon_t \triangleq \delta X_{t-1} + \epsilon_t$ ,  $(\delta = \phi 1)$ .
- The null hypothesis is that  $\delta = 0$  (i.e., unit root and non-stationarity), while the alternative is  $\delta < 0$ .
- The test also works for  $X_t = c_0 + c_1 t + \phi X_{t-1} + \epsilon_t$ .
- DF tests have low statistical power, i.e., they cannot distinguish between  $\delta = 0$  and  $\delta \approx 0$ .
- The Augmented Dickey Fuller test uses an AR(p) model of order *p* determined from the data.

#### More on Unit Root tests

- The Phillips-Perron is a unit-root test, which, instead of introducing higher-order lags, provides a non-parametric modification of the Dickey-Fuller t-test.
- The ADF and PP tests are asymptotically equivalent but differ in terms of small-sample performance.
- Unit root tests fail to consistently distinguish between highly persistent stationary and nonstationary processes.
- Tests using models that include a constant term and trend have less resolving power than tests that include only a constant term.
- It is possible to define more efficient tests.

#### Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

- The null hypothesis is that the series is stationary around a deterministic trend (i.e., trend stationary) while the alternative is the presence of a unit root (non-stationary).
- Trend-stationary processes revert to the mean ⇒ shocks have transient effect.
- Unit-root processes (processes with stochastic trends) => shocks have permanent effects.
- KPSS finds a unit root but ADF does not: the series can be made stationary by differencing.
- ADF finds a unit root but KPSS does not: the series can be viewed as *trend stationary* and requires detrending.

#### Impact of shocks



#### Heteroskedasticity

- Heteroskedasticity: the variance of the time series changes over time.
- Statistical analyses based on ordinary least squares (OLS) regression models often assume homoskedasticity.



 Engle's ARCH test: Null hypothesis → A series of residuals exhibits no conditional heteroscedasticity (ARCH effects). Alternative hypothesis → the series is described by an ARCH(p) model.

## **Nonlinear Transforms**

#### Transforms for Heteroskedastic Time Series

- Heteroskedastic time series exhibit non-constant variance in time. The variance depends on the (local) mean value of the process (proportional effect).
- Let us assume that the variance of  $Y_t$  depends non-linearly on the expectation:  $\mathbb{V}ar[Y_t] \triangleq h(\mu_t)$ .
- We seek a nonlinear transform  $g(\cdot)$  such that the time series  $X_t \triangleq g(Y_t)$  is homoskedastic, i.e.,  $\mathbb{V}ar[X_t] = \sigma^2$ .
- We can use the delta method which expands functions of a random variable into Taylor series around the mean.
- The second-order Taylor expansion is:

 $X_t = g(Y_t) \approx g(\mu_t) + g'(\mu_t) \left(Y_t - \mu_t\right).$ 

• Hence the variance of  $X_t$  becomes  $\mathbb{Var}[X_t] \approx (g'(\mu_t))^2 \mathbb{Var}[Y_t]$ .

#### Transforms for Heteroskedastic Time Series

Based on the previous slide it holds that X<sub>t</sub> ≜ g(Y<sub>t</sub>),

 $\mathbb{V}ar[Y_t] = h(\mu_t), \quad \mathbb{V}ar[X_t] \approx \left(g'(\mu_t)\right)^2 \mathbb{V}ar[Y_t].$ 

• Hence, since  $\mathbb{V}ar[X_t] = \sigma^2$  we obtain:

$$\begin{split} h(\mu_t) &\approx \frac{\mathbb{V}\mathrm{ar}[X_t]}{(g'(\mu_t))^2} = \frac{\sigma^2}{(g'(\mu_t))^2} \\ \implies g'(\mu_t) = \frac{\sigma}{\sqrt{h(\mu_t)}} \implies g(\mu_t) = \int_0^{\mu_t} \frac{\sigma}{\sqrt{h(u)}} \, du \, . \end{split}$$

Special cases of heteroskedasticity correctionVariance  $\mathbb{V}ar[Y_t] \triangleq h(\mu_t)$  $c \mu_t^4$  $c \mu_t^3$  $c \mu_t^2$  $c \mu_t$ Stabilizing transform:  $X_t \triangleq g(Y_t)$  $Y_t^{-1}$  $Y_t^{-1/2}$  $\log Y_t$  $Y_t^{1/2}$ 

#### **Box-Cox Transforms**

A nonlinear transform commonly used for time series with positive values is the so-called Box-Cox transform:

$$X_t = \left\{ \begin{array}{ll} \frac{Y_t^{\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \ln Y_t, & \lambda = 0. \end{array} \right\}$$

The transform for  $\lambda = 0$  is obtained using Taylor series expansion:

$$\frac{Y_t^{\lambda} - 1}{\lambda} = \frac{\exp\left(\lambda \ln Y_t\right) - 1}{\lambda} = \frac{1 + \lambda \ln Y_t + O(\lambda^2) - 1}{\lambda} = \ln Y_t + O(\lambda).$$

 $O(\lambda)$  denotes the most important term near  $\lambda \approx 0$ . Based on the above, we obtain the following:

$$\lim_{\lambda \to 0} \frac{Y_t^{\lambda} - 1}{\lambda} = \ln Y_t.$$

#### **Box-Cox Transforms**

The modified Box-Cox transform if  $Y_t$  also takes non-positive values:

$$X_t = \begin{cases} \frac{(Y_t - y_0)^{\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \ln(Y_t - y_0), & \lambda = 0. \end{cases}$$

- In this case there are two transformation parameters:  $y_0, \lambda$ .
- In general,  $y_0 < y_{\min} = \min(y_1, ..., y_N)$ , so that  $Y_t y_0 > 0$ .
- The parameters  $\lambda, y_0$  are estimated by maximizing the likelihood of the data.
- A new transform method based on deformed logarithmic functions is well-defined even for y < 0 values. (Kaniadakis Functions beyond Statistical Mechanics: Weakest-Link Scaling, Power-Law Tails, and Modified Lognormal Distribution.)

## Intermittent processes: Censoring and Nonlinear Transform

Intermittent Model using Nonlinear Transform of AR(1) process

Let X<sub>t</sub> represent an AR(1) process given by

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

• The censored process  $X_t^c = h(X_t)$ , where  $h(\cdot)$  is the unit step function, satisfies

$$X_t^c = \begin{cases} X_t, & \text{for } X_t > 0, \\ 0, & \text{for } X_t \le 0. \end{cases}$$

• Generate a skewed, intermittent process using a nonlinear, monotonic transformation  $G(\cdot)$ :

$$Y_t = G(X_t^c) - G(0)$$

where  $G(\cdot)$  is the inverse Box-Cox transform  $G(z) = (1 + \lambda z)^{1/\lambda}$ ,  $\lambda \ge 0$ , or the modified exponential transform  $G(z) = (\sqrt{1 + \kappa^2 z^2} + \kappa z)^{1/\kappa}$ ,  $\kappa \ge 0$ .

## Intermittent processes: Censoring and Nonlinear Transform

#### Censored AR(1) process



(GSLAB)

# Intermittent processes: Censoring and Nonlinear Transform

#### Intermittent process—transformation of cenored AR(1)



(GSLAB)

## **Parameter Estimation**

#### What are statistical moments?

Ensemble-based definition:

$$\mu_X^p(t) \triangleq \mathbb{E}[X_t^p], \ p \in \mathbb{N} \quad \sigma_t^2 \triangleq \mathbb{E}[X_t^2] - \mathbb{E}^2[X_t].$$

Sample based definition ( $x_i \triangleq x(t_i)$ ):

$$\overline{\hat{\mu}_X^p} \triangleq \frac{1}{N} \sum_{i=1}^N x_i^p \quad \overline{\sigma_X^2} \triangleq \frac{1}{N-1} \sum_{i=1}^N \left( x_i - \overline{x} \right)^2, \text{ where } \overline{x} = \overline{\mu_X}.$$

Ensemble-based definition (two-point moments):

$$c_X(t,t+\tau) \triangleq \mathbb{E}[X_{t+\tau} X_t] - \mathbb{E}[X_{t+\tau}] \mathbb{E}[X_t], \ \tau \in \mathsf{R}.$$

Sample-based definition (two-point moments):

$$\hat{c}_X(\tau) \triangleq \frac{1}{N - |\tau|} \sum_{i=1}^{N - |\tau|} (x_{i+\tau} - \overline{x}) (x_i - \overline{x}) .$$

D. Hristopulos: dchristopoulos@tuc.gr

#### How does the method of moments work?

- The ensemble moments depend on the model's parameters.
- The sample moments are fully determined from the data.
- MoM is based on equating ensemble and sample moments.
- In order for this to work, some form of stationarity and ergodicity are typically required (especially if only one realization is available).
- MoM formulates and solves a moment equation system that comprises as many equations as there are model parameters.

#### Example: AR(1) Model

- Let us consider  $X_t = \phi X_{t-1} + \epsilon_t$ , where  $|\phi| < 1$ .
- The ACF (ensemble-based) at lag one is given by  $\rho_1 = \phi$ .
- Since the AR(1) system only has one parameter, the MoM reduces to

 $\hat{\rho_1} = \rho_1(\phi) = \phi.$ 

- The solution of this system for  $\phi$  is  $\hat{\phi} = \hat{\rho_1}$ .
- Determining the innovation variance:

$$\hat{\sigma}_{\epsilon}^2 = s_X^2 \left( 1 - \hat{\phi}^2 \right) = s_X^2 \left( 1 - \hat{\rho}_1^2 \right)$$

#### Example: AR(2) Model

- Let us consider  $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$ .
- Unknown parameters:  $\phi_1, \phi_2$ .
- The ACF (ensemble-based) at lags one and two is given from the Yule-Walker equations:

$$\rho_1 = \phi_1 + \phi_2 \rho_2, \ \rho_2 = \phi_2 + \phi_1 \rho_1.$$

• Solving the Yule-Walker system for  $\phi_1, \phi_2$  leads to

$$\phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}, \ \phi_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}.$$

• The MoM estimates  $\hat{\phi}_1$  and  $\hat{\phi}_2$  are obtained by replacing  $\rho_1, \rho_2$  in the above equations with the sample-based estimates  $\hat{\rho}_1, \hat{\rho}_2$ .

#### MoM with MA(1)

• MA(1) Model:  $X_t = \epsilon_t - \theta \epsilon_{t-1}$ .

• Lag-1 ACF: 
$$\rho_1 = -\theta/(1+\theta^2)$$
.

- $\rho_1\theta^2 + \theta + \rho_1 = 0.$
- Solution of quadratic MoM equation:

$$\hat{\theta} = \frac{-1 + \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1} \quad \hat{\theta} = \frac{-1 - \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1}$$

- Problems:
  - Nonlinear moment equation.
  - 2 Two solutions for  $|\hat{\rho}_1| < 0.5$ , only one  $\implies$  invertible model.
  - 3 No invertible solution for  $\hat{\rho}_1 = \pm 0.5$ .
  - No real-valued solutions for  $|\hat{\rho}_1| > 0.5$ .

#### What about higher-order MA models?

- MoM equations for MA(q) models are nonlinear.
- Many solutions, only one invertible model.
- There are no-real valued solutions of the MoM system if the ACF estimates take certain values not allowed by the model.
- Overall, MoM is not the best method for estimating parameters of MA(q) models.

#### Estimating the Innovation variance

- The sample variance  $s_X^2 = \frac{1}{N-1} \sum_{t=1}^N (x_t \overline{x})^2$  is estimated from the time series data.
- The innovation variance is linked to the time series variance by means of

$$\sigma_X^2 = \sigma_\epsilon^2 \, g(\boldsymbol{\theta}),$$

where  $\theta$  is the vector of model parameters and  $g(\cdot)$  is a function that depends on the model.

Hence, in the MoM method the innovation variance is given by

$$\hat{\sigma}_{\epsilon}^2 = \frac{s_X^2}{g(\hat{\theta})} \,,$$

where  $\hat{\theta}$  is the estimate of the other (besides the innovation variance) model parameters (as discussed above).

#### Examples: Innovation variance for certain models

AR(p) model

$$\hat{\sigma}_{\epsilon}^2 = s_X^2 \, \left(1 - \hat{\phi}_1 \hat{
ho}_1 - \hat{\phi}_2 \hat{
ho}_2 - \ldots - \hat{\phi}_p \hat{
ho}_p 
ight)$$

MA(q) models

$$\hat{\sigma}_{\epsilon}^2 = \frac{s_X^2}{1 + \hat{\theta}_1^2 + \hat{\theta}_2^2 + \dots \hat{\theta}_p^2}$$

ARMA(1,1) model

$$\hat{\sigma}_{\epsilon}^2 = \frac{s_X^2 \left(1 - \hat{\phi}^2\right)}{1 - 2\hat{\phi}\hat{\theta} + \hat{\theta}^2} \,. \label{eq:sigma_eq}$$

#### Advantages of MLE

- Efficient (low variance estimates).
- Often the Gaussian assumption is reasonable.
- Even if X<sub>t</sub> does not follow the Gaussian distribution, the asymptotic distribution of the MLE estimates is the same as in the Gaussian case.

#### Disadvantages of MLE

- Difficult optimization problem.
- Need to choose a "good" starting point which is often determined by using other estimators.

Stochastic Processes with Special Reference to Methods and Applications, M.S. Bartlett

## Ordinary Least Squares (OLS) Estimation

#### OLS Estimation for AR Models

- Assumptions: Stationary time series.
- AR(1) model:  $X_t \mu = \phi (X_{t-1} \mu) + \epsilon_t \equiv X_t = \phi_0 + \phi X_{t-1} + \epsilon_t$ .
- The least-squares method aims to minimize the conditional sum of squares (CSS) with respect to the parameters φ<sub>0</sub>, φ:

$$S_c(\phi_0, \phi) = \sum_{t=2}^{N} (x_t - \phi_0 - \phi x_{t-1})^2$$

• By setting  $\partial S_c(\phi_0, \phi)/\partial \phi_0 = \partial S_c(\phi_0, \phi)/\partial \phi = 0$  we obtain

$$\hat{\phi}_0 = \frac{1}{(N-1)} \left( \sum_{t=2}^N x_t - \hat{\phi} \sum_{t=2}^N x_{t-1} \right)$$

$$\hat{\phi} = \frac{\sum_{t=2}^{N} (x_t - \overline{x}) (x_{t-1} - \overline{x}_*)}{\sum_{t=2}^{N} (x_{t-1} - \overline{x}_*)^2}, \ \overline{x} = \frac{1}{N-1} \sum_{t=2}^{N} x_t, \ \overline{x}_* = \frac{1}{N-1} \sum_{t=2}^{N} x_{t-1}$$

 At the asymptotic limit N → ∞ the OLS and MoM estimates are practically the same for AR(p) models.

## Ordinary Least Squares (OLS) Estimation

### OLS estimates for MA(q) and ARMA(p,q) models

The main idea is to minimize the conditional sum of squares (CSS)

$$S(\theta_1, \dots, \theta_q \mid \epsilon_0, \dots, \epsilon_{-q+1}) = \sum_{t=1}^N \epsilon_t^2 \mid \epsilon_0, \epsilon_{-1}, \dots, \epsilon_{-q+1}.$$

- Initial conditions: Assume  $\epsilon_0 = \epsilon_{-1} = \ldots = \epsilon_{-q+1} = 0$ .
- Then, based on  $\epsilon_t = X_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}$ , the innovations can be recursively estimated at every *t* using the observed time series  $\{x_1, \ldots, x_N\}$ .
- The minimization with respect to θ can be carried out numerically.
- Similar equations are obtained for ARMA(p,q) models with initial conditions  $\epsilon_p, \epsilon_{p-1}, \ldots, \epsilon_{p+1-q} = 0$ .
- Problem: For short samples the initial conditions may significantly impact the estimates.
- Exercise: Express the CSS for the MA(1) model.

#### General concepts

- The maximum likelihood estimation (MLE) method addresses some of the pitfalls of MoM and OLS.
- The likelihood,  $\mathcal{L}$ , is a function  $\mathcal{L}(\theta; x_1, \dots, x_N)$ , where  $\theta$  is the parameter vector.
- Conditioning information:  $\mathcal{I}_t$  represents the available information at time  $t: \mathcal{I}_{t-1} \triangleq \{x_{t-1}, x_{t-2}, \dots, x_1\}.$
- Two approaches: Conditional MLE and exact MLE.
- In conditional MLE we do not consider the initial values (depending on the order of the model) of the time series (we condition on the past).

#### Conditional MLE for AR(p) Models

• The  $\mathcal{L}(\theta; x_1, \dots, x_N)$  is given by joint PDF of the time series:

$$f(x_1,\ldots,x_N;\boldsymbol{\theta}) = \left(\prod_{t=p+1}^N f(x_t \mid \mathcal{I}_{t-1};\boldsymbol{\theta})\right) f(x_1,\ldots,x_p;\boldsymbol{\theta})$$

•  $\prod_{t=p+1}^{N} f(x_t \mid \mathcal{I}_{t-1}; \theta)$ : Product rule for joint probability.

•  $f(x_1, \ldots, x_p; \theta)$ : Joint PDF at the left-boundary points.

• Conditional log-likelihood:

$$\ell_c(\boldsymbol{\theta}; x_{p+1}, \dots, x_N) = \sum_{t=p+1}^N \log f(x_t \mid \mathcal{I}_{t-1}; \boldsymbol{\theta}),$$

• Conditional MLE:  $\hat{\theta} = \arg \max_{\theta} \sum_{t=p+1}^{N} \log f(x_t \mid \mathcal{I}_{t-1}; \theta)$ .

#### Conditional MLE for AR(1) Model

- The parameter vector is  $\boldsymbol{\theta} = (\phi_0, \phi, \sigma_{\epsilon}^2)^{\top}$ .
- The AR(1) conditional likelihood is given by

$$\ell_c(\theta; \mathbf{x}) = \sum_{t=2}^N \log f(x_t \mid x_{t-1}; \theta) = -\frac{N-1}{2} \log(2\pi) - \frac{N-1}{2} \log(\sigma_{\epsilon}^2) - \frac{1}{2\sigma_{\epsilon}^2} \sum_{t=2}^N (x_t - \phi x_{t-1} - \phi_0)^2$$

- The solution is obtained by  $\frac{\partial \ell_c(\boldsymbol{\theta};\mathbf{x})}{\partial \phi_0} = \frac{\partial \ell_c(\boldsymbol{\theta};\mathbf{x})}{\partial \sigma_{\epsilon}^2} = \frac{\partial \ell_c(\boldsymbol{\theta};\mathbf{x})}{\partial \phi} = 0.$
- The above conditions lead to

$$\hat{\phi}_0 = \hat{\phi}_{0;\text{OLS}}, \ \hat{\phi} = \hat{\phi}_{\text{OLS}}, \ \hat{\sigma}_{\epsilon}^2 = \frac{1}{N-1} \sum_{t=2}^N \left( x_t - \hat{\phi}_0 - \hat{\phi} x_{t-1} \right)^2$$

#### Exact MLE for AR(1) Model

- In order to take into account the information provided by  $x_1$  we need to use the full likelihood  $\ell(\theta; \mathbf{x}) = \ell_c(\theta; \mathbf{x}) + \log f(x_1; \theta)$ .
- The marginal log-likelihood at the boundary point  $t_1$  is given by

$$\log f(x_1; \boldsymbol{\theta}) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_X^2) - \frac{1}{2\sigma_X^2} (x_1 - \phi_0)^2$$

- The variance at the boundary (due to stationarity) is given by  $\sigma_X^2 = \sigma_\epsilon^2/(1-\phi^2)$ .
- The exact likelihood is given by  $\ell(\theta; \mathbf{x}) = \ell_c(\theta; \mathbf{x}) + \log f(x_1; \theta)$ .
- In general, there are no closed-form solutions for the maximum of the exact log-likelihood.
- Numerical optimization methods are used for exact MLE.

#### What constitutes a good forecast?

- Forecast:  $\hat{X}_N(k)$ .
- Forecast error:  $\varepsilon_N(k) = X_{N+k} \hat{X}_N(k)$ .
- Unbiased forecast:  $\mathbb{E}[\varepsilon_N(k)] = 0.$
- Efficiency: Minimize  $\mathbb{V}ar[\varepsilon_N(k)]$ .
- Combine low bias and efficiency in MMSE: Minimize  $\mathbb{E}[\varepsilon_N^2(k)]$ .
- A probabilistic forecast involves the optimal prediction, i.e., x̂<sub>N</sub>(k) and a measure of uncertainty based on σ<sub>ε<sub>N</sub>(k)</sub>.
- Prediction interval at confidence level  $1 \alpha$  (e.g.,  $\alpha = 0.05$ ):

$$[\hat{x}_N(k) - z_{\alpha/2}\sigma_{\varepsilon_N(k)}, \ \hat{x}_N(k) + z_{\alpha/2}\sigma_{\varepsilon_N(k)}]$$

where  $z_{\alpha/2} = 1.96$  for  $\alpha = 0.05$ .

#### Deterministic trend with noise

- Model:  $X_t = \mu_t + \epsilon_t$ .
- Prediction:  $\hat{x}_N(k) = \mu_{N+k}$ .
- The trend is known *a priori* or estimated, e.g. by fitting the data to a global or a local polynomial model.
- Error:  $\varepsilon_N(k) = \epsilon_{N+k}$ .

#### Deterministic trend, Seasonal term and noise

- Model:  $X_t = \mu_t + S_t + \epsilon_t$ .
- Prediction:  $\hat{x}_N(k) = \mu_{N+k} + S_{N+k}$ .
- The trend and seasonal terms are either known a priori or estimated by fitting the data to a deterministic model.

• Error: 
$$\varepsilon_N(k) = \epsilon_{N+k}$$
.

#### Types of Forecasts

- Forecasts can be either in-sample (interpolation) or out-of sample (extrapolation).
- In general, out-of sample forecasts are a stricter test of model performance, as the prediction is compared with data outside the range used for model estimation.
- In the case of out-of-sample forecasts we have the option of multistep-ahead and one-step ahead prediction.
- The prediction uncertainty tends to increase with the forecasting

(GSLAB)



#### Forecasting Time Series in R. Jocelyn Barker

#### Commonly used prediction strategies

- Multistep ahead prediction: The model is determined based on  $\{x_1, x_2, \ldots, x_N\}$ ; predictions are constructed for  $\hat{x}_N(1), \hat{x}_N(2), \ldots, \hat{x}_N(k)$ .
- One-step ahead prediction:
  - **(**) Estimate the model based on  $\{x_1, x_2, \ldots, x_N\}$
  - Predict each time  $t_{N+1}, \ldots, t_{N+k}$  using all the data up to the previous time, i.e.,  $\hat{x}_N(1), \hat{x}_{N+1}(1), \ldots, \hat{x}_{N+k-1}(1)$ .
- Rolling-window forecast: A model is estimated for different window subsamples; a forecasting horizon h is defined; the model which returns an optimal prediction performance metric over all subsamples is selected.


### **Conditional Forecasting**

The conditional expectation of the time series at time N + k ( $k \ge 1$ ) is based on all the information available up to and including the time  $t_N = N\delta t$ :

 $\hat{x}_N(k) \doteq \mathbb{E}[X_{N+k} \mid \mathcal{I}_N], \text{ where } \mathcal{I}_N = \{X_N = x_N, X_{N-1} = x_{N-1}, \dots, X_1 = x_1\}.$ 

### Forecast of ARMA(p,q) Model

$$\begin{split} \Phi(B)X_t &= \Theta(B)\epsilon_t, \\ x_{N+k} &= \sum_{i=1}^p \phi_i x_{N+k-i} + \epsilon_{N+k} - \sum_{j=1}^q \theta_j \epsilon_{t+k-j} \\ \hat{x}_N(k) &= \sum_{i=1}^p \phi_i \hat{x}_N(k-i) + \hat{\epsilon}_N(k) - \sum_{j=1}^q \theta_j \hat{\epsilon}_N(k-j) \end{split}$$

 $\hat{x}_{N}(j) = \mathbb{E}[X_{N+j} \mid \mathcal{I}_{N}], \ j \ge 1$  $\hat{x}_{N}(j) = x_{N+j}, \ j \le 0,$  $\hat{\epsilon}_{N}(j) = 0, \ j \ge 1,$  $\hat{\epsilon}_{N}(j) \triangleq \epsilon_{N+j}, \ j \le 0$  $\epsilon_{N+j} = x_{N+j} - \hat{x}_{N+j}$ 

### Conditional Forecasting: AR(1) Model

AR(1) model:  $x_t = \phi x_{t-1} + \epsilon_t$ 

• 
$$t \to n+1$$
:  $x_{n+1} = \phi x_n + \epsilon_{n+1}$ 

• Prediction:  $\hat{x}_n(1) = \phi x_n$ 

• 
$$t \rightarrow n+2$$
:  $x_{n+2} = \phi x_{n+1} + \epsilon_{n+2}$ 

•  $t \to n+k$ :  $x_{n+k} = \phi x_{n+k-1} + \epsilon_{n+k}$ Prediction:  $\hat{x}_n(k) = \phi \hat{x}_n(k-1) = \phi^k x_n$ 

$$\rightarrow n+2: x_{n+2} = \phi x_{n+1} + \epsilon_{n+2}$$

Prediction: 
$$\hat{x}_n(2) = \phi \hat{x}_n(1) = \phi^2 x_n$$

Innovation: GWN ~  $\mathcal{N}(0, \sigma_{\epsilon}^2)$ 

• Error: 
$$\varepsilon_n(1) = \epsilon_{n+1}$$

• Variance: 
$$\mathbb{V}ar[\varepsilon_n(1)] = \sigma_{\epsilon}^2$$

• Error:  

$$\varepsilon_n(2) = \phi \epsilon_{n+1} + \epsilon_{n+2}$$

• Variance: 
$$\mathbb{V}\mathrm{ar}[\varepsilon_n(2)] = (\phi^2 + 1)\sigma_\epsilon^2$$

Error:  $\varepsilon_n(k) = \sum_{i=1}^{k-1} \phi^{k-i} \epsilon_{n+i}.$ 

• Variance:  

$$\mathbb{V}ar[\varepsilon_n(k)] = \frac{1-\phi^{2k}}{1-\phi^2}\sigma_{\epsilon}^2$$

۲

### Conditional Forecasting: AR(p) Model

 $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t$ 

- Innovation: GWN  $\sim \mathcal{N}(0, \sigma_{\epsilon}^2)$
- $x_{n+1} = \phi_1 x_n + \dots \phi_p x_{n-p+1} + \epsilon_{n+1}$
- $\hat{x}_n(1) = \phi_1 x_n + \phi_2 x_{n-1} + \ldots + \phi_p x_{n-p+1}$
- $x_{n+2} = \phi_1 x_{n+1} + \ldots + \phi_p x_{n-p+2} + \epsilon_{n+2}$
- $\hat{x}_n(2) = \phi_1 \hat{x}_n(1) + \phi_2 x_n \dots + \phi_p x_{n-p+2}$
- $x_{n+k} = \phi_1 x_{n+k-1} + \ldots + \phi_p x_{n-p+k} + \epsilon_{n+k}$
- $\hat{x}_n(k) = \phi_1 \hat{x}_n(k-1) + \ldots + \phi_p \hat{x}_n(k-p),$

where  $\hat{x}_n(j) = x_{n+j}$ , if  $j \leq 0$ 

- Error:  $\varepsilon_n(1) = \epsilon_{n+1}$
- Variance:  $\mathbb{V}ar[\varepsilon_n(1)] = \sigma_{\epsilon}^2$

• 
$$\varepsilon_n(2) = \phi_1 \varepsilon_n(1) + \epsilon_{n+2}$$

• 
$$\operatorname{Var}[\varepsilon_n(2)] = (\phi_1^2 + 1)\sigma_\epsilon^2$$

• 
$$\varepsilon_n(k) = \sum_{j=1}^{k-1} \phi_j \varepsilon_n(k-j) + \epsilon_{n+k}$$

• 
$$\operatorname{Var}[\varepsilon_n(k)] = \left(1 + \sum_{j=1}^{k-1} \phi_j^2\right) \sigma_\epsilon^2$$

Conditional Forecasting: MA(1) Model	
MA(1) model: $x_t = \epsilon_t + \theta \epsilon_{t-1}$	Innovation: GWN $\sim \mathcal{N}(0,\sigma_{\epsilon}^2)$
• $t \to n + 1$ : $x_{n+1} = \epsilon_{n+1} + \theta \epsilon_n$ • Prediction: $\hat{x}_n(1) = \theta \epsilon_n$	• Error: $\varepsilon_n(1) = \epsilon_{n+1}$ • Variance: $\mathbb{V}ar[\varepsilon_n(1)] = \sigma_{\epsilon}^2$
• $x_{n+2} = \theta \epsilon_{n+1} + \epsilon_{n+2}$ • Prediction: $\hat{x}_n(2) = 0$	• Error: $\varepsilon_n(2) = x_{n+2}$ • Variance: $\mathbb{V}ar[\varepsilon_n(2)] = \sigma_x^2$
<ul> <li>For k &gt; 1: x<sub>n+k</sub> = θε<sub>n+k-1</sub> + ε<sub>n+k</sub></li> <li>Prediction: x̂<sub>n</sub>(k) = 0</li> <li>The innovation ε<sub>n</sub> can be estimated if  θ  &lt; 1 using invertibility: ε<sub>n</sub> = Σ<sup>∞</sup><sub>i=0</sub> θ<sub>j</sub>x<sub>n-j</sub></li> </ul>	• Error: $\varepsilon_n(k) = x_{n+k}$ • Variance: $\mathbb{V}ar[\varepsilon_n(k)] = \sigma_x^2$

### Conditional Forecasting: MA(q) Model

MA(q) model:  $x_t = \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}$ 

$$x_{n+1} = \epsilon_{n+1} + \sum_{i=1}^{q} \theta_q \epsilon_{n+1-i}$$

• 
$$\hat{x}_n(1) = \theta_1 \epsilon_n + \ldots + \theta_q \epsilon_{n-q+1}$$

• 
$$x_{n+2} = \sum_{i=1}^{q} \theta_i \epsilon_{n+2-i} + \epsilon_{n+2}$$

• 
$$\hat{x}_n(2) = \theta_2 \epsilon_n + \ldots + \theta_q \epsilon_{n-q+2}$$

• 
$$x_{n+k} = \sum_{i=1}^{q} \theta_i \epsilon_{n+k-i} + \epsilon_{n+k}, \ k > 1$$

• 
$$\hat{x}_n(k) = \sum_{i=k}^q heta_i \epsilon_{n+k-i}$$
, for  $k \leq q$ 

• 
$$\hat{x}_n(k) = 0$$
, for  $k > q$ 

Innovation: GWN ~  $\mathcal{N}(0, \sigma_{\epsilon}^2)$ 

• 
$$\varepsilon_n(1) = \epsilon_{n+1}$$
  
•  $\operatorname{Var}[\varepsilon_n(1)] = \sigma_e^2$ 

• 
$$\varepsilon_n(2) = \epsilon_{n+2} + \theta_1 \epsilon_{n+1}$$
  
•  $\mathbb{V}[\mathbf{r}_n(2)] = (\theta_1^2 + \theta_1) - \theta_1^2$ 

• 
$$\operatorname{Var}[\varepsilon_n(2)] = (\theta_1^2 + 1)\sigma_\epsilon^2$$

• 
$$\varepsilon_n(k) = \epsilon_{n+k} + \sum_{i=1}^{k-1} \theta_i \epsilon_{n+k-i}$$
  
•  $\operatorname{Var}[\varepsilon_n(k)] = \sigma_{\epsilon}^2 \left(1 + \sum_{i=1}^{k-1} \theta_i^2\right).$ 

# Forecasting Non-Stationary Time Series

### Stationarity-restoring transformation and ARMA

- Apply nonlinear transformation X = g(Y) to the data  $\{y_1, \ldots, y_N\}$  to restore stationarity in the transformed series  $\{x_1, \ldots, x_N\}$ .
- Use ARMA(p,q) or other model to predict  $\{\hat{x}_{N+1}, \ldots, \hat{x}_{N+P}\}$ .
- Use inverse the transformation,  $Y = g^{-1}(X)$ , to obtain forecasts  $\{\hat{y}_{N+1}, \ldots, \hat{y}_{N+P}\}$  in the original domain.

#### Normalizing transformation and ARMA/ARIMA

- Apply nonlinear transformation *X* = *g*(*Y*) to the data {*y*<sub>1</sub>,...,*y<sub>N</sub>*} to obtain normal distribution of {*x*<sub>1</sub>,...,*x<sub>N</sub>*}.
- If  $\{x_1, \ldots, x_N\}$  is stationary, use ARMA(p,q) to predict.
- If the transformed series is non-stationary use an ARIMA model.
- Invert transformation,  $Y = g^{-1}(X)$ , to forecast  $\{\hat{y}_{N+1}, \dots, \hat{y}_{N+P}\}$  in original domain.

### Forecasting Non-Stationary Time Series

### Forecasting for ARIMA(p,1,q)

- k-step-ahead forecast  $(k \ge 1)$ .
- First, calculate  $y_t = x_t x_{t-1}$ , for  $t = 2, 3, \ldots$
- Using ARMA(p,q) estimate  $\hat{y}_N(k)$ .
- Invert differencing:  $\hat{x}_N(k) = \hat{x}_N(k-1) + \hat{y}_N(k)$ , where  $\hat{x}_N(0) = x_N$ .
- Hence, forecasting  $\hat{x}_N(k)$  requires iteratively forecasting  $\hat{x}_N(1) \mapsto \hat{x}_N(2) \dots \mapsto \hat{x}_N(k-1)$ .
- Since  $\hat{x}_N(1) = x_N + \hat{y}_N(1) \implies \varepsilon_{X;N}(1) = \varepsilon_{Y;N}(1)$ .
- One can then show recursively that  $\varepsilon_{X;N}(k) = \sum_{j=1}^{k} \varepsilon_{Y;N}(j)$ .

# Forecasting Non-Stationary Time Series

### Nonlinear transformations do not preserve the mean

- Caution: The forecasts  $\hat{y}_{N+p}$  obtained by inverting a nonlinear transformation are not necessarily the conditional expectations of the process  $Y_t$ .
- Monotonic nonlinear transformations preserve quantiles of the marginal predictive distribution but not the mean.
- If X is normally distributed, the conditional mean  $\hat{x}_{N+p}$  is equivalent to the median of the marginal predictive distribution  $f_X(x_{N+p} | \mathcal{I}_N)$ .
- Then  $\hat{y}_{N+p}$  is equivalent to the median of the marginal predictive distribution  $f_Y(y_{N+p} | \mathcal{I}_N)$ .
- More generally, the principle of *quantile invariance* holds: If the transformation  $g(\cdot)$  is a *monotonically increasing function*, it preserves ordering, i.e.,  $y_1 \leq y_2 \Leftrightarrow g(y_1) \leq g(y_2)$ . Such transformations also preserve quantiles: If  $\hat{y}_p$  is *q*-th quantile of the marginal distribution  $F_y(y)$ , then  $F_x(\hat{x}_p)$ , where  $\hat{x}_p = g^{-1}(\hat{y}_p)$ corresponds to the *q*-th quantile of the marginal distribution  $F_x(x)$ .

- EWMA is a simple smoothing and forecasting method that works reasonably well with minimal assumptions and parameter tuning.
- EWMA 1-step-ahead predictive equation  $(\hat{x}_n(1) \equiv \hat{x}_{n+1})$ :

$$\hat{x}_n(1) = \sum_{k=0}^{n-1} w_k x_{n-k}, \quad w_0 \ge w_1 \ge w_2 \ge \dots 0, \quad \sum_{k=0}^{n-1} w_k = 1.$$

• Weights:  $w_k = \alpha(1-\alpha)^k$  where  $\alpha < 1$  and  $1-\alpha$  is the smoothing factor.

• 
$$w_k = \alpha (1 - \alpha)^k$$
,  $k = 0, 1, 2, ...$ 

Recursive estimate of non-deterministic mean (smoothing)

$$\hat{\mu}_t = \alpha x_t + (1 - \alpha)\hat{\mu}_{t-1}, \ t = 1, 2, \dots$$

- $\alpha \approx 1$ : little smoothing (variations attributed to the mean).
- $\alpha \approx 0$ : considerable smoothing (variations attributed to stochastic term).
- $\hat{\mu}_t$  is known as the *level*.

### Example of smoothing



#### EWMA used in Forecasting

- Forecast for any time horizon k:  $\hat{x}_n(k) = \hat{\mu}_n, \ k = 1, 2, ...$
- One-step-ahead forecast:

$$\hat{x}_{n+1} \equiv \hat{x}_n(1) = \sum_{k=0}^{n-1} w_k x_{n-k} = \sum_{k=0}^{n-1} \alpha (1-\alpha)^k x_{n-k}$$
$$= \alpha x_n + (1-\alpha) \sum_{l=0}^{n-2} \alpha (1-\alpha)^l x_{n-l-l} = \alpha x_n + (1-\alpha) \hat{x}_{n-1}(1)$$

• The next-step forecast based on the augmented dataset  $\{x_1, x_2, \ldots, x_n, x_{n+1}\}$  is:

 $\hat{x}_{n+2} \equiv \hat{x}_{n+1}(1) = \alpha x_{n+1} + (1-\alpha)\hat{x}_n(1)$ 

• An "optimal  $\alpha$ " minimizes the sum of squared 1-step (ahead) prediction errors SS1PE: SS1PE =  $\sum_{t=2}^{n} \varepsilon_t^2 \triangleq \sum_{t=2}^{n} [x_t - \hat{x}_{t-1}(1)]^2$ 

#### Example of EWMA One-step-ahead forecasting



(GSLAB)

# Generalization of EWMA model

### Holt-Winters Model

The Holt-Winters model extends EWMA by adding deterministic trend and seasonality effects:

$$\mu_t = \alpha(x_t - s_{t-T}) + (1 - \alpha)(\mu_{t-1} + b_{t-1}) \tag{4}$$

$$b_t = \beta \left( \mu_t - \mu_{t-1} \right) + (1 - \beta) \, b_{t-1} \tag{5}$$

$$s_t = \gamma \left( x_t - \mu_t \right) + \left( 1 - \gamma \right) s_{t-T} \tag{6}$$

- $\mu_t$  is the "level" of the process.
- $\mu_{t-1} + b_{t-1}$  is the 1-step level forecast at t-1.
- *s<sub>t</sub>* is the periodic component with period *T*.
- $b_t$  is the local slope.
- Initialization:  $\mu_1 = x_1$ ;  $b_1 = 0, s_1, \dots, s_T = 0$  or  $\{s_i\}_{i=1}^T$  are set equal to the mean over the observations for the respective "season".
- Forecast  $(t \rightarrow n)$ :  $\hat{x}_n(k) = \mu_n + kb_n + s_{n+k-T}$ .
- "Typical" parameter values:  $\alpha = \beta = \gamma = 0.2$ .

### Model Assessment

### Model Selection Criteria

- Typically, data are fitted to different time series models.
- The models are compared in terms of model selection criteria that quantify the goodness of fit of each model.
- N: Number of points in time series; K: number of parameters; L: likelihood of the model.
- Akaike information criterion (AIC): AIC =  $2K - 2\log \mathcal{L} \approx 2K + N\log \hat{\sigma}_{\epsilon}^2$ .
- Bayesian information criterion (BIC): BIC =  $K \log N - 2 \log \mathcal{L} \approx K \log N + N \log \hat{\sigma}_{\epsilon}^2$ .
- For ARMA(p,q) models K = p + q + 1.
- Models with lower AIC (BIC) values are preferred to models with higher AIC (BIC) values.
- Cross-validation can also be used for model selection.

### Model Selection Criteria

### Preference for Parsimonious Models (Occam's razor)

- The aim is to produce a model that is parsimonious, i.e., it has as few parameters as possible (*principle of parsimony*-Occam's razor), while it also passes the diagnostic checks.
- A parsimonious model is desirable because (i) it is more interpretable and (ii) including irrelevant lags in the model increases the coefficients' standard errors.
- High-order models that incorporate too many lags tend to forecast poorly, because they fit the data too closely adapting to random (non-systematic) features in the data (compare to overfitting in linear regression problem).
- Combining weighted forecasts (e.g., ARIMA with EWMA) can improve accuracy.

# **Cross-validation Analysis**

#### Types of Cross-Validation

In cross-validation (CV) we use part of the N data points to train the model and the remaining data to test performance. Different CV approaches include:

- k-fold CV: partitioning the dataset in k folds of which k - 1 are used for training; repeated k times.
- Leave-*p*-out CV: all ways of splitting the dataset into a training set of N p points and a validation set with *p* points.
- Leave-one-out CV: leave-p-out CV with p = 1.
- Holdout CV: random splitting of dataset into training and validation sets.
- Rolling CV Time series CV: see figure (from Rob J. Hyndman's website).

# Rolling-forecast cross-validation:



# **Cross-validation Analysis**

### Mean Squared Error (MSE)

- Let us assume that the time series is divided into two parts:
- The training set includes the first N time series values  $\{x_1, \ldots, x_N\}$
- The validation set includes the remaining M values  $\{x_{N+1}, \ldots, x_{N+M}\}$
- The mean square error (MSE) is then defined as

MSE = 
$$\frac{1}{M} \sum_{k=1}^{M} [x_{N+k} - \hat{x}_N(k)]^2$$

 Other assessment measures are also used, e.g., RMSE = sqrt(MSE), mean error (ME), mean absolute error (MAE), coefficient of determination R<sup>2</sup>.

# **Cross-validation Analysis**

#### Tests of Forecast Accuracy

- Tests of forecast accuracy are based on the difference between the forecast of the variables value at time t and the actual value at time t (forecast error)
- Statistical forecast measures include:
  - Mean Absolute Error
  - 2 Mean Average Prediction Error
  - Ohow's test for predictive failure (structural break in the time series)
  - Orrelation (linear, Spearman) between true values and forecasts
  - Soot Mean Square Error (the square root of the MSE)
  - Percentage of interval coverage (what percent of validation data fall inside the prediction intervals)

### **Residuals Diagnostics**

### Are the residuals consistent with the model assumptions?

- Are the residuals zero-mean (unbiased forecast)?
- Are the residuals normally distributed?
- Are they uncorrelated?
- Are they stationary (homoskedastic)?

### What do do next?

- If the residuals satisfy the desired properties, then the model is consistent with the data.
- If the residuals fail one or more of the above properties, then a more suitable model may be necessary for the dataset.

### **Spectral Analysis**

# **Regarding Spectral Density Estimation**

### Estimation methods

- Non-parametric methods are data-driven; they estimate the PSD directly from the time series without making assumptions.
- Parametric methods assume that the time series can be represented as the response of a linear system to an innovation process. They can be more efficient than non-parametric methods if the time series is short.
- Subspace methods, also known as *high-resolution methods* estimate the spectral content by diagonalizing the auto-correlation matrix.

# **Regarding Spectral Density Estimation**

### Synopsis of different spectral methods

Method	Features	
Periodogram	Raw PSD estimation	
Welch (modified peri- odogram) Multitaper methods Autoregressive	Average periodogram calculated over overlap- ping segments Estimates combine multiple estimation win- dows (tapers) Use of AR model	
Burg Modified autocovari-	AR model based on mininization of forward lin- ear forecasting errors Regression model based on mininization of for- word/backward forecasting errors	
Multiple Signal Classifi- cation (Music) Eigenvector	Pseudo-spectrum estimation based on eigen- vectors of auto-correlation matrix Pseudo-spectrum estimation based on	
	weighted version of Music.	

#### The periodogram estimator Estimate the discrete Fourier transform of the time series

$$\tilde{X}_n = \sum_{t=0}^{N-1} X_t \, e^{-j \, 2\pi \, n \, t/N} \,, \quad n = 0, \dots, N-1.$$

The discrete frequencies used are

$$f_n = \frac{2f_c n}{N} = \frac{f_s n}{N}, \quad n = 0, 1, \dots, \frac{N}{2} - 1$$

f<sub>s</sub> = 1/δt: sampling frequency; f<sub>c</sub> = f<sub>s</sub>/2: Nyquist (critical) frequency.
 For real-valued time series, the one-sided periodogram is used

$$\hat{P}_X(f_0) = \frac{1}{N^2} \|\tilde{X}_0\|^2, \ \hat{P}_X(f_c) \triangleq \hat{P}_X(f_{N/2}) = \frac{1}{N^2} \|\tilde{X}_{N/2}\|^2.$$
$$\hat{P}_X(f_n) = \frac{1}{N^2} \left[ \|\tilde{X}_n\|^2 + \|\tilde{X}_{N-n}\|^2 \right], \ n = 1, 2, \dots, N/2 - 1.$$

#### Properties of the Periodogram

- Spectral leakage: Mixing of frequencies due to finite observation window.
- $X_{obs}(t) = X(t) \Pi_N(t)$ , where  $\Pi_N(t) = 1$  if  $t = 0, \delta t, \dots, (N-1) \delta t$  and  $\Pi_N(t) = 0$  for all other t is the window function.
- Hence,  $\tilde{X}_{obs}(f) = \tilde{X}(f) * \tilde{\Pi}_N(f)$
- FT of the window function:

$$\tilde{\Pi}_N(f) = \sum_{k=0}^{N-1} e^{j2\pi f_n k\delta t} = e^{j\pi f_n (N-1)\delta t} \left[ \frac{\sin(\pi N f_n \delta t)}{\sin(\pi f_n \delta t)} \right], \ f_n = \frac{n}{N\,\delta t}$$

Hence,

$$\hat{P}_X(f) = \|\tilde{X}_{obs}(f)\|^2 = \|\tilde{X}(f) * \tilde{\Pi}_N(f)\|^2$$

• Consequently,  $\hat{P}_X(f) = \|\tilde{X}(f)\|^2$  only if  $\tilde{\Pi}_N(f) = \delta(f - f')$ .



Periodogram of the superposition of two sine functions with frequencies 140 Hz and 150 Hz contaminated with white noise exhibiting spectral leakage.

#### Other properties of the periodogram

- Resolution: It determines the ability to distinguish between neighboring frequencies.
- Two frequencies can be resolved if  $\Delta f \ge f_{\rm res} = \frac{f_s}{N}$ .
- Bias: The periodogram is biased due to spectral leakage.
- Asymptotic lack of bias: However, at the limit  $\lim_{N\to\infty} \tilde{\Pi}_N(f) = \delta(f f')$  and the periodogram is *asymptotically unbiased*.
- Precision: The variance of the periodogram is proportional to the square of the expected spectral density and independent of *N*.
- However, this is not a problem for time series with a strong deterministic periodic component.

### Regarding the use of windows

- The raw periodogram estimator suffers from: spectral leakage and low precision.
- Precision can be improved by segmenting the time series in different windows, estimating the periodogram in each window, and then averaging over all windows.
- The spectral leakage is caused by the finite observation time.



physics-for-understanding-fmri

Spectral leakage is reduced by using spectrum analysis windows.

#### Method of Welch (Modified periodogram)

- The main idea: Introduce leakage (in terms of weights) in the time domain to reduce spectral leakage.
- Modified Fourier transform

$$\breve{X}_n = \sum_{k=0}^{N-1} X_k w_k e^{-i 2\pi k n/N}, \quad k = 0, \dots, N-1$$

Modified periodogram estimator

$$\hat{P}_X(f_0) = \frac{1}{W_{ss}} \|\breve{X}_0\|^2, \quad \hat{P}_X(f_c) \equiv \hat{P}_X(f_{N/2}) = \frac{1}{W_{ss}} \|\breve{X}_{N/2}\|^2$$

$$\hat{P}_X(f_n) = \frac{1}{W_{ss}} \left[ \| \breve{X}_n \|^2 + \| \breve{X}_{N-n} \|^2 \right] \quad n = 1, 2, \dots, N/2 - 1,$$

• Normalization factor:  $W_{ss} = N \sum_{n=0}^{N-1} w_n^2$  .

### Spectral windows

• Spectral form for a general window function  $\{w_k\}_{k=0}^{N-1}$ :

$$W_T(f_n) = \frac{1}{W_{ss}} \| \sum_{k=0}^{N-1} e^{j 2\pi f_n k \delta t} w_k \|^2,$$

• Welch window:

$$w_k = 1 - \left(\frac{k - N/2}{N/2}\right)^2.$$

• Different window models are available, and they have different properties concerning the reduction of spectral leakage.



#### Improving precision

- The periodogram estimator has high variance.
- This can be reduced by segmenting a time series of length N = 2K M into K segments of length 2M each.
- The periodogram is evaluated separately for each of the K segments.
- An average is evaluated over the *K* periodograms.
- This reduced the variance by 1/K.

### Hamming Window in Time and Frequency Domains



### Impact of Windowing on Spectral Density Estimation



# **Spectral Analysis**

#### Estimation of Power Spectral Density: Sunspot numbers

```
sspf = ssp - mean(ssp); % Centering
cl = 0.95; % Confidence level
fs = 1; % Sampling frequency
Nfft = 256; % No. FFT points
& Define window function
window = hamming(length(ssp));
% Estimate periodogram
[pxx, f, pxxc] = periodogram(sspf, window,
... Nfft, fs, 'ConfidenceLevel', cl);
pxxdb = 10 \times log10 (pxx);
pxxcdb = 10*log10(pxxc);
plot(f,pxxdb,'b-','LineWidth',2);
hold on;
plot(f,pxxcdb,'r-.');
axis tight; xlabel('Hz')
ylabel('Power (dB)');
title('Sunspot number');
grid on; set(gca, 'FontSize', 14);
```

Why do we remove the mean value of the measurements?

Why do we use the Hamming window;

# **Spectral Analysis**

#### Estimation of power spectral density

```
% Find dominant peak
[pks, flocs]=findpeaks(pxxdb,f
    . . . .
   'SortStr', 'descend');
plot(flocs(1), pks(1), 'mo',
   'MarkerFaceColor', 'g', ...
    'MarkerSize', 8);
% Dominant period
period=round(1/flocs(1));
labt=[num2str(period) ' Years'
   1;
gtext(labt, 'FontSize', 14);
```


#### Fitting a Polynomial Trend Model to Time series Data

```
tt <- as.numeric(time(co2))
fit <-lm(co2~poly(tt,degree
    =1,raw=TRUE))
fit2 <-lm(co2~poly(tt,degree
    =2,raw=TRUE))
plot(co2)
lines(tt,predict(fit),col='
    red')
lines(tt,predict(fit2),col='
    blue')</pre>
```



### Removing the Trend from Time series Data

```
trend2 <- predict(fit2)
fit4 <- lm(co2~poly(tt,
    degree=4, raw=TRUE))
trend4 <- predict(fit4)
co2fluc1 = co2-trend
co2fluc2 = co2-trend2
co2fluc4 = co2-trend4
plot(co2fluc)
lines(co2fluc2, col='green')
lines(co2fluc4, col='red')</pre>
```

- Residuals based on 3 polynomial trend models
- Linear fit, -second degree, -fourth degree



#### Testing the Residuals for Correlations

acf(ts(co2fluc4,freq=1),lag.max=40,main="Autocorrelation
Function for CO2 Data after Removing 4th-order Polynomial",
ylim=c(-1,1))



#### Constructing a Model for the Periodic Component

- (Intercept) -2.125e-09 3.474e-02 0.000 1
  s1 2.748e+00 4.913e-02 55.933 < 2e-16 \*\*\*
  s2 -3.968e-01 4.913e-02 -8.076 5.78e-15 \*\*\*
  --Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
  Residual standard error: 0.7516 on 465 degrees of freedom
  Multiple R-squared: 0.8729, Adjusted R-squared: 0.8724
  F-statistic: 1597 on 2 and 465 DF, p-value: < 2.2e-16</pre>
- The smaller the p-value for a variable, the more significant the specific variable is for the model.
- The p-value is the probability of obtaining the observed statistic (here, t-value) if the null hypothesis (i.e., constant value) holds.

### Testing for Correlations after Removing the Periodic Component



There are remaining periodic correlations but their magnitude is reduced.

(GSLAB)

### Testing the Model so Far (Trend and Periodic Component)

```
co2e <- periodic + trend4
plot(tt,co2)
lines(tt,co2e,col='red')</pre>
```



#### Estimate the trend and periodic models simultaneously

```
s1<-sin(tt*2*pi/1)
s2<-cos(tt*2*pi/1)
fit.complete <- lm(formula = co2~poly(tt,degree=2,raw=TRUE) +
    s1 + s2)
summary(fit.complete)</pre>
```

```
call:
lm(formula = co2 \sim poly(tt, degree = 2, raw = TRUE) + s1 + s2)
Residuals:
              10 Median 30
    Min
                                       Max
-2.87720 -0.66136 0.00051 0.62583 2.36405
Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)
                                 4.771e+04 1.447e+03 32.979 < 2e-16 ***
polv(tt, degree = 2, raw = TRUE)1 -4.920e+01 1.463e+00 -33.642 < 2e-16 ***
poly(tt, degree = 2, raw = TRUE)2 1.277e-02 3.696e-04 34.539 < 2e-16 ***
s1
                                 2.773e+00 5.927e-02 46.780 < 2e-16 ***
52
                                 -3.902e-01 5.926e-02 -6.584 1.24e-10 ***
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.9065 on 463 degrees of freedom
Multiple R-squared: 0.9964, Adjusted R-squared: 0.9963
F-statistic: 3.171e+04 on 4 and 463 DF. p-value: < 2.2e-16
```

### Plot the trend+periodic model versus the original data

```
plot (co2)
```

lines(tt,pred.complete, col='blue')



• The R-squared (coefficient of determination) is the percent of total variance in the data explained by the model:

$$R^2 = 1 - \frac{RSS}{TSS}$$

$$RSS = \sum_{t=1}^{n} (x_t - \hat{x}_t)^2 ,$$
$$TSS = \sum_{i=t}^{n} (x_t - \overline{x})^2$$

 The adjusted R-squared is a modified version of R-squared which accounts for insignificant predictors in regression models.

#### Airline Passengers data

- R dataset AirPassengers contains the monthly airline passengers from 1949 to 1960 (From Shumway & Stoffer, 2017).
- The time series analysis will be carried out with the ATSA R package.
- The package is installed using install.packages("astsa")



We compute (i) the logarithm (ii) differences of the logarithms and (iii) seasonal differences of order 12 (to remove periodicity).

### Testing the Time Series for Correlations



### Estimating a SARIMA Model

The ACF and PACF plots indicate (p, d, q) = (1, 1, 1) and (P, D, Q) = (0, 1, 1).

sarima(lx,1,1,1,0,1,1,12) # Model A



### Testing residuals for correlations

#### Ljung-Box (Portmanteau) test

- Null hypothesis  $H_0: \rho_{\tau} = 0, \tau = 1, ..., k$ . In portmanteau statistical tests only the null hypothesis is well-defined.
- Sampling function used for L-B test:

$$Q = N(N+2)\sum_{\tau=1}^{k} \frac{\hat{\rho}_{\tau}^{2}}{N-\tau}$$

Q follows the  $\mathcal{X}_k^2$  distribution with k degrees of freedom, i.e.,  $Q \sim \mathcal{X}_k^2$ .



- Rejection zone of  $H_0$  for significance level  $\alpha$ :  $Q > \chi^2_{k;1-\alpha}$ .  $\chi^2_{k;1-\alpha}$  is the critical value of the distribution  $\mathcal{X}^2_k$  that corresponds to the  $1 \alpha$  quantile :  $P(\mathcal{X}^2_k \le \chi^2_{k;1-\alpha}) = 1 \alpha$ .
- Practical rule-of-thumb:
  - For time series without seasonality it is recommended  $k = \min(10, N/5)$ .
  - For periodic time series with period T it is recommended that

 $k = \min(2T, N/5).$ 

sarima(lx,1,1,1,0,1,1,12) # Model A

### Statistics for Model A: (p, d, q) = (1, 1, 1), $(P, D, Q) = (0, 1, 1)_{12}$

	Estimate	SE	t.value	p.value
ar1	0.1960	0.2475	0.7921	0.4298
ma1	-0.5784	0.2132	-2.7127	0.0076
sma1	-0.5643	0.0747	-7.5544	0.0000

Table: The AR1 term is not statistically significant (high p-value).



- AIC = -3.678622
- AICc = -3.677179
- BIC = -3.59083

### SARIMA Model B: $(p, d, q) = (1, 1, 0), (P, D, Q) = (0, 1, 1)_{12}$

#### sarima(lx,1,1,0,0,1,1,12) # Model B



52/68

sarima(lx,1,1,0,0,1,1,12) # Model B

### Statistics for Model B: (p, d, q) = (1, 1, 0), $(P, D, Q) = (0, 1, 1)_{12}$

	Estimate	SE	t.value	p.value
ar1	-0.3395	0.0822	-4.129	1e-04
sma1	-0.5619	0.0748	-7.5109	0.0000

Table: The AR1 term is statistically significant.

Selection criteria for Model B: (p, d, q) = (1, 1, 0),  $(P, D, Q) = (0, 1, 1)_{12}$ 

- AIC = -3.675493
- AICc = -3.674777

• BIC = -3.609649 lower than Model A:  $(p, d, q) = (1, 1, 1), (P, D, Q) = (0, 1, 1)_{12}$ 

#### Diagnostic Checks of SARIMA Residuals for Model B



### SARIMA Prediction for Model A: $(1, 1, 1) \times (0, 1, 1)_{12}$

lxp1 = sarima.for(lx, 12, 1,1,1, 0,1,1, 12) # Model A



(GSLAB)

### SARIMA Prediction Model A: $(1, 1, 0) \times (0, 1, 1)_{12}$

lxp2 = sarima.for(lx, 12, 1,1,0, 0,1,1, 12) # Model B



(GSLAB)

### Comparing SARIMA Predictions of Logarithmic Air Passenger Numbers

 Jan
 Feb
 Mar
 Apr
 May
 Jun
 Jul
 Aug
 Sep
 Oct
 Nov
 Dec

 0.0366
 0.0430
 0.0474
 0.0512
 0.0547
 0.0580
 0.0611
 0.0668
 0.0695
 0.0721
 0.0746

 0.0369
 0.0444
 0.0528
 0.0594
 0.0656
 0.0711
 0.0763
 0.0811
 0.0857
 0.0900
 0.0941
 0.0981

Table: Standard Prediction Errors for 12 months in 1961. Row 1: (p, d, q) = (1, 1, 1),  $(P, D, Q) = (0, 1, 1)_{12}$ . Row 2: (p, d, q) = (1, 1, 0),  $(P, D, Q) = (0, 1, 1)_{12}$ .

Selection Criteria prefer Model B.

Model A (higher complexity) has smaller standard errors than Model B. Is this good or bad?

### Nonlinear time series models

# Threshold Auto-regressive (TAR) models

- In threshold auto-regressive (TAR) models the process transitions between regimes with different values of AR parameters based on certain threshold(s).
- In self-exciting TAR (SETAR) models, the regime for  $x_t$  is determined by the value of the time series in the past,  $x_{t-d}$ , where d > 0 is the delay parameter.
- System of equations for K regimes:
- $x_t = \sum_{i=1}^p \phi_i^{(k)} x_{t-i} + \sigma_{\epsilon}^{(k)} \epsilon_t$ , if  $x_{t-d} \in R_k$

• 
$$R_k = (r_{k-1}, r_k)$$
 where  $-\infty \triangleq r_0 < r_1 < r_2 < \ldots < r_K \triangleq \infty$ 

# Threshold Auto-regressive (TAR) models

### Example of SETAR Model

 $x_t = 3 + 0.9 x_{t-1} + \epsilon_t \text{ if } x_{t-1} \le 0, \quad x_t = -2 - 0.5 x_{t-1} + 0.5 \epsilon_t \text{ if } x_{t-1} > 0$ 



# Threshold Auto-regressive (TAR) models

### Other TAR Models

- It is possible to switch regimes for x<sub>t</sub> based on the values of an exogenous variable z<sub>t</sub> (exogenous TAR models).
- Aleatory (random) or periodic regime switching are also possible.
- Transitions in TAR models are abrupt.
- Smooth-transition auto-regressive models (STAR) involve variable-rate state transitions.

# Markov Switching Model for Time Series

### Markov or probabilistic regime switching

- The model contains regimes that can represent very different behavior.
- Change in behavior represents structural changes in the time series.
- Regime switching allows capturing dynamic patterns.
- Markov property: The transition probability to any particular state (regime) depends only on the current state, not any of the past states.
- In structural models (cf. Chow's test) there is usually one regime change. In Markov switching, multiple changes are likely.

### Markov Switching Model for Time Series

### Two-state Markov Switching AR Model (Matlab)

- Assume two regimes, denoted by  $s_t = 0$  and  $s_t = 1$ .
- The two-state Markov switching model is then determined by

$$\begin{aligned} x_t &= \sum_{i=1}^{p_0} \phi_i^{(0)} x_{t-i} + \sigma_{\epsilon}^{(0)} \epsilon_t, \text{ if } s_t = 0, \text{ AR}(p_0) \text{ model} \\ x_t &= \sum_{i=1}^{p_1} \phi_i^{(1)} x_{t-i} + \sigma_{\epsilon}^{(1)} \epsilon_t, \text{ if } s_t = 1, \text{ AR}(p_1) \text{ model} \end{aligned}$$

• Markov chain transition model,  $p_{i,j} = \text{Prob}(s_t = i \mid s_{t-1} = j)$ :

$$\mathbb{P} = \left[ \begin{array}{cc} p_{0,0} & p_{0,1} \\ p_{1,0} & p_{1,1} \end{array} \right]$$

# Example of Markov Switching Model

```
P = [0.9 \ 0.1; \ 0.3 \ 0.7];
mc=dtmc(P,'StateNames', ...
["Expansion" "Recession"]);
mdl1=arima('Constant',5, ...
'AR', [0.3 0.2], 'Variance', 2);
mdl2=arima('Constant',-5, ...
 'AR',0.1, 'Variance',1);
mdl=[mdl1; mdl2];
Mdl=msVAR(mc,mdl);
[y,e,sp] = simulate(Mdl,50);
figure
subplot(3,1,1); plot(y)
ylabel('Response');
grid on
subplot(3,1,2); plot(e)
ylabel('Innovation')
grid on
subplot(3,1,3); plot(sp,'m')
ylabel('State'); yticks([1 2])
yticklabels(Mdl.StateNames)
```



### **ARCH Models**

The ARCH model involves multiplicative noise with variance given by an AR(p) model

$$x_t = \sigma_t \epsilon_t$$
, where  $\sigma_t^2 = \phi_0 + \phi_1 x_{t-1}^2 + \dots \phi_p x_{t-p}^2$ ,  $\sigma_t$ : volatility,

 $\epsilon_t \sim N(0,1)$ : standard Gaussian white noise (GWN)

Constraints:  $\phi_0 > 0, \ \phi_i \ge 0, \ i = 1, \dots, p$ 

#### Generalized ARCH (GARCH) Models

In the generalized ARCH model the volatility has double AR dependence:

$$x_t = \sigma_t \epsilon_t$$
, where  $\sigma_t^2 = \phi_0 + \sum_{i=1}^p \phi_i x_{t-i}^2 + \sum_{j=1}^q \psi_j \sigma_{t-j}^2$ 

Constraints:  $\phi_0 > 0, \ \phi_i \ge 0, \ i = 1, \dots, p \ , \ \psi_j \ge 0, \ j = 1, \dots, q$ 

### ARCH(1) Model

ARCH(1) Equation

$$X_t = \sqrt{\phi_0 + \phi_1 X_{t-1}^2} \,\epsilon_t$$

• Conditional mean (on  $I_{t-1}$ , information at time t-1):

$$\mathbb{E}\left[X_t \mid I_{t-1}\right] = \mathbb{E}[\epsilon_t] \sqrt{\phi_0 + \phi_1 x_{t-1}^2} = 0.$$

Unconditional mean

$$\mathbb{E}[X_t] = \mathbb{E}[\epsilon_t] \mathbb{E}\left[\sqrt{\phi_0 + \phi_1 X_{t-1}^2}\right] = 0,$$

Conditional variance

$$\begin{aligned} \mathbb{V}\mathrm{ar}[X_t \mid I_{t-1}] &= \mathbb{E}\left[X_t^2 \mid I_{t-1}\right] - \mathbb{E}^2\left[X_t \mid I_{t-1}\right] = \mathbb{E}[\epsilon_t^2] \left(\phi_0 + \phi_1 x_{t-1}^2\right) - 0 \\ &= \phi_0 + \phi_1 x_{t-1}^2 \end{aligned}$$

### Variance of ARCH(1) Model

- The conditional variance depends on time. However, this does not mean that the ARCH model is non-stationary.
- Unconditional variance (using the law of total variance)

$$\begin{aligned} \mathbb{V}\operatorname{ar}[X_t] &= \mathbb{E}[\mathbb{V}\operatorname{ar}[X_t \mid I_{t-1}]] + \mathbb{V}\operatorname{ar}\left[\mathbb{E}\left(X_t \mid I_{t-1}\right)\right] \\ &= \mathbb{E}\left[\phi_0 + \phi_1 X_{t-1}^2\right] + \mathbb{V}\operatorname{ar}(0) \\ &= \phi_0 + \phi_1 \mathbb{E}[X_{t-1}^2] = \phi_0 + \phi_1 \mathbb{V}\operatorname{ar}[X_{t-1}] \end{aligned}$$

• Assuming stationarity  $Var[X_t] = Var[X_{t-1}]$  and thus

$$\mathbb{V}\mathrm{ar}[X_t] = \frac{\phi_0}{1 - \phi_1} \,.$$

• Hence, stationarity is possible only if  $\phi_1 < 1$ .

### The ARCH(1) process is leptokurtic

Leptokurtic means thinner core, fatter tails than the Gaussian

• Let  $\mathbb{K}urt(X_t)$  represent the kurtosis of  $X_t$ 

$$\mathbb{K}\mathrm{urt}(X_t) \triangleq \frac{\mathbb{E}[X_t^4]}{\mathbb{V}\mathrm{ar}^2[X_t]}$$

The fourth-order moment is obtained using the law of total expectation as

$$\begin{split} \mathbb{E}[X_t^4] = & \mathbb{E}\left[\mathbb{E}\left(X_t^4 \mid I_{t-1}\right)\right] = \mathbb{E}\left[\left(\phi_0 + \phi_1 X_{t-1}^2\right)^2\right] \mathbb{E}[\epsilon_t^4] \\ = & 3\left(\phi_0^2 + 2\phi_0\phi_1 \mathbb{E}[X_{t-1}^2] + \phi_1^2 \mathbb{E}[X_{t-1}^4]\right) \\ = & 3\left(\phi_0^2 + 2\phi_0\phi_1 \mathbb{Var}[X_t] + \phi_1^2 \mathbb{E}[X_t^4]\right) \\ \Rightarrow & \mathbb{E}[X_t^4] = 3\frac{\phi_0^2}{1 - \phi_1}\frac{1 + \phi_1}{1 - 3\phi_1^2}. \quad \text{Also use } \mathbb{Var}[X_t] = \frac{\phi_0}{1 - \phi_1} \end{split}$$

• Hence, it follows that  $\mathbb{K}\mathrm{urt}(X_t) = 3\frac{1-\phi_1^2}{1-3\phi_1^2} > 3$ .

### Simulation of GARCH(1,1) model

```
\sigma_t^2 = 0.01 + 0.25 \, x_{t-i}^2 + 0.7 \, \sigma_{t-j}^2
```

```
% Step 1. Specify a GARCH model
Mdl = garch('Constant',0.01,'GARCH',0.7,'ARCH',0.25)
```

```
% Step 2. Simulate from the model without using presample data
rng default; % For reproducibility
[Vn, Yn] = simulate(Mdl, 100, 'NumPaths', 5);
Vn(1,:) % Display variances
figure
subplot (2, 1, 1)
plot (Vn)
xlim([0,100])
title ('Conditional Variances')
subplot (2, 1, 2)
plot (Yn)
xlim([0,100])
title('Innovations')
```

#### Simulation of GARCH(1,1) model

 $\sigma_t^2 = 0.01 + 0.25 \, x_{t-i}^2 + 0.7 \, \sigma_{t-j}^2$ 

%Step 3. Simulate from the model using presample data.

```
[Vw,Yw] = simulate(Mdl,100,'NumPaths',5, 'E0',0.05,'V0',0.001);
Vw(1,:)
```

#### figure

```
subplot(2,1,1)
plot(Vw)
xlim([0,100])
title('Conditional Variances')
```

```
subplot(2,1,2)
plot(Yw)
xlim([0,100])
title('Innovations')
```

### Simulation of GARCH(1,1) model $\sigma_t^2 = 0.01 + 0.25 x_{t-i}^2 + 0.7 \sigma_{t-i}^2$ %Step 4. Look at the unconditional variances. sig2 = 0.01/(1-0.7-0.25);rng default; [V,Y] = simulate(Mdl, 500, 'NumPaths', 10000); figure **plot** (var(Y, 0, 2), 'Color', [.7, .7, .7], 'LineWidth', 1.5) xlim([0,500]) hold on plot(1:500, ones(500, 1) \*sig2, 'k--', 'LineWidth', 2) legend('Simulated','Theoretical','Location','NorthWest') title('Unconditional Variance') hold off

### The ARCH(1) process is uncorrelated

• We use the Law of Iterated Expectation (LIE):

$$\mathbb{E}[X_t X_{t-1}] = \mathbb{E}_{I_{t-1}} \left[ \mathbb{E} \left[ X_t X_{t-1} \mid I_{t-1} \right] \right] \\ = \mathbb{E}_{I_{t-1}} \left[ X_{t-1} \mathbb{E} \left[ X_t \mid I_{t-1} \right] \right] = \mathbb{E} \left[ X_t \cdot 0 \right] = 0.$$

- Similarly, we can show that  $\mathbb{E}[X_t X_{t-n}] = 0$  for  $n \ge 1$ .
- Since  $\mathbb{E}[X_t] = 0$ ,  $\mathbb{V}ar[X_t] = \phi_0/(1 \phi_1)$ ,  $\mathbb{E}[X_t X_{t-n}] = 0$  for  $n \ge 1$ , the process  $X_t$  is stationary if  $0 < \phi_1 < 1$ .
- Similarly, ARCH(p) processes are stationary if  $\sum_{i=1}^{p} \phi_i < 1$ .

### ARCH(1) and Efficient Market Hypothesis

- Hence,  $X_t$  cannot be predicted based on the past.
- This property resonates with the efficient market hypothesis (EMH).
- EMH: Market operation is efficient, meaning it is not possible to make excess profits by investing since stocks are fairly and accurately priced at any time.
- EMH is controversial. Some investors (e.g., Warren Buffett) consistently outperform the markets.
## Auto-regressive conditional heteroskedasticity (ARCH)

### Nonetheless, $X_t^2$ can be predicted!

- Note that  $\mathbb{E}[X_t^2] = \mathbb{V}ar[X_t]$ .
- The conditional (on the past) variance of  $X_t^2$  is:  $Var[X_t \mid I_{t-1}] = \phi_0 + \phi_1 x_{t-1}^2$ .

#### The R script arch\_wmt applies GARCH modeling to Walmart stock prices.

### Auto-regressive conditional heteroskedasticity (ARCH)

```
library("readx1")
library("tseries")
WMT <- read excel("Dionisis laptop/TEACHING HMMY/MLDS Course Time Series Analysis/R-coc
View(WMT)
p = ts(WMT[.2]) \# opening price
ret = diff(log(p)) # return
ret2 = ret^2 # squared return
plot(p)
arima(x = p. order = c(1, 0, 0)) # autoregressive coefficient very close to 1
plot(ret) # plot return
mean(ret) # return series is not trending
Box.test (ret. lag = 1. type="Liung") # Test for correlations
# Based on Box test H0 cannot be rejected
fit_r <- arima(x = ret, order = c(1, 0, 0)) # Apply ARIMA to returns. Can we predict?
# The above command gives an OLS estimate
summarv(fit r)
r1 = acf(ret) # ACF of returns
r2 = acf(ret2) # ACF of squared returns
Box.test (ret2, lag = 1, type="Ljung") # Reject null hypothesis of GWN
arima(x = ret2, order = c(1, 0, 0), method = "CSS") # ARIMA model of squared returns
qfit < -qarch(x = ret, order = c(0, 1)) # Fit an ARCH(1) model using MLE
summary(gfit) # Liung-Box shows that the residuals have no remaining correlations
# Jarque-Berra test shows that the normal hypothesis. HO, is rejected.
qfit2 < -qarch(x = r, order = c(1, 1))
summarv(qfit2) # ARCH(1.1) is sufficient
```

### Auto-regressive conditional heteroskedasticity (ARCH)

- ARMA models are used to model conditional mean.
- ARCH models are used to represent conditional variance.
- ARCH models capture volatility clustering: i.e., periods of large (or small) values of volatility tend to be followed by periods of similar values.

# A Summary of Time Series Analysis

- Exploratory data analysis
  - Are there correlations in the data?
  - White noise
- Model determination
  - correlations or signal + noise
  - identify p, d, q and seasonal components
  - Nonlinear dependence? SETAR, ARCH/GARCH ?
  - Estimating parameters of the signal and for residuals
- Model checks
  - Testing of residuals
  - Examining realizations and their characteristics
  - Obtain forecasts, spectral estimates, use cross-validation
- General remarks
  - Nonlinear models are more flexible but identification is difficult and over-fitting is possible.

### Multivariate time series

• The cross-covariance functions of two scalar processes  $X_i(t)$  and  $X_j(t)$  are defined by the expectation (for i, j = 1, ..., D)

 $C_{i,j}(t,t+\tau) \triangleq \mathbb{C}\mathsf{ov}\{X_i(t+\tau), X_j(t)\} = \mathbb{E}[X'_i(t+\tau) X'_j(t)],$ 

where  $\tau$  is the temporal lag.

- For i = j the auto-covariance functions are obtained.
- The processes  $X_i(t)$  and  $X_j(t)$  are jointly stationary (in the weak sense) if (i) they are both stationary and (ii) the cross-covariance function  $C_{i,j}(\cdot)$  is purely a function of  $\tau$ .
- The cross-correlation functions (CCF) are given by

$$\rho_{i,j}(\tau) = \frac{C_{i,j}(\tau)}{\sigma_i \sigma_j},$$

where  $\sigma_i^2 = C_{i,i}(0) = \mathbb{V}arX_i(t)$  is the variance of the *i*-th component.

• The auto-correlation functions are defined by  $\rho_{i,i}(\tau) = \frac{C_{i,i}(\tau)}{\sigma_i^2}$ .

• Sample cross-covariance:

$$\hat{C}_{i,j}(k) = \frac{1}{N-k} \sum_{n=k+1}^{N} x_{i,n-k} x_{j,n} - \overline{x_i} \, \overline{x_j}, \quad \text{for } k \in \{0, \pm 1, \pm 2, \ldots\}$$

Sample cross-correlations:

$$\hat{\rho}_{i,j}(k) \triangleq \frac{\hat{C}_{i,j}(k)}{\sqrt{\hat{C}_{i,i}(0)\,\hat{C}_{j,j}(0)}}$$

Symmetry property:

$$C_{i,j}(\tau) = C_{j,i}(-\tau)$$

• Contrast with the ACF:

$$C_{i,i}(\tau) = C_{i,i}(-\tau)$$

### Review of Bochner's theorem for scalar process

A function  $C(\tau) : \mathbb{R} \to \mathbb{R}$  is nonnegative definite if the following conditions hold:

- The Fourier transform  $\widetilde{C}(\omega) \triangleq \operatorname{FT}[C(\tau)]$  exists.
- 2  $\widetilde{C}(\omega) \ge 0$ , for all  $\omega \in \mathbf{R}$ .
- Solution The spectral integral  $\frac{1}{2\pi} \int_{\mathbf{R}} d\omega \widetilde{C}(\omega)$  is a finite number which represents the variance of the stochastic process.

#### Cramer's Theorem for multivariate processes

 $C(\tau) : \mathbb{R} \to \mathbb{R}^D \times \mathbb{R}^D$  is a valid matrix covariance function for a continuous, stationary, stochastic vector process, if the following conditions hold:

- The FTs  $\widetilde{C}_{i,j}(\omega)$  exist for all  $i, j = 1, \ldots, D$ .
- 2 Non-negative *auto-spectral densities*:  $\tilde{C}_{i,i}(\omega) \ge 0$  for all i = 1, ..., D.
- The integrals of the auto-spectral densities are finite:

$$\int_{\mathsf{R}} \frac{d\omega}{2\pi} \, \tilde{C}_{i,i}(\omega) = \sigma_i^2, \quad i = 1, \dots, D.$$

- The cross-spectral densities have bounded variation, i.e., for all  $i \neq j = 1, ..., D \int_{\mathsf{R}} d\omega \left| \widetilde{C}_{i,j}(\omega) \right|$  are finite.
- So The spectral density matrix  $\widetilde{\mathbf{C}}(\omega)$ ,  $[\widetilde{\mathbf{C}}(\omega)]_{i,j} = \widetilde{C}_{i,j}(\omega)$ , is nonnegative definite for all  $\omega \in \mathbf{R}$ .

#### A Simple example of CCF

- Simple system with time-delayed interaction
- $Y_t = A X_{t-\ell} + \epsilon_t$
- X leads Y if  $\ell > 0$  and X lags Y if  $\ell < 0$ , respectively.
- $C_{Y,X}(\tau) = \mathbb{C}\text{ov}(Y_{t+\tau}, X_t) =$   $\mathbb{C}\text{ov}(A X_{t+\tau-\ell} + \epsilon_{t+\tau}, X_t) =$   $A \mathbb{C}\text{ov}(X_{t+\tau-\ell}, X_t) =$  $AC_X(\tau - \ell).$

Simulated system (ccf.R)



### What is the distribution of the sample cross-correlation?

#### Theorem

Large-Sample Probability Distribution of Cross-Correlation:

Let us consider the time series  $X_t$  and  $Y_t$ . If  $N \to \infty$ , and at least one of the series  $X_t$ ,  $Y_t$  is independent white noise, then

$$\hat{\rho}_{X,Y}(\tau) \sim \mathcal{N}\left(0, \frac{1}{\sqrt{N}}\right)$$

- Nice result, but usually none of the series is i.i.d. See Cross-correlation.R
- If both processes have auto-correlations, the variance of  $\hat{\rho}_{X,Y}(\tau)$  involves an annoying infinite series of ACF products.
- Solution: Prewhitening. See CCF\_synthetic.R

#### Permutation Method for CCF testing

- Null hypothesis: The series  $X_t$  and  $Y_t$  are not correlated.
- What is the probability that the sample cross-correlation,  $\hat{\rho}_{X,Y}(\tau)$  is realized if the null hypothesis is true?
- Generate  $M_s$  randomized states  $X_t^{(m)}$ , where  $m = 1, \ldots, M_s$ .
- These states are obtained from  $M_s$  random permutations of N values of  $X_t$  (or  $Y_t$ ).
- Permutations destroy the temporal ordering.
- Hence, the estimated sample cross-correlation  $\hat{\rho}_{X^{(m)},Y}(\tau)$  based on the shuffled time series  $X_m(t)$  represents random fluctuations.
- The *p*-value for cross-correlation testing at lag  $\tau$  is the percentage of the  $M_s$  permutation states for  $|\hat{\rho}_{X^{(m)},Y}(\tau)| > |\hat{\rho}_{X,Y}(\tau)|$ .
- The observed cross-correlation,  $\hat{\rho}_{X,Y}(\tau)$  is statistically significant if the *p*-value is less than the specified significance level (typically 0.1%-5%).

### Example of permutation testing



(GSLAB)

#### Model properties

- Transfer function models allow describing the interaction (dependence) between time series
- They involve two main assumptions.
- First, the input (source) series influences the output (receiver) series but there is no feedback in the opposite direction.
- Second, the interaction between input and output involves only previous (past) times of the input series.
- Furthermore, it is assumed that both the input and response series are stationary.

### Model equations

- Assume that  $X_t$  is the input and  $Y_t$  is the output.
- TF model:  $Y_t \mu_Y = \sum_{j=0}^{\infty} w_j (X_{t-j} \mu_X) + Z_t$ .
- $\sum_{j=0}^{\infty} |w_j| < \infty$ .
- $Z_t$  is assumed to be a zero-mean residual (error term).
- In practice, the interaction is cut off after some maximum lag T.
- The input series X<sub>t</sub> may also be governed by a stochastic model ARMA(p,q) model.
- The residual  $Z_t$  is uncorrelated with the input process  $X_{t'}$  for all t, t', i.e.,  $Z_t \perp X_{t'}$ .
- The residuals  $Z_t$  and  $Z_{t'}$  are usually correlated.
- In the following, we assume that  $\mu_X = \mu_Y = 0$ .

#### Estimation of cross-covariance

• Assume that  $X_t$  is stationary.

- $c_{Y,X}(\tau) = \mathbb{C}\mathrm{ov}\{Y_{t+\tau}, X_t\} = \sum_{j=0}^{\infty} w_j \mathbb{C}\mathrm{ov}\{X_{t+\tau-j}, X_t\} + \mathbb{C}\mathrm{ov}\{Z_{t+\tau}, X_t\}$
- Since  $Z_t \perp X_{t'}$  for all t, t', the cross-covariance  $c_{Y,X}(\tau)$  is given by
- $c_{Y,X}(\tau) = \sum_{j=0}^{\infty} w_j c_{X,X}(\tau-j).$
- The auto-covariance function can be replaced by its sample estimates.
- If  $X_t$  is white noise then  $c_{Y,X}(\tau) = w_{\tau}c_{X,X}(0)$ .
- Hence,  $w_{\tau} = \frac{c_{Y,X}(\tau)}{c_{X,X}(0)}$ .
- However, this is not the typical case since  $X_t$  is usually correlated.

### Considering a Realistic Example

- Assume that X<sub>t</sub> is a stationary process described by an ARMA(p,q) model.
- Consider that both  $X_t$  and  $Y_t$  are centered so that  $\mu_X = \mu_Y = 0$ .
- TF Model:  $Y_t = \sum_{j=m_1}^{m_2} w_j X_{t-j} + Z_t$ ,  $t = m_2 + 1, \dots, n$ .
- Typically,  $z_t$  is an ARMA( $p_z, q_z$ ) model.
- The above is known as the transfer model, the finite distributed lag model, or the dynamic regression model.
- Positive  $m_1, m_2$  indicate that the input  $X_t$  influences the response  $Y_t$  but not vice versa.
- Based on the observations of  $X_t$  we estimate an "optimal" model ARMA(p,q):  $\phi(B)X_t = \theta(B)\epsilon_t$  (omit "hats" for simplicity).

### Pre-whitening using Filtering

- Assuming invertibility of the ARMA process, it holds that  $\epsilon_t = \theta^{-1}(B)\phi(B) X_t \triangleq \pi(B) X_t$ .
- $\pi(B) = 1 \sum_{i=1}^{\infty} \pi_i B^i$  is the linear filter function.
- Then,  $\tilde{X}_t = \pi(B)X_t$  is a white noise series (use  $\tilde{X}_t$  instead of  $\epsilon_t$ ).
- By applying  $\pi(B)$  to both sides of the TF model we obtain:

$$\tilde{Y}_{t} = \sum_{j=m_{1}}^{m_{2}} w_{j} \tilde{X}_{t-j} + \tilde{Z}_{t}, \text{ where}$$

$$\tilde{X}_{t} = X_{t} - \pi_{1} X_{t-1} - \pi_{2} X_{t-2} - \dots$$

$$\tilde{Y}_{t} = Y_{t} - \pi_{1} Y_{t-1} - \pi_{2} Y_{t-2} - \dots$$

$$\tilde{Z}_{t} = Z_{t} - \pi_{1} Z_{t-1} - \pi_{2} Z_{t-2} - \dots$$

• Since  $\tilde{X}_t$  is white noise, it holds that  $w_{\tau} = \frac{c_{\tilde{Y},\tilde{X}}(\tau)}{c_{\tilde{X},\tilde{X}}(0)}$ , for  $\tau = m_1, \ldots, m_2$ .

#### Summary

- Sind an appropriate ARMA(p,q) model for the input series  $X_t$ .
- **2** Pre-whiten the input data with the corresponding linear filter:  $X_t \mapsto \tilde{X}_t$ .
- Solution Transform the output data  $Y_t \mapsto \tilde{Y}_t$  by applying the same filter as for the input data.
- Construct cross-correlation plots between  $\tilde{X}_t$  and  $\tilde{Y}_t$ .
- Based on the CCF plots deduce candidate values for the interaction lags m<sub>1</sub> and m<sub>2</sub>.
- Solution Using the estimated  $m_1$  and  $m_2$ , construct fits of the data to  $Y_t = \sum_{j=m_1}^{m_2} w_j X_{t-j} + Z_t$ ,  $t = m_2 + 1, \dots, n$  for different model orders  $p_z$  and  $q_z$  of the ARMA residuals  $Z_t$ .
- Choose the "optimal" model suggested by an appropriate selection criterion (e.g., lowest AIC).

### Multivariate ARMA models

### Multivariate ARMA model for stationary time series

Vector time series x<sub>t</sub>

$$\mathbf{x}_t = (x_{1,t}, x_{2,t}, \dots, x_{D,t})^{\top}, \quad -\infty < t < \infty$$

VARMA equation

$$\mathbf{x}_t = \phi_1 \mathbf{x}_{t-1} + \phi_2 \mathbf{x}_{t-2} + \dots + \phi_p \mathbf{x}_{t-p} \ + \boldsymbol{\epsilon}_t - \boldsymbol{ heta}_1 \boldsymbol{\epsilon}_{t-1} - \boldsymbol{ heta}_2 \boldsymbol{\epsilon}_{t-2} - \dots - \boldsymbol{ heta}_q \boldsymbol{\epsilon}_{t-q} \; ,$$

Matrices of auto-regressive and moving average components

 $\phi_k: D \times D$  matrix,  $k = 1, \ldots, p$ 

 $\boldsymbol{\theta}_k: D \times D$  matrix,  $k = 1, \ldots, q$ 

• Vector Gaussian white noise:  $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{2,t}, \dots, \epsilon_{D,t})^{\top}$ 

$$\mathbb{E}[\boldsymbol{\epsilon}_t] = 0, \ \mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t'}^\top] = 0, \ t \neq t', \ \mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top] = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$$

## Multivariate ARMA models

#### Example of VARMA(2,1) system

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} \phi_{2,11} & \phi_{2,12} \\ \phi_{2,21} & \phi_{2,22} \end{pmatrix} \begin{pmatrix} x_{1,t-2} \\ x_{2,t-2} \end{pmatrix} + \\ \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \begin{pmatrix} \theta_{1,11} & \theta_{1,12} \\ \theta_{1,21} & \theta_{1,22} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix}$$

• VARMA models: Complicated estimation procedure.

### VAR(p) Models

$$\mathbf{x}_t = \phi_1 \mathbf{x}_{t-1} + \phi_2 \mathbf{x}_{t-2} + \dots + \phi_p \mathbf{x}_{t-p} + \boldsymbol{\epsilon}_t$$

- VAR(p) models are VARMA(p,0) models.
- Parameter estimation possible by multivariate least squares and maximum likelihood.

## Multivariate ARMA models

#### Example of VAR(2) system

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \phi_{2,11} & \phi_{2,12} \\ \phi_{2,21} & \phi_{2,22} \end{pmatrix} \begin{pmatrix} x_{t-2} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

- The diagonal elements of the matrices φ<sub>k</sub> couple each time series to its previous values.
- $\phi_{1,11}$  and  $\phi_{2,11}$  are lag-1 and lag-2 coefficients for the input series  $X_t$ .
- $\phi_{1,22}$  and  $\phi_{2,22}$  are lag-1 and lag-2 coefficients for the output series  $Y_t$ .
- The lower off-diagonal terms represent the influence of  $X_t$  on  $Y_t$  (also known as transfer mechanism).
- The upper off-diagonal terms measure the influence of  $Y_t$  on  $X_t$  (also known as feedback).

- The brain, climate, flock behavior, and finance, are all examples of complex systems.
- Complex systems are driven, nonlinear, dissipative, and comprise many interacting parts.
- How to model, predict, and control responses in complex systems?
- Causality analysis offers a data-driven approach for studying interactions in such systems.



- Given two time series  $X_t$  and  $Y_t$ , are there interactions of the form  $X_t \rightarrow Y_t, Y_t \rightarrow X_t, X_t \leftrightarrow Y_t$  (arrows imply directional interaction), or are  $X_t$  and  $Y_t$  independent ?
- Linear correlation measures (e.g., Pearson correlation, cross-correlation function) are inadequate for the following reasons:
  - Lack of correlation does not imply lack of dependence.
  - 2 Correlation is symmetric, i.e.,  $R_{X,Y} = R_{Y,X} \implies$  no directionality.
- Cause-effect relationships (Eichler, PRSA, 2013).
  - **1** *Temporal precedence:* a cause precedes its effects in time.
  - *Physical influence:* Manipulations of the cause change the effects.

### A Simple example

• Solve numerically the following VAR(2) system:

$$x_t = 0.2x_{t-1} - 0.4x_{t-2} + \epsilon_t$$

$$y_t = 0.25y_{t-1} + 1.5x_{t-1} + \epsilon'_t,$$

where  $\epsilon_t$  and  $\epsilon'_t$  are independent innovation terms (Gaussian white noise).

• There is only  $Y_t \rightarrow X_t$  dependence in the above system.

• However, the correlation coefficients  $R_{X,Y}$  and  $R_{Y,X}$  are both non-zero.

### Some Measures used in Causality Analysis

- Wiener-Granger causality
- Transfer entropy (Schreiber, 2000), which is a directed version of Shannon's mutual information
- For Gaussian variables transfer entropy ⇔ Granger causality
- Nonlinear (kernel-based) Granger causality (Amblard et al., 2012; Zaremba & Aste, 2014)
- Convergent cross mapping (CCM) (Sugihara et al., 2012)

# Wiener-Granger Prediction

#### Main ideas

- Granger prediction aims to determine whether including information from a time series *Y* in a different time series *X* improves the predictions of *X*—compared to those obtained without information from *Y*.
- First, an AR model is constructed for *X*. The fit of this model is assessed by means of the estimated innovation variance σ<sup>2</sup><sub>ϵ.x</sub>.
- Then, a VAR model is constructed for *X* which includes both *X* and *Y*. The innovation variance,  $\sigma_{\epsilon,y \to x}^2$ , for this model is also calculated.
- If  $\sigma_{\epsilon,y\to x}^2$  is significantly lower than  $\sigma_{\epsilon,x}^2$ , then *Y* Granger causes *X*.
- Granger causality essentially refers to directed functional interactions or information flow.

### Wiener-Granger Prediction

#### The Granger prediction equations for two time series

$$x_t = \sum_{i=1}^{P} a_i x_{t-i} + \epsilon_{x,t}, \quad \epsilon_{x,t} \sim N(0, \sigma_{\epsilon,x}^2)$$
(7)

$$y_t = \sum_{i=1}^p b_i y_{t-i} + \epsilon_{y,t}, \quad \epsilon_{y,t} \sim N(0, \sigma_{\epsilon,y}^2)$$
(8)

$$x_t = \sum_{i=1}^p a'_i x_{t-i} + \sum_{i=1}^p c_i y_{t-i} + \epsilon_{y \to x,t}, \quad \epsilon_{y \to x,t} \sim N(0, \sigma_{\epsilon,y \to x}^2)$$
(9)

$$y_t = \sum_{i=1}^p b'_i y_{t-i} + \sum_{i=1}^p d_i x_{t-i} + \epsilon_{x \to y,t}, \quad \epsilon_{x \to y,t} \sim N(0, \sigma_{\epsilon,x \to y}^2)$$
(10)

$$GP_{y \to x} = \log\left(\frac{\sigma_{\epsilon,x}^2}{\sigma_{\epsilon,y \to x}^2}\right), \quad GP_{x \to y} = \log\left(\frac{\sigma_{\epsilon,y}^2}{\sigma_{\epsilon,x \to y}^2}\right), \tag{11}$$

If  $\operatorname{GP}_{y \to x} \approx 0 \implies Y$  does not Granger-predict X.

#### Generate two time series based on AR(2) model

```
close all; clearvars;
L = 250; % Length of time series
p = 2; % Order of AR model
% define time series X and Y
x=randn(2,1); y=randn(2,1); % random initial conditions
for i=3: L
    x(i) = .2 * x(i-1) - .4 * x(i-2) + randn;
    y(i) = .25 * y(i-1) - .8 * x(i-2) + 1.5 * x(i-1) + randn;
end
figure
plot (x, 'mo-', 'linewid', 1, 'markerface', 'k', 'markersize', 6)
hold on;
plot(y,'go-','linewid',1,'markerface','k','markersize',6)
title('Stationary bivariate autoregression')
axis tight;
```

#### Vector AR(2) model with two components



#### Determine optimal *p* order from data using BIC

```
armaxp = 15; % Maximum AR order tested
x = x(:)'; v = v(:)';
Nr = 1; % One realization used.
% Determine optimal AR model order based on
% Bayesian Information Criterion (BIC)
bicm = zeros(armaxp,1);
for bici=1: armaxp
    [Axy_test, E_test] = armorf([x; y], 1, L, bici);
    bicm(bici) = \log(\det(E_test)) + (\log(L) \star bici \star 2^2)/L;
end
figure;
plot(1:1:armaxp, bicm, 'b--');
hold on;
plot(1:1:armaxp, bicm, 'ro', 'MarkerFaceColor', 'q');
xlabel('AR order')
ylabel('BIC');
[minbic, popt] = min(bicm);
```

#### BIC for Vector AR(2) model with two components



#### Use optimal AR order from BIC to check GP

```
p = popt;
[Ax,Ex] = armorf(x, Nr, L, p);
[Ay,Ey] = armorf(y, Nr, L, p);
[Axy,E] = armorf([x; y], 1, L, p);
y2x =log(Ex/E(1,1));
x2y =log(Ey/E(2,2));
disp(['ln(var AR(p)/var VAR(p) Y-> X)=' num2str(y2x)]);
disp(['ln(var AR(p)/var VAR(p) X-> Y)=' num2str(x2y)]);
```

If y2x > 0, *Y* is said to Granger cause *X*. In practice, we need to perform a statistical test to decide which non-zero values are significant.

Similarly, if x2y > 0, X is said to Granger cause Y.

armorf.m implements Morf's modified Locally Weighted Regression method for AR coefficient estimation: M. Morf, A. Vieira, D.T. Lee, and T. Kailath. Recursive multichannel maximum entropy spectral estimation. *IEEE Trans Geosci Electron.* 1978; 16:85–94.

#### Results of Granger Prediction for the VAR(2) system

```
>> granger_causality
```

```
log(variance AR(p) / variance VAR(p) Y-> X)=0.0075709
```

```
log(variance AR(p) / variance VAR(p) X-> Y)=1.1439
```

```
H0: y does not Granger cause x: p value=0.34062
```

```
Test result: Y does not Granger-cause X
```

```
H0: x does not Granger cause y: p value=0
```

```
Test result: X Granger-causes Y
```

Note that  $y2x \approx 0.0076 > 0$  but the actual value is quite small, and the statistical test rejects  $Y \rightarrow X$  causation.

#### Limitations of Granger causality

- Linearity
- Stationarity
- Time invariance
- Exogenous inputs
- Noise, insufficient sampling rates
- Temporal/spatial aggregation generated by neural data acquisition

P. A. Stokes, P. L. Purdon (2017). A study of problems encountered in Granger causality analysis from a neuroscience perspective. PNAS, 114(34), E7063–E7072. https://doi.org/10.1073/pnas.1704663114. L.

Barnett, A.B. Barrett, A.K. Seth (2018). Misunderstandings regarding the application of Granger causality in neuroscience. PNAS, 115, E6676–7.


## Some References

## For more information ...

- Shumway and Stoffer (2016). Time Series Analysis and Its Applications: With R Examples 3rd Edition, *Springer*, 4th edition.
- Peter Bartlet (2010). Introduction to Time Series
- Marcel Dettling (2017). Applied Time Series Analysis
- D. Kougioumtzis (2019). Time Series Analysis
- Box, G.E.P. and G.M. Jenkins (1970). *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.
- Makridakis, S., S.C. Wheelwright, and R.J. Hyndman (1998). *Forecasting: methods and applications*, New York: John Wiley & Sons.
- Tsay, Ruey, S. (2010). *Analysis of Financial Time Series*, 3rd edition, New York: John Wiley & Sons.