Univariate Time Series Analysis

Klaus Wohlrabe[1] and Stefan Mittnik

[1] Ifo Institute for Economic Research, wohlrabe@ifo.de

SS 2017

Table of content I

Table of content II

Introduction

Time series analysis:

- Focus: Univariate Time Series and Multivariate Time Series Analysis.
- A lot of theory and many empirical applications with real data
- Organization:
  - 25.04. - 30.05.: Univariate Time Series Analysis, *six* lectures (Klaus Wohlrabe)
  - 28.04. - 02.06.: Fridays: Tutorials with Malte Kurz
  - 13.06. - End of Semester: Multivariate Time Series Analysis (Stefan Mittnik)
- $\Rightarrow$ **Lectures and Tutorials are complementary!**

Tutorials and Script

- Script is available at: *moodle* website (see course website)
- Password: armaxgarchx
- Script is available at the day before the lecture (noon)
- All datasets and programme codes
- Tutorial: Mixture between theory and R - Examples

Literature

- **Shumway and Stoffer (2010): Time Series Analysis and Its Applications: With R Examples**
- Box, Jenkins, Reinsel (2008): Time Series Analysis: Forecasting and Control
- Lütkepohl (2005): Applied Time Series Econometrics.
- Hamilton (1994): Time Series Analysis.
- Lütkepohl (2006): New Introduction to Multiple Time Series Analysis
- Chatham (2003): The Analysis of Time Series: An Introduction
- Neusser (2010): Zeitreihenanalyse in den Wirtschaftswissenschaften

Examination

- Evidence of academic achievements: Two hour written exam both for the univariate and multivariate part
- Schedule for the Univariate Exam: tba.

Prerequisites

- Basic Knowledge (ideas) of OLS, maximum likelihood estimation, heteroscedasticity, autocorrelation.
- Some algebra

Software

Where you have to pay:

- STATA
- Eviews
- Matlab (Student version available, about 80 Euro)

Free software:

- R (www.r-project.org)
- Jmulti (www.jmulti.org) (Based on the book by Lütkepohl (2005))

Tools used in this lecture

- standard approach (as you might expected)
- derivations using the whiteboard (not available in the script!)
- live demonstrations (examples) using Excel, Matlab, Eviews, Stata and JMulti
- live programming using Matlab

Outline

- Introduction
- Linear Models
- Modeling ARIMA Processes: The Box-Jenkins Approach
- Prediction (Forecasting)
- Nonstationarity (Unit Roots)
- Financial Time Series

Goals

After the lecture you should be able to ...

- ... identify time series characteristics and dynamics
- ... build a time series model
- ... estimate a model
- ... check a model
- ... do forecasts
- ... understand financial time series

Questions to keep in mind

General Question	**Follow-up Questions**
All types of data	
How are the variables defined?	What are the units of measurement? Do the data comprise a sample? If so, how was the sample drawn?
What is the relationship between the data and the phenomenon of interest?	Are the variables direct measurements of the phenomenon of interest, proxies, correlates, etc.?
Who compiled the data?	Is the data provider unbiased? Does the provider possess the skills and resources to enure data quality and integrity?
What processes generated the data?	What theory or theories can account for the relationships between the variables in the data?
Time Series data	
What is the frequency of measurement	Are the variables measured hourly, daily monthly, etc.? How are gaps in the data (for example, weekends and holidays) handled?
What is the type of measurement?	Are the data a snapshot at a point in time, an average over time, a cumulative value over time, etc.?
Are the data seasonally adjusted?	If so, what is the adjustment method? Does this method introduce artifacts in the reported series?

Table of content I

Table of content II

Goals and methods of time series analysis

The following section partly draws upon Levine, Stephan, Krehbiel, and Berenson (2002), *Statistics for Managers*.

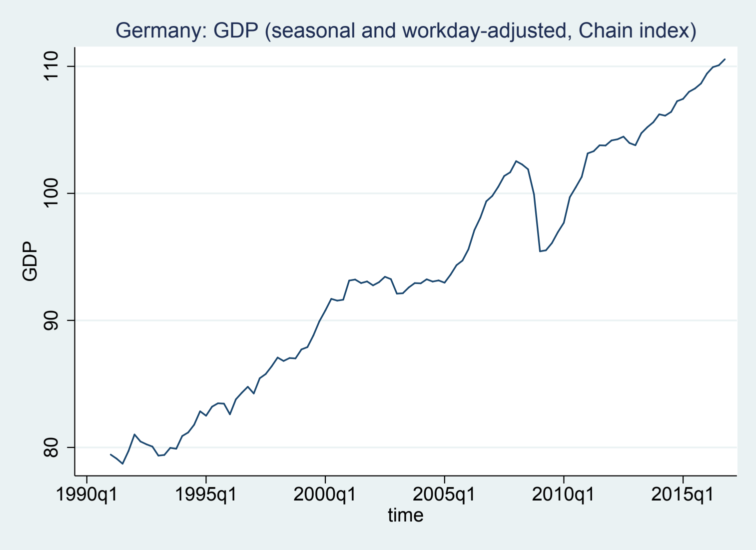
Goals and methods of time series analysis

- understanding time series characteristics and dynamics
- necessity of (economic) forecasts (for policy)
- time series decomposition (trends vs. cycle)
- smoothing of time series (filtering out noise)
    - moving averages
    - exponential smoothing

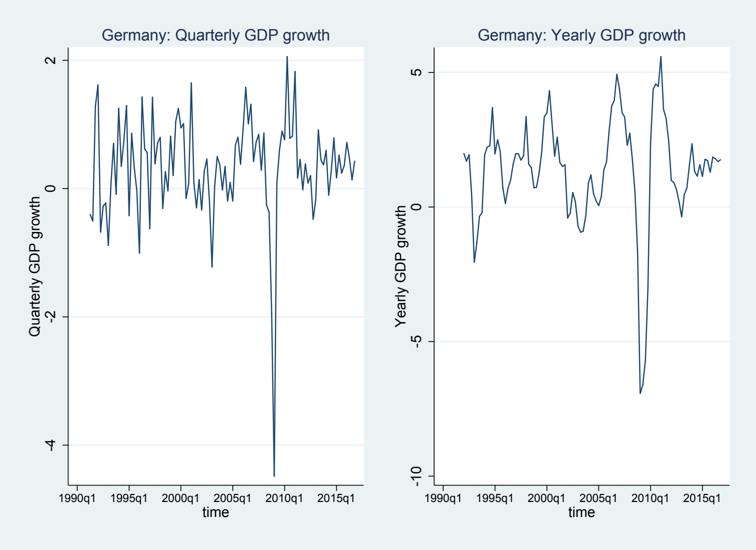
Time Series

- A time series is timely ordered sequence of observations.
- We denote y_t as an observation of a specific variable at date t.
- A time series is list of observations denoted as $\{y_1, y_2, \ldots, y_T\}$ or in short $\{y_t\}_{t=1}^{T}$.
- **What are typical characteristics of times series?**

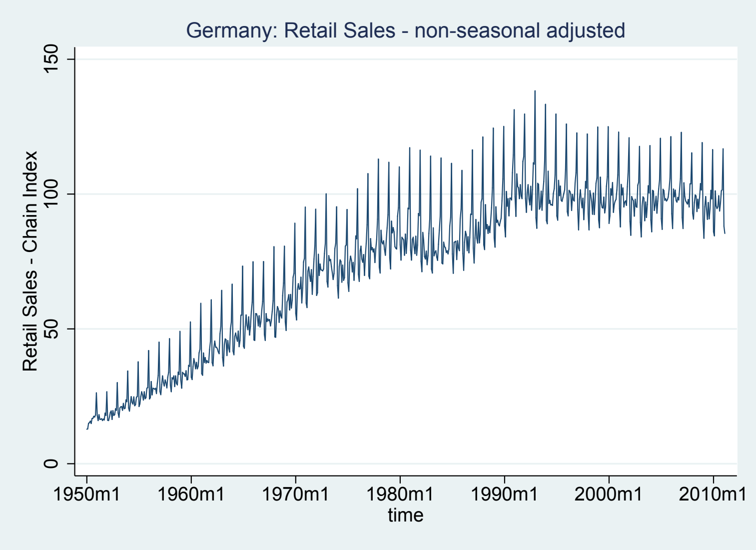
Economic Time Series: GDP I



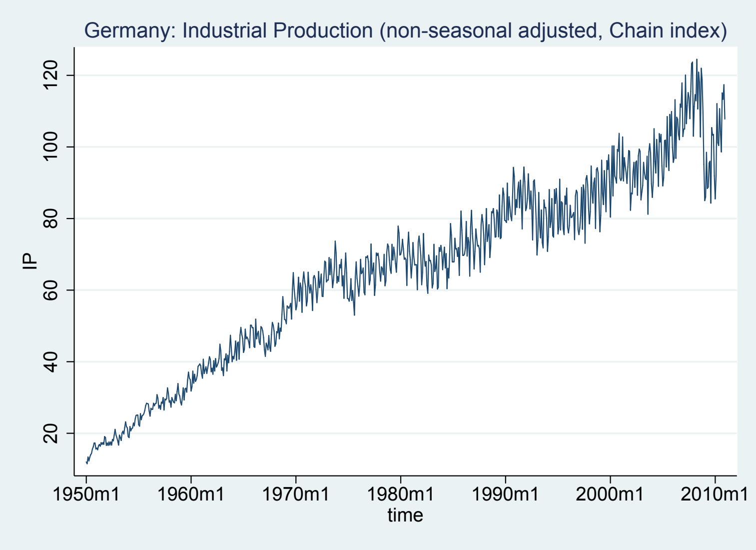
Economic Time Series: GDP II



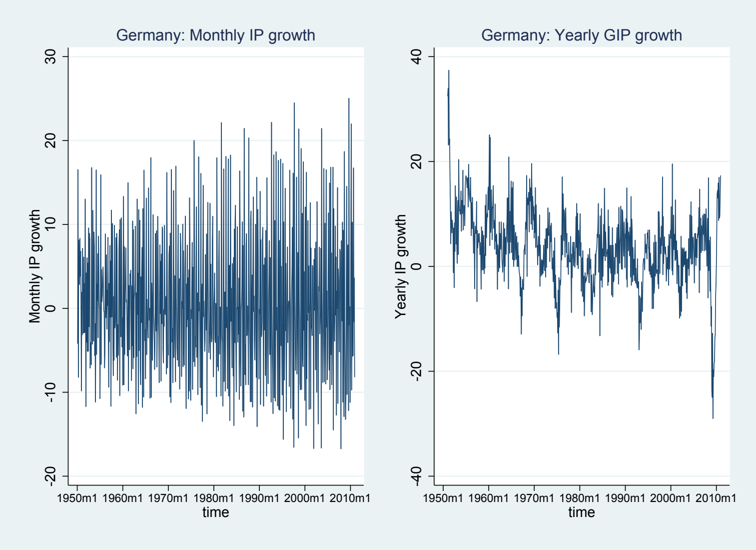
Economic Time Series: Retail Sales



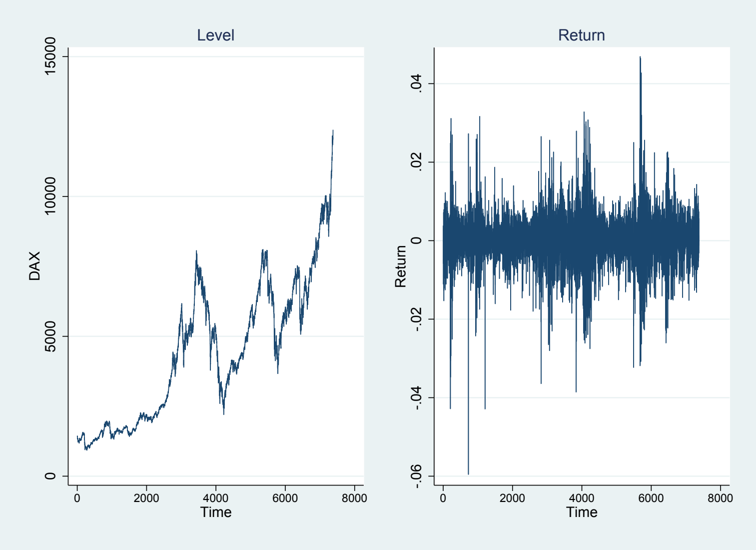
Economic Time Series: Industrial Production



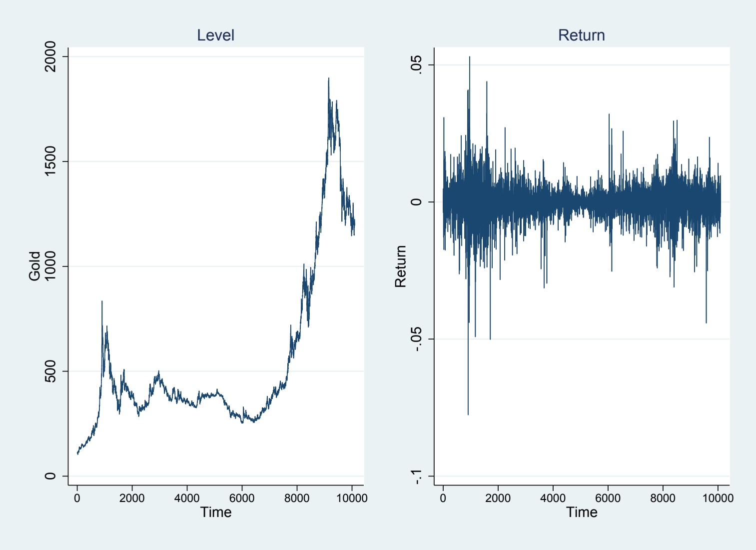
Economic Time Series: Industrial Production



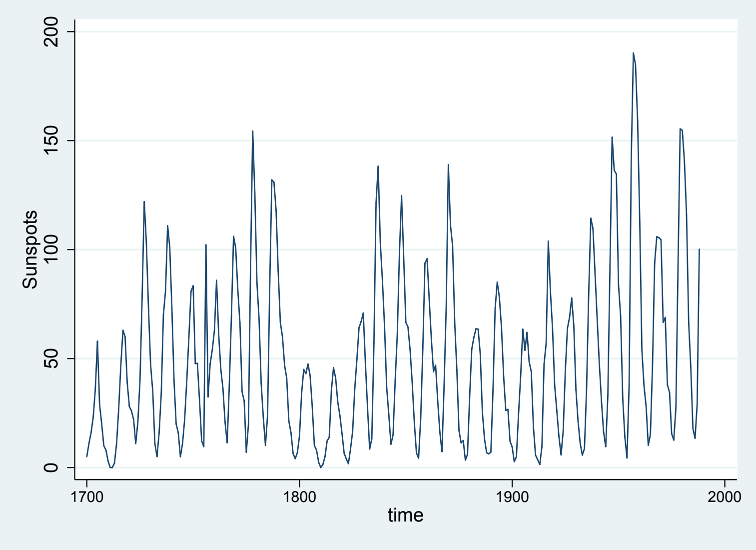
Economic Time Series: The German DAX



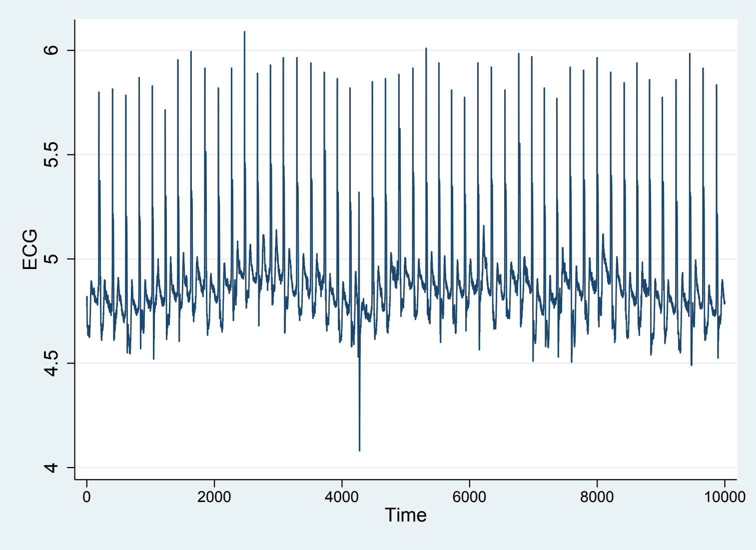
Economic Time Series: Gold Price



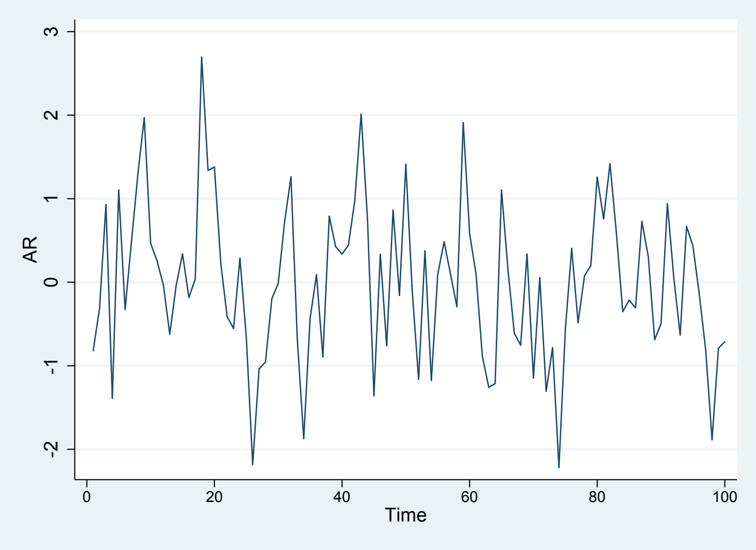
Further Time Series: Sunspots



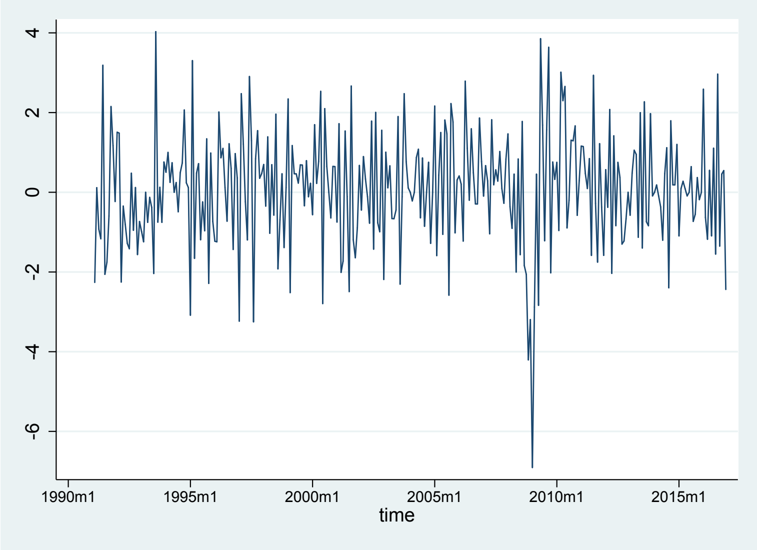
Further Time Series: ECG



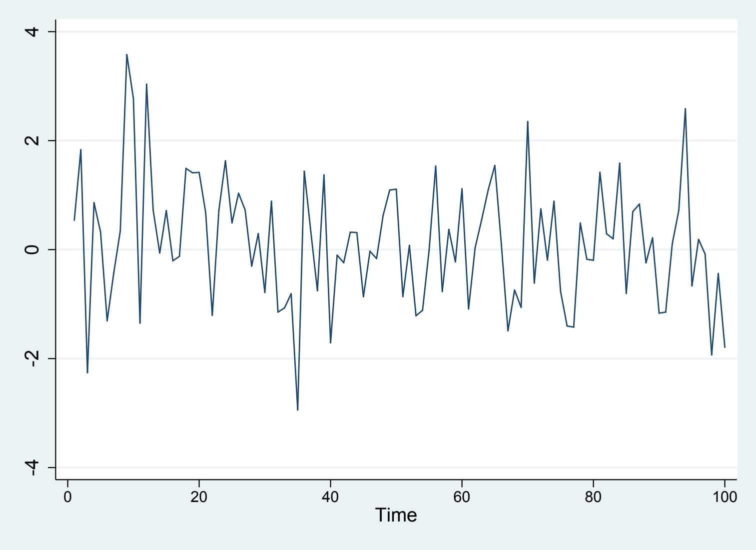
Further Time Series: Simulated Series: AR(1)



Further Time Series: Chaos or a real time series?



Further Time Series: Chaos?



Characteristics of Time series

- Trends
- Periodicity (cyclicality)
- Seasonality
- Volatility Clustering
- Nonlinearities
- Chaos

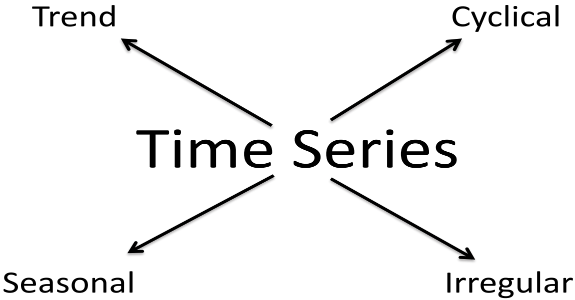
Necessity of (economic) Forecasts

- For political actions and budget control governments need forecasts for macroeconomic variables
  GDP, interest rates, unemployment rate, tax revenues etc.
- marketing need forecasts for sales related variables
  - future sales
  - product demand (price dependent)
  - changes in preferences of consumers

Necessity of (economic) Forecasts

- retail sales company need forecasts to optimize warehousing and employment of staff
- firms need to forecasts cash-flows in order to account of illiquidity phases or insolvency
- university administrations needs forecasts of the number of students for calculation of student fees, staff planning, space organization
- migration flows
- house prices

Time series decomposition



Time series decomposition

Classical additive decomposition:

$$y_t = d_t + c_t + s_t + \epsilon_t \tag{1}$$

- d_t trend component (deterministic, almost constant over time)
- c_t cyclical component (deterministic, periodic, medium term horizons)
- s_t seasonal component (deterministic, periodic; more than one possible)
- ϵ_t irregular component (stochastic, stationary)

Time series decomposition

Goal:

- Extraction of components d_t, c_t and s_t
- The irregular component

$$\epsilon_t = y_t - d_t - c_t - s_t$$

  should be stationary and ideally white noise.
- Main task is then to model the components appropriately.
- Data transformation maybe necessary to account for heteroscedasticity (e.g. log-transformation to stabilize seasonal fluctuations)

Time series decomposition

The multiplicative model:

$$y_t = d_t \cdot c_t \cdot s_t \cdot \epsilon_t \quad (2)$$

will be treated in the tutorial.

Simple Filters

- $$\textit{series} = \textit{signal} + \textit{noise} \tag{3}$$
- The statistician would says
  $$\textit{series} = \textit{fit} + \textit{residual} \tag{4}$$
- At a later stage:
  $$\textit{series} = \textit{model} + \textit{errors} \tag{5}$$
  $\Rightarrow$ mathematical function plus a probability distribution of the error term

Linear Filters

A linear filter converts one times series (x_T) into another (y_t) by the linear operation

$$y_t = \sum_{r=-q}^{+s} a_r x_{t+r}$$

where a_r is a set of weights. In order to smooth local fluctuation one should chose the weight such that

$$\sum a_r = 1$$

The idea

$$y_t = f(t) + \epsilon_t \tag{6}$$

We assume that $f(t)$ and ϵ_t are well-behaved.
Consider N observations at time t_j which are reasonably close in time to t_i. One possible smoother is

$$y^*_{t_i} = 1/N \sum y_{t_j} = 1/N \sum f(t_j) + 1/N \sum \epsilon_{t_j} \approx f(t_i) + 1/N \sum \epsilon_{t_j} \tag{7}$$

if $\epsilon_t \sim N(0, \sigma^2)$, the variance of the sum of the residuals is σ^2/N^2.
The smoother is characterized by

- span, the number of adjacent points included in the calculation
- type of estimator (median, mean, weighted mean etc.)

Moving Average

- Used for time series smoothing.
- Consists of a series of arithmetic means.
- Result depends on the window size L (number of included periods to calculate the mean).
- In order to smooth the cyclical component, L should exceed the cycle length
- L should be uneven (avoids another cyclical component)

Moving Average

$$
\begin{aligned}
MA(y_t) &= \frac{1}{2q+1} \sum_{r=-q}^{+q} y_{t+r} \\
L &= 2q+1
\end{aligned}
$$

where the weights are given by

$$
a_r = \frac{1}{2q+1}
$$

Moving Average

Two-Sided MA:

$$MA(y_t) = \frac{1}{2q+1} \sum_{r=-q}^{+q} y_{t+r}$$

One-sided MA:

$$MA(y_t) = \frac{1}{q+1} \sum_{r=0}^{q} y_{t-r}$$

Moving Average

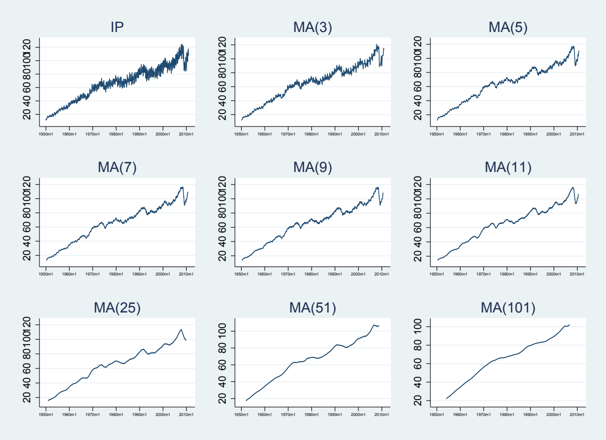
Example: Moving Average (MA) over 3 Periods

- First MA term: $MA_2(3) = \frac{y_1+y_2+y_3}{3}$
- Second MA term: $MA_3(3) = \frac{y_2+y_3+y_4}{3}$

Moving Average

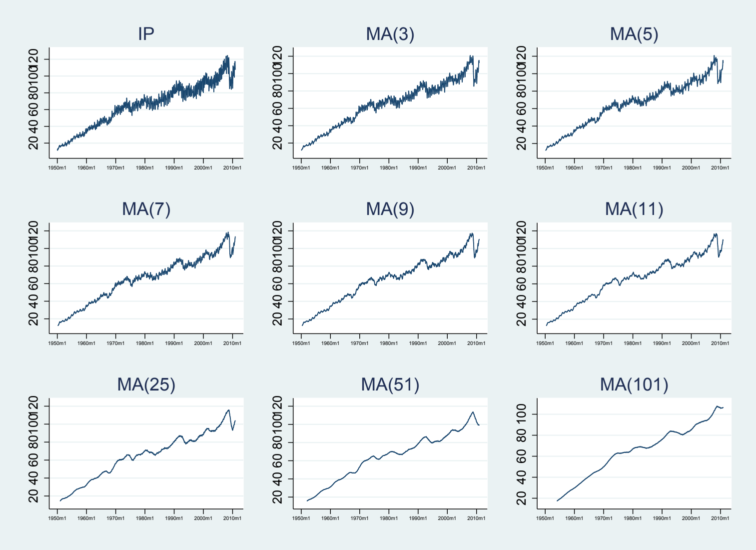
Year	Projects	MA(3) L=3
2005	2	
2006	5	3
2007	2	3
2008	2	3.67
2009	7	5
2010	6	

Moving Average Example - TWO-sided

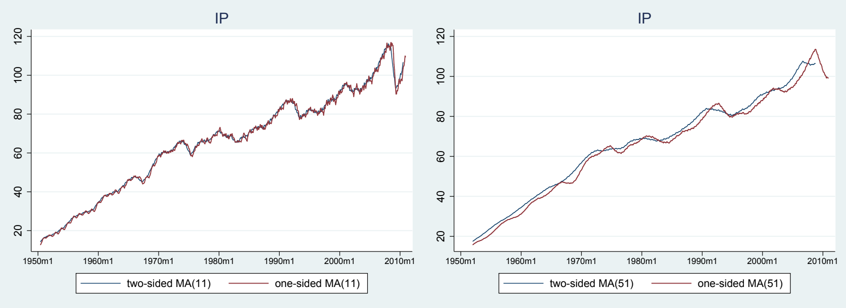


⇒ the larger L the smoother and shorter the filtered series

Moving Average Example - One-sided



Moving Average Example - Comparison of One- and two-sided



EXAMPLE

Generate a random time series (normally distributed) with $T = 20$

- Quick and dirty: Moving Average with Excel
- Nice and Slow: Write a simple Matlab program for calculating a moving average of order L
- Additional Task: Increase the number of observations to $T = 100$, include a linear time trend and calculate different MAs
- Variation: Include some outliers and see how the calculations change.

Exponential Smoothing

- weighted moving averages
- latest observation has the highest weight compared to the previous periods

$$\hat{y}_t = wy_t + (1-w)\hat{y}_{t-1}$$

Repeated substitution gives:

$$\hat{y}_t = w\sum_{s=0}^{t-1}(1-w)^s \hat{y}_{t-s}$$

$\Rightarrow$ that's why it is called exponential smoothing, forecasts are the weighted average of past observations where the weights decline exponentially with time.

Exponential Smoothing

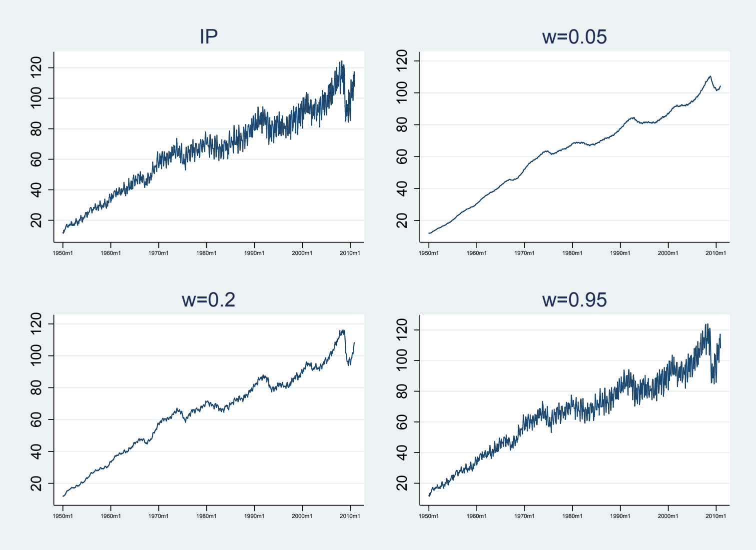
- Is used for smoothing and short–term forecasting
- Choice of w:
    - subjective or through calibration
    - numbers between 0 and 1
      Close to 0 for smoothing out unpleasant cyclical or irregular components
      Close to 1 for forecasting

Exponential Smoothing

$$\hat{y}_t = w y_t + (1 - w)\hat{y}_{t-1} \quad w = 0.2$$

Year	Projects	Smothed Value	Forecast
2005	2	2	-
2006	5	0.2*5+0.8*2 =2.6	2.000
2007	2	0.2*2+0.8*2.6 =2.48	2.600
2008	2	0.2*2+0.8*2.48 =2.3684	2.480
2009	7	0.2*7+0.8*2.384=3.307	2.384
2010	6	0.2*6+0.8*3.307 =3.846	3.307

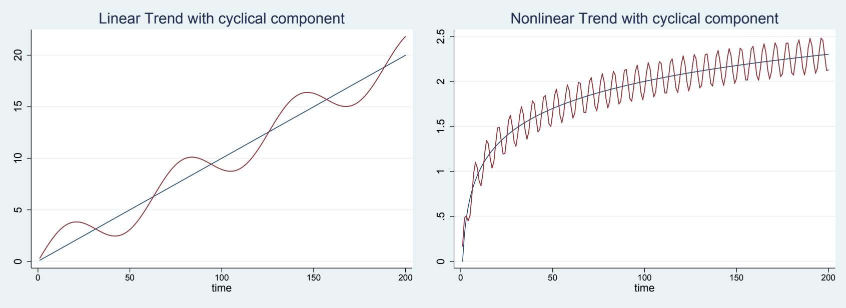
Exponential Smoothing



Trend Component

- positive or negative trend
- observed over a longer time horizon
- linear vs. non–linear trend
- smooth vs. non–smooth trends
- $\Rightarrow$ trend is ’unobserved’ in reality

Trend Component: Example



Why is trend extraction so important?

The case of detrending GDP

- trend GDP is denoted as **potential output**
- The difference between trend and actual GDP is called the **output gap**
- Is an economy below or above the current trend? (Or is the output gap positive or negative?)
  $\Rightarrow$ consequences for economic policy (wages, prices etc.)
- Trend extraction can be highly controversial!

Linear Trend Model

Year	Time (x_t)	Turnover (y_t)
05	1	2
06	2	5
07	3	2
08	4	2
09	5	7
10	6	6

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

Linear Trend Model

Estimation with OLS

$$\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t = 1.4 + 0.743 x_t$$

Forecast for 2011:

$$\hat{y}_{2011} = 1.4 + 0.743 \cdot 7 = 6.6$$

Quadratic Trend Model

Year	Time (x_t)	Time^2 (x_t^2)	Turnover (y_t)
05	1	1	2
06	2	4	5
07	3	9	2
08	4	16	2
09	5	25	7
10	6	36	6

$$y_t = \alpha + \beta_1 x_t + \beta_2 x_t^2$$

Quadratic Trend Model

$$\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t + \hat{\beta}_2 x_t^2 = 3.4 - 0.757143 x_t + 0.214286 x_t^2$$

Forecast for 2011:

$$\hat{y}_{2011} = 3.4 - 0.757143 \cdot 7 + 0.214286 \cdot 7^2 = 8.6$$

Exponential Trend Model

Year	Time (x_t)	Turnover (y_t)
05	1	2
06	2	5
07	3	2
08	4	2
09	5	7
10	6	6

$$y_t = \alpha \beta_1^{x_t}$$

$\Rightarrow$ Non-linear Least Squares (NLS) or
Linearize the model and use OLS:

$$\log y_t = \log \alpha + \log(\beta_1) x_t$$

$\Rightarrow$ 'relog' the model

Exponential Trend Model

Estimation via **NLS**:

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1^{\,x_t} = 0.08 \cdot 1.93^{x_t}$$

Forecast for 2011:

$$\hat{y}_{2011} = 0.08 \cdot 1.93^7 = 15.4$$

Logarithmic Trend Model

Year	Time (x_t)	log(*Time*)	Turnover (y_t)
05	1	log(1)	2
06	2	log(2)	5
07	3	log(3)	2
08	4	log(4)	2
09	5	log(5)	7
10	6	log(6)	6

Logarithmic Trend:

$$y_t = \alpha + \beta_1 \log x_t$$

Logarithmic Trend Model

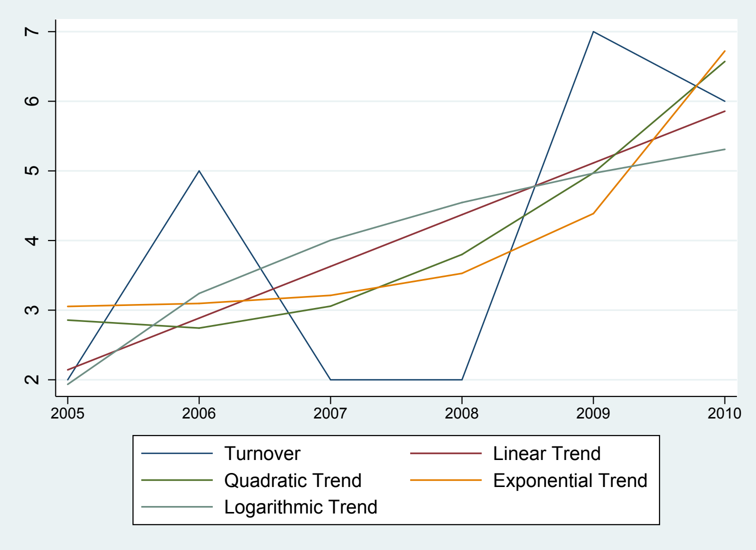
Estimation via **OLS**:

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 \log x_t = 1.934675 + 1.883489 \cdot \log y_t$$

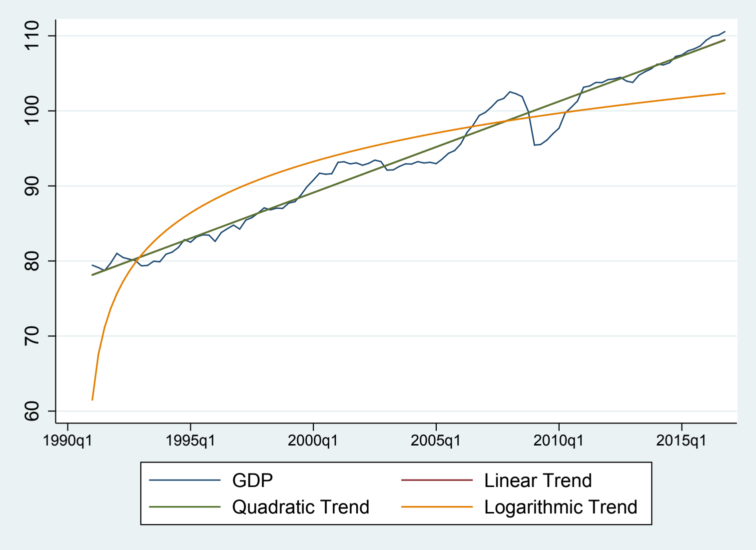
Forecast for 2011:

$$\hat{Y}_{2011} = 1.934675 + 1.883489 \cdot \log(7) = 5.6$$

Comparison of different trend models



Detrending GDP



Which trend model to choose?

- Linear Trend model, if the first differences

$$y_t - y_{t-1}$$

  are stationary
- Quadratic trend model, if the second differences

$$(y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

  are stationary
- Logarithmic trend model, if the relative differences

$$\frac{y_t - y_{t-1}}{y_t}$$

  are stationary

The Hodrick-Prescott-Filter (HP)

The HP extracts a flexible trend. The filter is given by

$$\min_{\mu_t} \sum_{t=1}^{T} [(y_t - \mu_t)^2 + \lambda \sum_{t=2}^{T-1} \{(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})\}^2] \tag{8}$$

where μ_t is the flexible trend and λ a smoothness parameter chosen by the researcher.

- As λ approaches infinity we obtain a linear trend.
- Currently the most popular filter in economics.

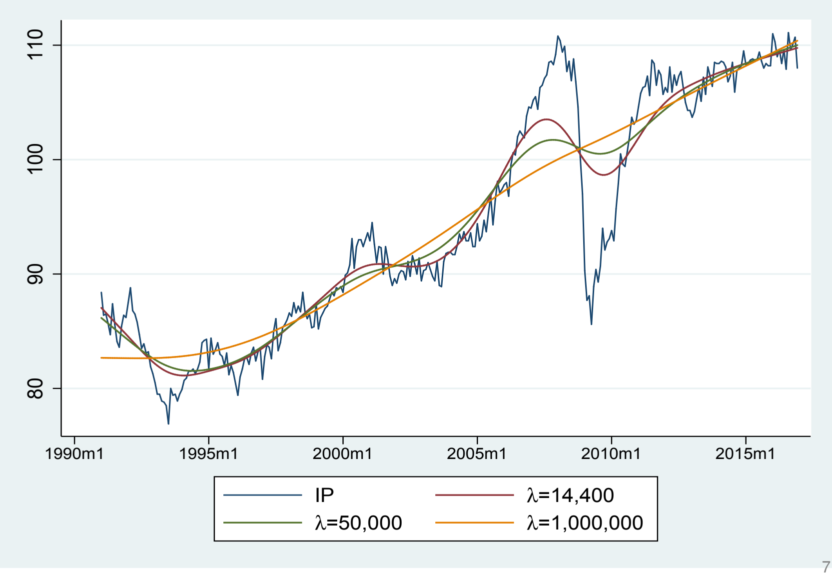
The Hodrick-Prescott-Filter (HP)

How to choose λ?
Hodrick-Prescot (1997) recommend:

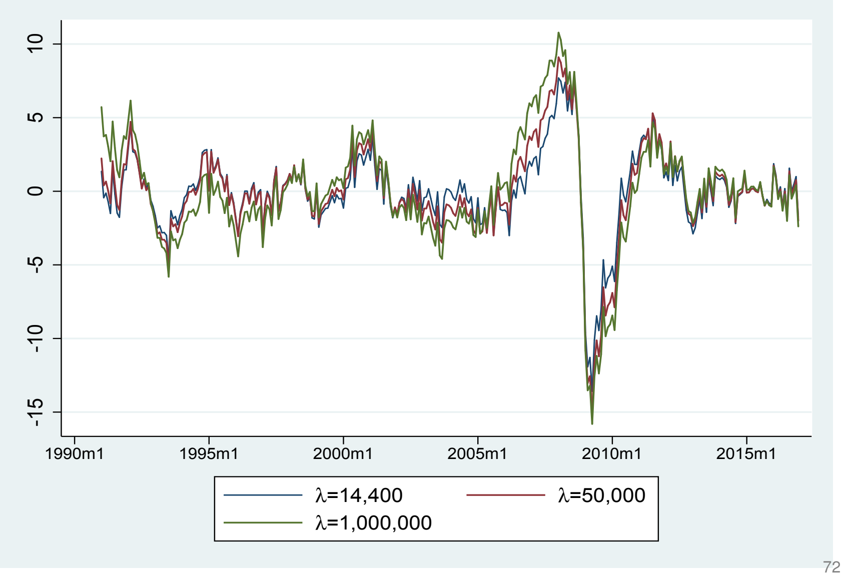
$$\lambda = \begin{cases} 100 \text{ for annual data} \\ 1600 \text{ for quarterly data} \\ 14400 \text{ for monthly data} \end{cases} \tag{9}$$

Alternative: Ravn and Uhlig (2002)

The Hodrick-Prescott-Filter (HP)



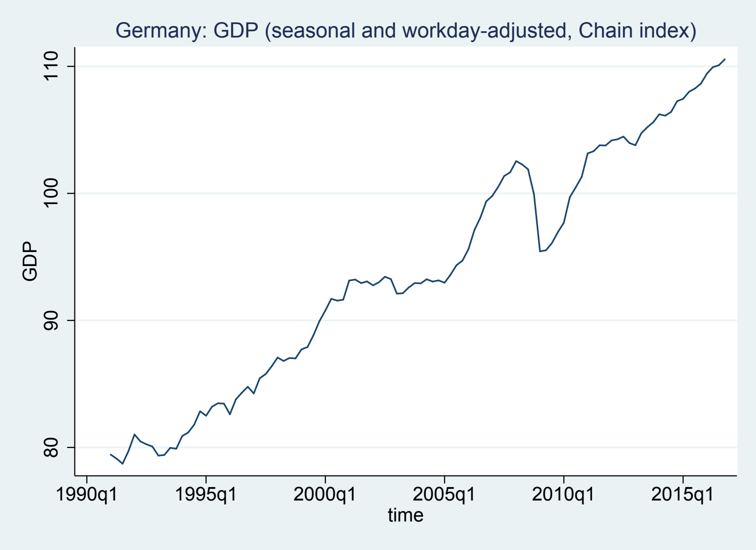
The Hodrick-Prescott-Filter (HP)



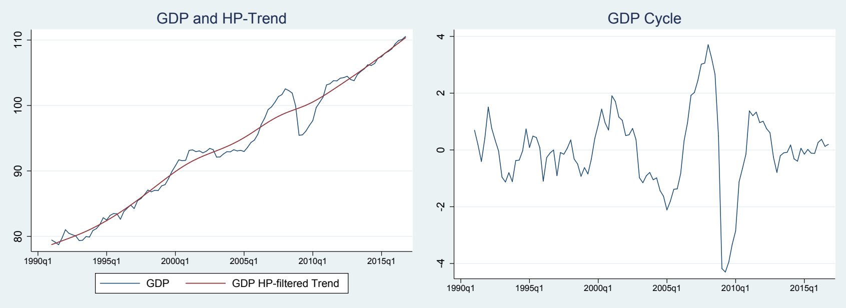
Problems with the HP-Filter

- λ is a 'tuning' parameter
- end of sample instability
  $\Rightarrow$ AR-forecasts

Case study for German GDP: Where are we now?



HP-Filter


GDP and HP-Trend
110
100
90
80
1990q1
1995q1
2000q1
2005q1
2010q1
2015q1
GDP
GDP HP-filtered Trend
GDP Cycle
4
2
0
-2
-4
1990q1
1995q1
2000q1
2005q1
2010q1
2015q1

Can we test for a trend?

- Yes and no
- Is the trend component significant?
- several trends can be significant
- Trend might be spurious
- Is it plausible to have a trend in the data?
- A priori information by the researcher
- unit roots

EXAMPLE

Time series: Industrial Production in Germany (1991:01-2016:12)

- Plot the time series and state which trend adjustment might be appropriate
- Prepare your data set in Excel and estimate various trends in Eviews
- Which trend would you choose?

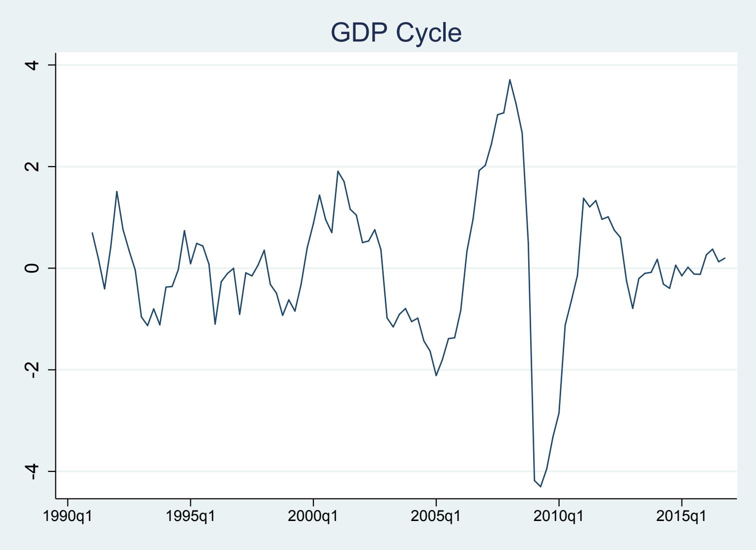
Cyclical Component

- is not always present in time series
- Is the difference between the observed time series and the estimated trend

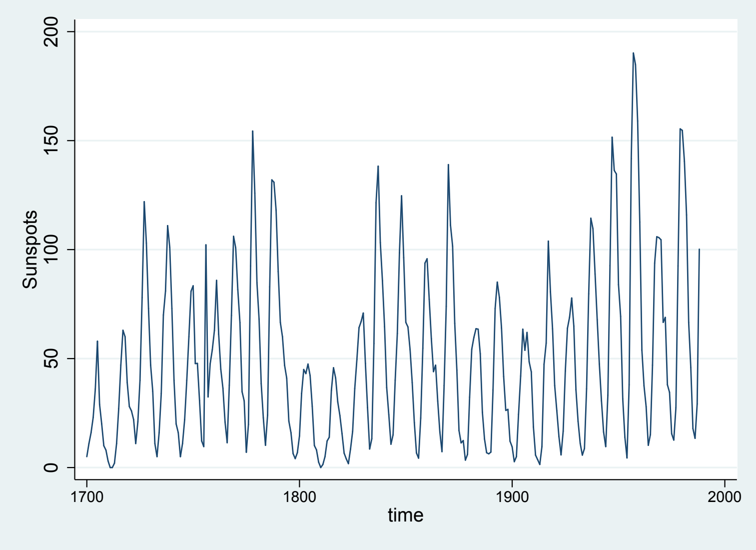
In economics

- characterizes the Business cycle
- different length of cycles (3-5 or 10-15 years)

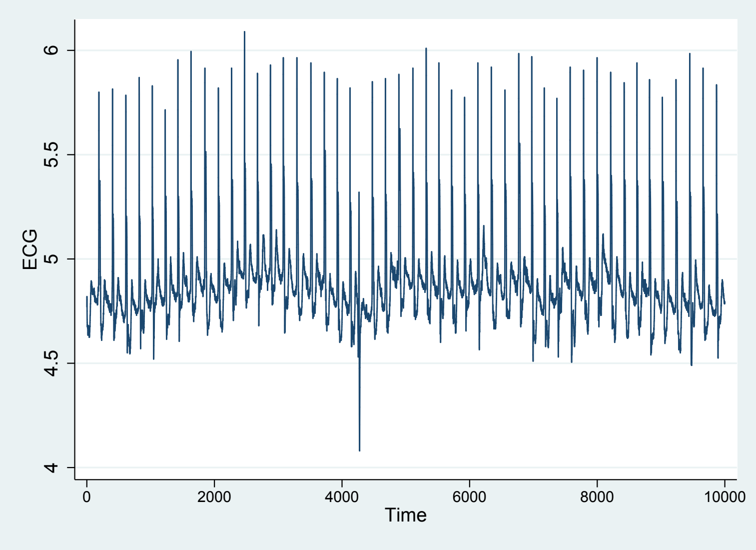
Cyclical Component: Example



Cyclical Component: Example II



Cyclical Component: Example III

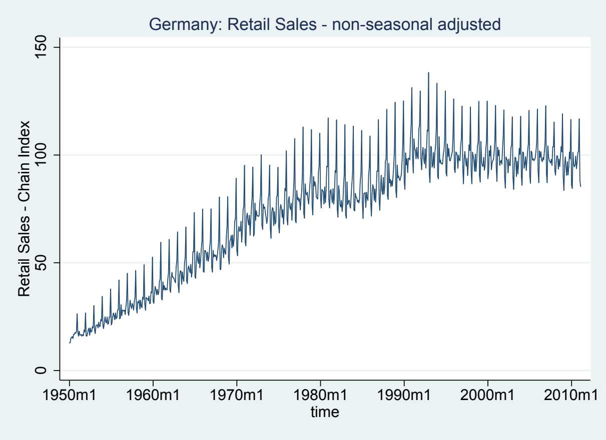


Can we test for a cyclical component?

- Yes and no
- see the trend section
- Does a cycle make sense?

Seasonal Component

- similar upswings and downswings in a fixed time interval
- regular pattern, i.e. over a year



Types of Seasonality

- A: $y_t = m_t + S_t + \epsilon_t$
- B: $y_t = m_t S_t + \epsilon_t$
- C: $y_t = m_t S_t \epsilon_t$

Model A is additive seasonal, Models B and C contains multiplicative seasonal variation

Types of Seasonality

- if the seasonal effect is constant over the seasonal periods
  $\Rightarrow$ additive seasonality (Model A)
- if the seasonal effect is proportional to the mean
  $\Rightarrow$ multiplicative seasonality (Model A and B)
- in case of multiplicative seasonal models use the logarithmic transformation to make the effect additive

Seasonal Adjustment

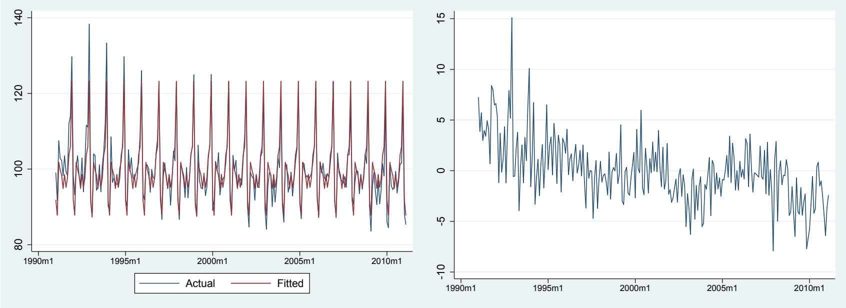
Simplest Approach to seasonal adjustment:

- Run the time series on a set of dummies without a constant (Assumes that the seasonal pattern is constant over time)
- the residuals of this regression are seasonal adjusted
- Example: Monthly data

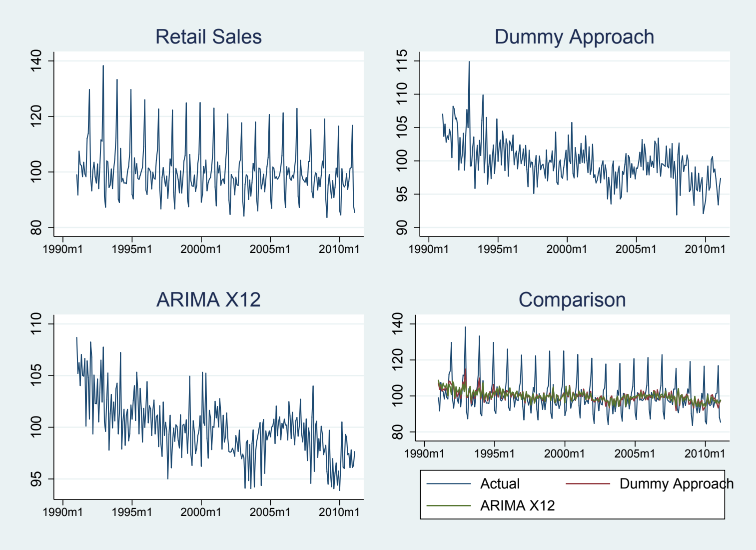
$$
\begin{aligned}
y_t &= \sum_{i=1}^{12} \beta_i D_i + \epsilon_t \\
\epsilon_t &= y_t - \sum_{i=1}^{12} \hat{\beta} D_i \\
y_{t,sa} &= \epsilon_t + mean(y_t)
\end{aligned}
$$

- The most well known seasonal adjustment procedure: CENSUS X12 ARIMA

Seasonal Adjustment: Dummy Regression Example



Seasonal Adjustment: Example



Seasonal Moving Averages

For monthly data one can employ the filter

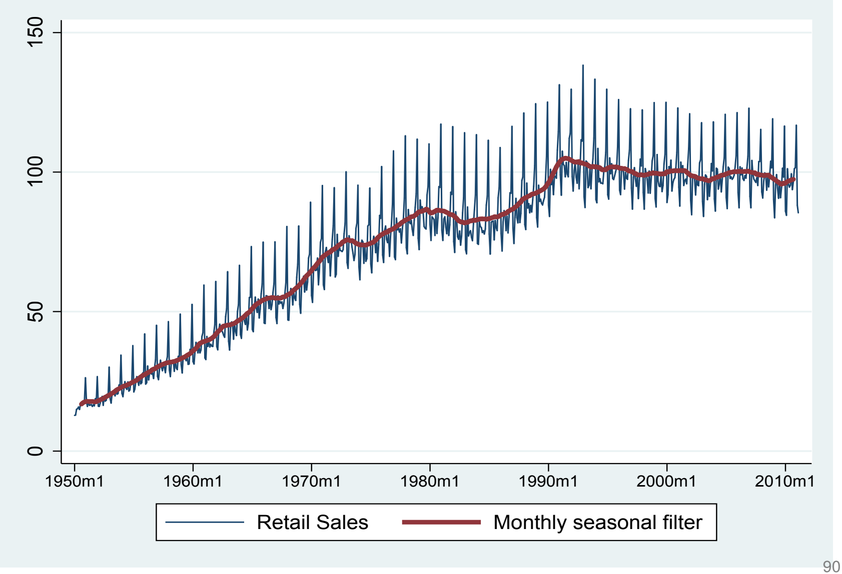
$$SMA(y_t) = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + y_{t-4} + \ldots + y_{t+6} + \frac{1}{2}y_{t+6}}{12}$$

or for quarterly data

$$SMA(y_t) = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}$$

- Note: The weights add up to one!
- Standard moving average not applicable

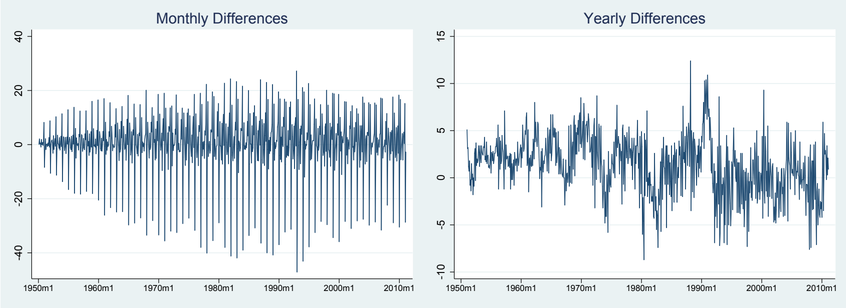
Seasonal Moving Averages: Retail Sales Example



Seasonal Differencing

- seasonal effect can be eliminated using the a simple linear filter
- in case of a monthly time series: $\Delta_{12} y_t = y_t - y_{t-12}$
- in case of a quarterly time series: $\Delta_4 y_t = y_t - y_{t-4}$

Seasonal Differencing: Retail Sales Example



Can we test for seasonality?

- Yes and no
- Does seasonality makes sense?
- Compare the seasonal adjusted and unadjusted series
- look into the ARIMA X12 output
- Be aware of changing seasonal patterns

EXAMPLE

Time series: seasonally unadjusted Industrial Production in Germany (1991:01-2011:02)

- Remove the seasonality by a moving seasonal filter
- Try the dummy approach
- Finally, use the ARIMAX12-Approach
- Start the sample in 1991:01 and compare all filters with the full sample

Irregular Component

- erratic, non-systematic, random "residual" fluctuations due to random shocks
  - in nature
  - due to human behavior
- no observable iterations

Can we test for an irregular component?

- YES
- several tests available whether the irregular component is a white noise or not

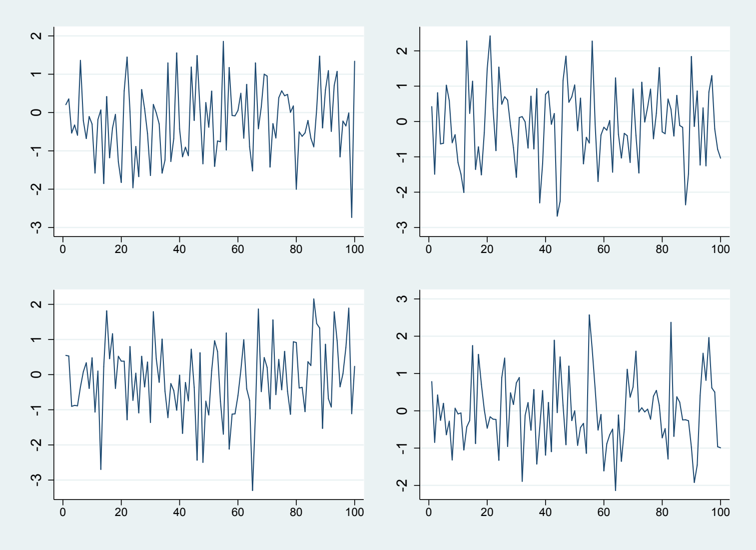
White Noise

A process $\{y_t\}$ is called **white noise** if

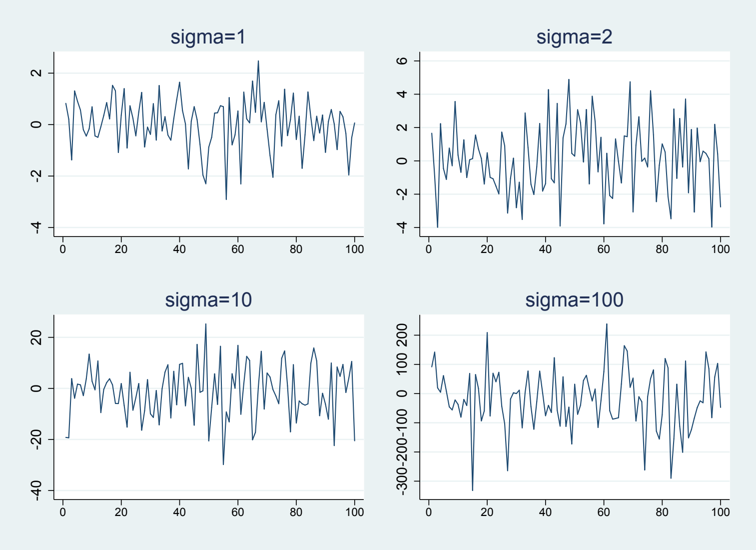
$$\begin{aligned} \mathrm{E}(y_t) &= 0 \\ \gamma(0) &= \sigma^2 \\ \gamma(h) &= 0 \text{ for } |h| > 0 \end{aligned}$$

$\Rightarrow$ all y_t's are uncorrelated. We write: $\{y_t\} \sim WN(0, \sigma^2)$

White Noise


2
1
0
-1
-2
-3
0
20
40
60
80
100
2
1
0
-1
-2
-3
0
20
40
60
80
100
2
1
0
-1
-2
-3
0
20
40
60
80
100
3
2
1
0
-1
-2
0
20
40
60
80
100

White Noise


sigma=1
2
0
-2
-4
0
20
40
60
80
100
sigma=2
6
4
2
0
-2
-4
sigma=10
20
0
-20
-40
sigma=100
200
100
0
-100
-200
-300

Random Walk (with drift)

A simple *random walk* is given by

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

By adding a constant term

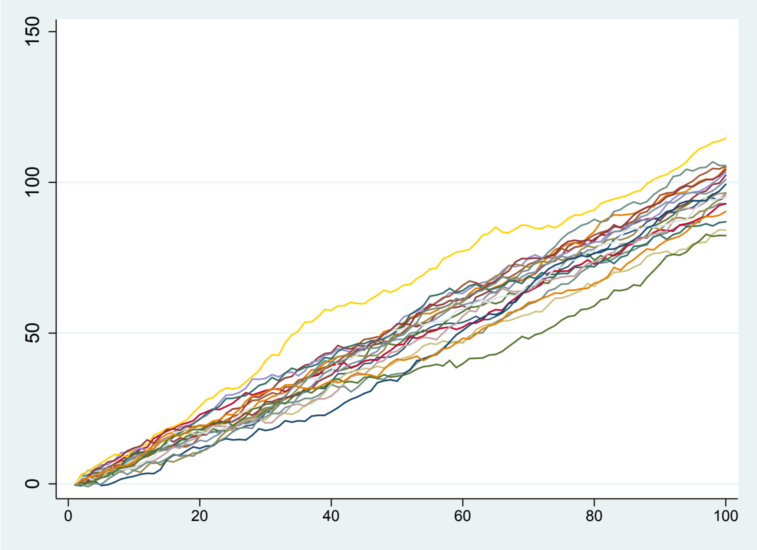
$$y_t = c + y_{t-1} + \epsilon_t$$

we get a *random walk with drift*. It follows that

$$y_t = ct + \sum_{j=1}^{t} \epsilon_j$$

Random Walk: Examples

Random Walk with Drift: Examples



EXAMPLE

Fun with Random Walks

- Generate 50 different random walks
- Plot all random walks
- Try different variances and distributions

Autoregressive processes

- especially suitable for (short-run) forecasts
- utilizes autocorrelations of lower order
  - 1st order: correlations of successive observations
  - 2nd order: correlations of observations with two periods in between
- Autoregressive model of order p

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \epsilon_t$$

Autoregressive processes

Number of machines produced by a firm

Year	Units
2003	4
2004	3
2005	2
2006	3
2007	2
2008	2
2009	4
2010	6

$\Rightarrow$ Estimation of an AR model of order 2

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t$$

Autoregressive processes

Estimation Table:

Year	Constant	y_t	y_{t-1}	y_{t-2}
2003	1	4		
2004	1	3	4	
2005	1	2	3	4
2006	1	3	2	3
2007	1	2	3	2
2008	1	2	2	3
2009	1	4	2	2
2010	1	6	4	2

$\Rightarrow$ OLS

$$\hat{y}_t = 3.5 + 0.8125 y_{t-1} - 0.9375 y_{t-2}$$

Autoregressive processes

Forecasting with an AR(2) model:

$$
\begin{aligned}
\hat{y}_t &= 3.5 + 0.8125 y_{t-1} - 0.9375 y_{t-2} \\
y_{2011} &= 3.5 + 0.8125 y_{2010} - 0.9375 y_{2009} \\
&= 3.5 + 0.8125 \cdot 6 - 0.9375 \cdot 4 \\
&= 4.625
\end{aligned}
$$

Table of content I

Table of content II

Stochastic Processes

A stochastic process can be described as ’a statistical phenomenon that evolvoes in time according to probabilistic terms’.

Stochastic Processes

- Let y_t be an index ($t \in Z$) random variable.
- The sequence $\{y_t\}_{t \in Z}$ is called a stochastic process.
- Stochastic processes can be studied both in the time and frequency domain.
  $\Rightarrow$ We focus on the time domain.
- For stochastic processes the expectation value, variance and covariance are the theoretical counterparts to the time series mean, variance and covariance.
- A time series is a realization of a stochastic process.
- In order to characterize stochastic processes we have to focus on stationary processes.
- An important class of stationary processes are linear ARIMA (autoregressive integrated moving average) processes.

Stochastic Processes

- most statistical problems are concerned with estimating the properties of a population from a sample
- the latter one is typically determined by the investigator, including sample size and whether randomness is incorporated into the selection process
- time series analysis is different, as it usually impossible to make more than one observation at any given time
- it is possible to increase the sample size by varying the *length* of the observed time series
- but there will be only a single outcome of the process and a single observation on the random variable at time t

Basic Approach to time series modeling

- time series are sampled either with regular (equidistant) or irregular intervals (non-equidistant)
- regular time intervals: yearly, quarterly, monthly, weekly, daily, hourly, etc. ($\Rightarrow$ continuous flow)
- irregular intervals: transaction prices of stocks

Basic Approach to time series modeling

- A time series $\{y_t, t = \ldots -1, 0, 1, \ldots\}$ can be interpreted as a realisation of a stochastic process
- For time series with finite first and second moments we define
  - mean function: $\mu(t) = \mathrm{E}(y_t)$
  - covariance function:

$$\begin{aligned}\gamma(t, t+h) &= \mathrm{Cov}(y_t, y_{t+h}) \\ &= \mathrm{E}[(y_t - \mu(t))(y_{t+h} - \mu(t))]\end{aligned}$$

  - the Autocorrelation function: $\rho(h) = \gamma(h)/\gamma(0) = \gamma(h)/\sigma^2$

Basic Approach to time series modeling

The concept of *stationarity* plays a central role in time series analysis.

- A time series $\{y_t\}$ is **weakly stationary**, if for all t:
  - $\mu(t) = \mu$, i.e., it does not depend on t, and
  - $\gamma(t + h, t) = \gamma(h)$, depends only on h and not on t
- This means, that for all h die time series $\{y_t\}$ moves in a similar way as the "shifted" time series $\{y_{t+h}\}$.

Basic Approach to time series modeling

- Assuming that y_t is weakly stationary, we define the **Autocovariance function** (ACVF) for lag h

  $$\gamma(h) = \gamma(t, t-h)$$

  and the autocorrelation function(ACF)

  $$\rho(h) = \gamma(h)/\gamma(0) = Corr(y_t, y_{t-h})$$

  The ACF is a sequence of correlation with the following characteristics

  $$-1 \leq \rho(h) \leq 1 \text{ mit } \rho(0) = 1.$$

Basic Approach to time series modeling

- The ACVF has the following properties:
  - $\gamma(0) \geq 0$,
  - $|\gamma(h)| \leq \gamma(0)$, for all h
  - $\gamma(h) = \gamma(-h)$, for all h

Basic Approach to time series modeling

- **Step 1**: Data inspection, data cleaning (exclusion of outliers), data transformation (e.g. seasonal or trend adjustment),
- **Step 2**: Choice of a specific model that accounts best for the (adjusted) data at hand
- **Step 3**: Specification and estimation of parameters of the model
- **Step 4**: Check the estimated model, if necessary go back to step 3, 2, or 1
- **Step 5**: Use the model in practice
  - compact description of the data
  - interpretation of the data characteristics
  - inference, testing of hypotheses (in-sample)
  - forecasting (out-of-sample)

Be careful!

- Basic Assumption: Characteristics of a time series remain constant also in the future.
- Forecasting with "mechanical" trend projections without considering experience and subjective elements ("judgemental forecasts")

Table of content I

Table of content II

Linear Difference Equations

Time series models can be represented or approximated by a linear difference equation. Consider the situation where a realization at time t, y_t, is a linear function of the last p realizations of y and a random disturbance term, denoted by ϵ_t.

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \epsilon_t. \tag{10}$$

$\Rightarrow$ $AR(p)$-Process

The Lag Operator

The *lag operator* (also called backward shift operator), denoted by L, is an operator that shifts the time index backward by one unit. Applying it to a variable at time t, we obtain the value of the variable at time $t-1$, i.e.,

$$Ly_t = y_{t-1}.$$

Applying the lag operator twice amount to lagging the variable twice, i.e., $L^2 y_t = L(Ly_t) = Ly_{t-1} = y_{t-2}$.

The Lag Operator

More formally, the lag operator transforms one time series, say $\{x_t\}_{t=-\infty}^{\infty}$, into another series, say $\{y_t\}_{t=-\infty}^{\infty}$, where $x_t = y_{t-1}$. Raising L to a negative power, we obtain a *delay* (or *lead*) *operator*, i.e.,

$$L^{-k} y_t = y_{t+k}.$$

The Lag Operator

The following statements hold for the lag operator L

$$
\begin{aligned}
Lc &= c \text{ for a constant c} && (11) \\
(L^j + L^i)y_t &= L^j y_t + L^i y_t \text{ (distributive law)} && (12) \\
L^i(L^j y_t) &= L^i y_{t-j} \text{ (associative law)} && (13) \\
aLy_t &= L(ay_t) = ay_{t-1} && (14)
\end{aligned}
$$

The Difference Operator

The *difference operator* Δ is used to express the difference between values of time series at different times. With Δy_t we denote the first difference of y_t, i.e.,

$$\Delta y_t = y_t - y_{t-1}.$$

It follows that

$$\begin{aligned} \Delta^2 y_t &= \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \end{aligned}$$

etc. The difference operator can expressed in terms of the lag operator by $\Delta = 1 - L$. Hence, $\Delta^2 = (1 - L)^2 = 1 - 2L + L^2$ and, in general, $\Delta^n = (1 - L)^n$.

Transforming the Expression of Time Series Models

The lag operator enables us to express higher–order difference equations more compactly in form of polynomials in lag operator L.

For example, the difference equation

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + c$$

can be written as

$$y_t = \alpha_1 L y_t + \alpha_2 L^2 y_t + \alpha_3 L^3 y_t + c,$$

$$(1 - \alpha_1 L - \alpha_2 L^2 - \alpha_3 L^3) y_t = c$$

or, in short,

$$a(L) y_t = c.$$

The Characteristic Equation

Replacing in polynomial $a(L)$ lag operator L by variable λ, we obtain the *characteristic equation* associated with difference equation (10):

$$a(\lambda) = 0. \tag{15}$$

A value of λ which satisfies characteristic equation (15) is called a *root* of polynomial $a(\lambda)$.
$\Rightarrow$ Will be important in later applications.

Solving Difference Equations

Expression (15) represents the so-called *coefficient form* of a characteristic equation, i.e.,

$$1 - \alpha_1 \lambda - \cdots - \alpha_p \lambda^p = 0.$$

An alternative is the *root form* given by

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_p - \lambda) = \prod_{i=1}^{p} (\lambda_i - \lambda) = 0.$$

Solving Difference Equations: An Example

Consider the difference equation

$$y_t = \frac{3}{2}y_{t-1} - \frac{1}{2}y_{t-2} + \epsilon_t.$$

The characteristic equation in coefficient form is given by

$$1 - \frac{3}{2}\lambda + \frac{1}{2}\lambda^2 = 0$$

or

$$2 - 3\lambda + 1\lambda^2 = 0,$$

which can be written in root form as

$$(1 - \lambda)(2 - \lambda) = 0.$$

Here, $\lambda_1 = 1$ and $\lambda_2 = 2$ represent the set of possible solutions for λ satisfying the characteristic equation $1 - \frac{3}{2}\lambda + \frac{1}{2}\lambda^2 = 0$.

Solving Difference Equations: An Example

Calculate the characteristic roots of the following difference equations

$$y_t = y_{t-1} - y_{t-2} + \epsilon_t \quad (16)$$

$$y_t = -y_{t-1} + y_{t-2} + \epsilon_t \quad (17)$$

$$y_t = 0.125 y_{t-3} + \epsilon_t \quad (18)$$

Autoregressive (AR) Processes

An *autoregressive process* of order p, or briefly an AR(p) process, is a process where realization y_t is a weighted sum of past p realizations, i.e., $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$, plus an additive, contemporaneous disturbance term, denoted by ϵ_t.
The process can be represented by the p-th order difference equation

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \epsilon_t. \tag{19}$$

Autoregressive (AR) Processes

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \epsilon_t. \tag{20}$$

We assume that ϵ_t, $t = 0, \pm 1, \pm 2 \ldots$, is a zero-mean, independently and identically distributed (iid) sequence with

$$\mathrm{E}(\epsilon_t) = 0, \qquad \mathrm{E}(\epsilon_s \epsilon_t) = \begin{cases} \sigma^2, & \text{if } s = t, \\ 0, & \text{if } s \neq t, \end{cases} \tag{21}$$

for all t and s. Sequence (21) is called a zero–mean *white–noise process*, or simply *white noise.*

Autoregressive (AR) Processes

Using the lag operator L, the AR(p) process (19) can be expressed more compactly as

$$(1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p) y_t = \epsilon_t$$

or

$$a(L) y_t = \epsilon_t, \tag{22}$$

where the autoregressive polynomial $a(L)$ is defined by $a(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p$.

The mean of a stationary AR(1) process

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t$$

Taking Expectations (E) we get

$$\begin{aligned} E(y_t) &= \alpha_0 + \alpha_1 E(y_{t-1}) + E(\epsilon_t) \\ E(y_t) &= \alpha_0 + \alpha_1 E(y_t) \\ E(y_t) &= \mu = \frac{\alpha_0}{1 - \alpha_1} \end{aligned}$$

The mean of a stationary AR(p) process

We the same technique one can obtain the mean of an AR(2) process

$$E(y_t) = \mu = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2}$$

and an AR(p) process

$$E(y_t) = \mu = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \ldots - \alpha_p}$$

Examples

Calculate the mean of the following AR processes

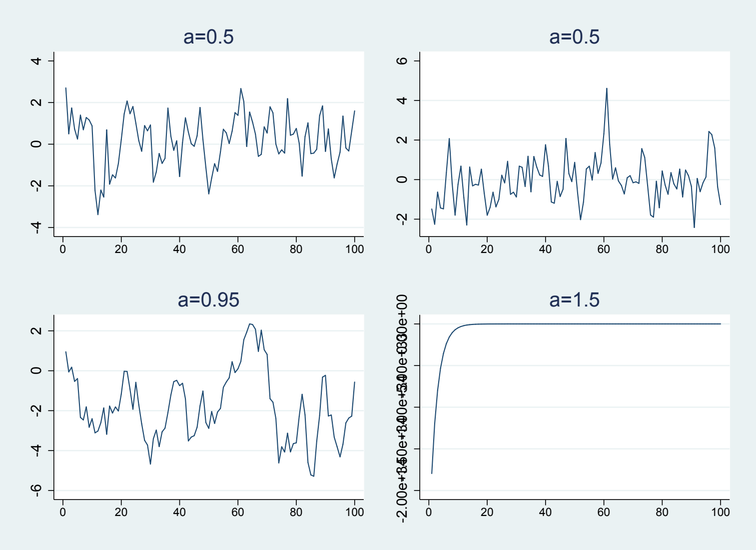
$$y_t = 0.5y_{t-1} + \epsilon_t \quad (23)$$

$$y_t = 0.5 + 0.5y_{t-1} + \epsilon_t \quad (24)$$

$$y_t = 0.5 - 0.5y_{t-1} + \epsilon_t \quad (25)$$

$$y_t = 0.5 + 0.5y_{t-1} + 0.5y_{t-2} + \epsilon_t \quad (26)$$

AR Examples


a=0.5
4
2
0
-2
-4
0
20
40
60
80
100
a=0.5
6
4
2
0
-2
0
20
40
60
80
100
a=0.95
2
0
-2
-4
-6
0
20
40
60
80
100
a=1.5
0
20
40
60
80
100

Moving Average (MA) Processes

A *moving average process* of order q, denoted by MA(q), is the weighted sum of the preceding q lagged disturbances plus a contemporaneous disturbance term, i.e.,

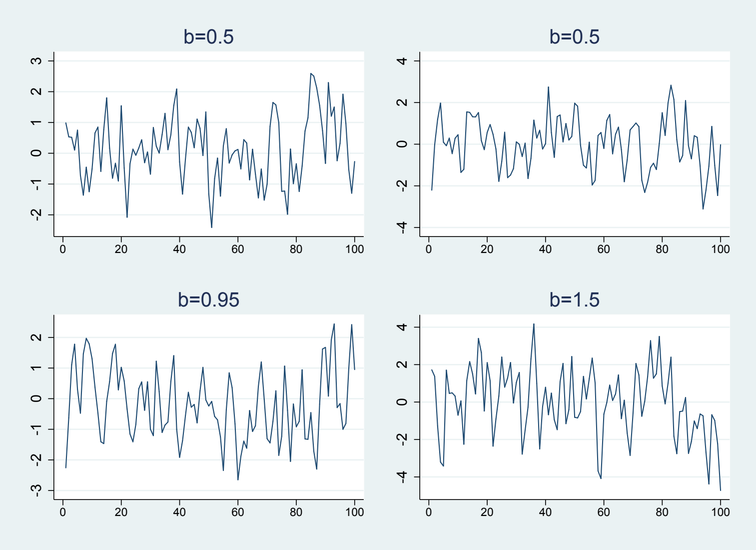
$$y_t = \beta_0 + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q} + \epsilon_t \tag{27}$$

or

$$y_t = b(L)\epsilon_t. \tag{28}$$

Here $b(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \ldots + \beta_q L^q$ denotes a moving average polynomial of degree q, and ϵ_t is again a zero-mean white noise process.

MA Examples



The mean of a stationary MA(q) process

$$y_t = \beta_0 + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q} + \epsilon_t$$

Taking expectations we get

$$\mathrm{E}(y_t) = \mu = \beta_0$$

because

$$\mathrm{E}(\epsilon_t) = \mathrm{E}(\epsilon_{t-1}) = \ldots = \mathrm{E}(\epsilon_{t-q}) = 0$$

Relationship between AR and MA

Consider the AR(1) process

$$y_t = \alpha_1 y_{t-1} + \epsilon_t$$

Repeated substitution yields

$$\begin{aligned} y_t &= \alpha_1(\alpha_1 y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \alpha_1^2 y_{t-2} + \alpha_1 \epsilon_{t-1} + \epsilon_t \\ &= \alpha_1^2(\alpha_1 y_{t-3} + \epsilon_{t-1}) + \alpha_1 \epsilon_{t-1} + \epsilon_t \\ &= \ldots \\ &= \sum_{j=1}^{\infty} \alpha_1^j \epsilon_{t-j} + \epsilon_t \end{aligned}$$

i.e., each stationary AR(1) process can be represented as an MA(∞) process.

The mean of a stationary AR(q) process

Whiteboard

Alternative derivation of the mean of an stationary AR(1) process

$$y_t = c + ay_{t-1} + \epsilon_t \tag{29}$$

with $|a| < 1$.

Relationship between AR and MA

For a general stationary AR(p) process

$$\begin{aligned} y_t &= \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \epsilon_t \\ a(L) y_t &= \epsilon_t \end{aligned}$$

we have

$$y_t = a(L)^{-1} \epsilon_t = \phi(L) \epsilon_t = \sum_{j=1}^{\infty} \phi_j \epsilon_{t-j} \qquad (30)$$

where $\phi(L)$ is an operator satisfying $a(L)\phi(L) = 1$.

Autoregressive Moving Average (ARMA) Processes

The AR and MA processes just discussed can be regarded as special cases of a mixed *autoregressive moving average process*, in short, an ARMA(p, q) process. It is written as

$$y_t = \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q} \qquad (31)$$

or

$$a(L) y_t = b(L) \epsilon_t. \qquad (32)$$

Clearly, ARMA($p, 0$) and ARMA($0, q$) processes correspond to pure AR(p) and MA(q) processes, respectively.

The mean of a stationary ARMA(p, q) process

For

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q} + \epsilon_t \quad (33)$$

we get

$$\mathrm{E}(y_t) = \mu = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \ldots - \alpha_p}$$

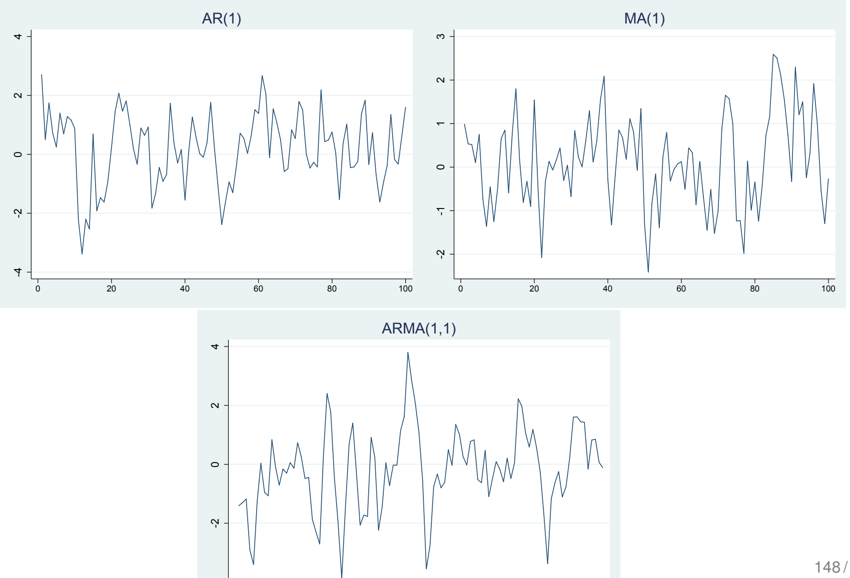
applying the previous arguments.

Examples

Calculate the mean of the following ARMA processes

$$
\begin{aligned}
y_t &= 0.5\epsilon_{t-1} + \epsilon_t && (34) \\
y_t &= 1500\epsilon_{t-1} + 0.5 + 0.75y_{t-1} + \epsilon_t - 0.8\epsilon_{t-2} && (35) \\
y_t &= 0.5 - 0.5y_{t-1} + 2\epsilon_{t-1} + 0.8\epsilon_{t-2} + \epsilon_t && (36) \\
y_t &= y_{t-1} + 0.5\epsilon_{t-1} + \epsilon_t && (37)
\end{aligned}
$$

ARMA Examples



ARMA Processes With Exogenous Variables (ARMAX Processes)

ARMA processes that also include current and/or lagged, exogenously determined variables are called *ARMAX processes*. Denoting the exogenous variable by y_t, an ARMAX process has the form

$$a(L)y_t = b(L)\epsilon_t + g(L)x_t. \tag{38}$$

Example: ARX-models for Forecasting

$$a(L)y_t = g(L)x_t + \epsilon_t \tag{39}$$

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=1}^{q} \gamma_j x_{t-j} + \epsilon_t \tag{40}$$

For example: Forecasting German Industrial Production with its own lagged values plus an exogenous indicator (e.g. the Ifo Business Climate)
$\Rightarrow$ Section about prediction

Integrated ARMA (ARIMA) Processes

Very often we observe that the mean and/or variance of economic time series increase over time. In this case, we say the series are nonstationary. However, a series of the *changes* from one period to the next, i.e., the first differences, may have a mean and/or variance that do not change over time.
$\Rightarrow$ Model the differenced series

Integrated ARMA (ARIMA) Processes

An ARMA model for the d-th difference of a series rather than the original series is called an *autoregressive integrated moving average process*, or an ARIMA (p, d, q), process and written as

$$a(L)\Delta^d y_t = b(L)\epsilon_t. \tag{41}$$

Further Aspects

Seasonal ARMA Processes

$$\alpha_s(L^s)(1 - L^s)^D y_t = \beta_s(L^s)\epsilon_t, \tag{42}$$

ARMA Processes with deterministic Components:
Adding a constant

$$a(L)y_t = c + b(L)\epsilon_t. \tag{43}$$

Or a linear Trend

$$a(L)y_t = c_0 + c_1 t + b(L)\epsilon_t.$$

The Concept of Stationarity

Stationarity is a property that guarantees that the essential properties of a time series remain constant over time. An important concept of stationarity is that of *weak stationarity*. Time series $\{y_t\}_{t=-\infty}^{\infty}$ is said to be weakly stationary if:

(1) the mean of y_t is constant over time, i.e., $E(y_t) = \mu$, $|\mu| < \infty$;

(2) the variance of y_t is constant over time, i.e., $Var(y_t) = \gamma_0 < \infty$;

(3) the covariance of y_t and y_{t-k} does not vary over time, but may depend on the lag k, i.e., $Cov(y_t, y_{t-k}) = \gamma_k$, $|\gamma_k| < \infty$.

$\Rightarrow$ A process is called *strongly (strictly) stationary* if the joint distribution of $(y_1, ...y_k)$ is identical to that of $(y_{1+t}, ...y_{k+t})$.

Stationarity of AR(p) processes

An AR(p) is stationary if the absolute values of all the roots of the characteristic equation

$$\alpha_0 - \alpha_1\lambda - \cdots - \alpha_p\lambda^p = 0.$$

are greater than 1 (with $\alpha_0 = 1$).

- This is in practice difficult to realize.
- What about forth order characteristic equations?
- Alternative: Employ the *Schur Criterion*

Stationarity of AR(p) processes: The Schur Criterion

If the determinants

$$A_1 = \begin{vmatrix} \alpha_0 & \alpha_p \\ \alpha_p & \alpha_0 \end{vmatrix}, A_2 = \begin{vmatrix} \alpha_0 & 0 & \alpha_p & \alpha_{p-1} \\ \alpha_1 & \alpha_0 & 0 & \alpha_p \\ \alpha_p & 0 & \alpha_0 & \alpha_1 \\ \alpha_{p-1} & \alpha_p & 0 & \alpha_0 \end{vmatrix} \dots$$

Stationarity of AR(p) processes: The Schur Criterion

and

$$A_p = \begin{vmatrix} \alpha_0 & 0 & \dots & 0 & \alpha_p & \alpha_{p-1} & \dots & \alpha_1 \\ \alpha_1 & \alpha_0 & \dots & 0 & 0 & \alpha_p & \dots & \alpha_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \alpha_{p-1} & \alpha_{p-2} & \dots & \alpha_0 & 0 & 0 & \dots & \alpha_p \\ \alpha_p & 0 & \dots & 0 & \alpha_0 & \alpha_1 & \dots & \alpha_{p-1} \\ \alpha_{p-1} & \alpha_{p-1} & \dots & 0 & 0 & \alpha_0 & \dots & \alpha_{p-2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ \alpha_1 & \alpha_2 & \dots & \alpha_p & 0 & 0 & \dots & \alpha_0 \end{vmatrix}.$$

are all positive, then an AR(p) process is stationary.

Stationarity of an AR(1) Process

Consider the AR(1) process

$$y_t = \alpha_1 y_{t-1} + \epsilon_t$$

The characteristic equation is

$$1 - \alpha_1 \lambda = 0$$

We have

$$\begin{aligned} A_1 &= \begin{vmatrix} \alpha_0 & \alpha_p \\ \alpha_p & \alpha_0 \end{vmatrix} = \begin{vmatrix} 1 & -\alpha_1 \\ -\alpha_1 & 1 \end{vmatrix} \\ &= 1 - \alpha_1^2 > 0 \iff \left|\alpha_1\right| < 1 \end{aligned}$$

Stationarity of AR(p) processes: **An Alternative Schur Criterion**

For the AR polynomial $a(L) = 1 - \alpha_1 L - \ldots - \alpha_p L^p$, the Schur criterion requires the construction two lower-triangular Toeplitz matrices, A_1 and A_2, whose first columns consist of the vectors $(1, -\alpha_1, -\alpha_2, \ldots, -\alpha_{p-1})'$ and $(-\alpha_p, -\alpha_{p-1}, \ldots, -\alpha_1)'$, respectively, i.e.,

$$A_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -\alpha_1 & 1 & & & 0 \\ -\alpha_2 & -\alpha_1 & \ddots & & \vdots \\ \vdots & & & & 0 \\ -\alpha_{p-1} & -\alpha_{p-2} & \cdots & -\alpha_1 & 1, \end{bmatrix}$$

Stationarity of AR(p) processes: **An Alternative Schur Criterion**

$$A_2 = \begin{bmatrix} -\alpha_p & 0 & \cdots & 0 & 0 \\ -\alpha_{p-1} & -\alpha_p & & & 0 \\ -\alpha_{p-2} & -\alpha_{p-1} & & & \vdots \\ \vdots & & \ddots & & 0 \\ -\alpha_1 & -\alpha_2 & \cdots & -\alpha_{p-1} & -\alpha_p \end{bmatrix}.$$

Then, the AR (p) process is covariance stationary if and only if the so-called *Schur matrix*, defined by

$$S_a = A_1 A_1' - A_2 A_2', \tag{44}$$

is positive definite.

Stationarity of AR(1) processes: **An Alternative Schur Criterion**

For

$$y_t = \alpha_1 y_{t-1} + \epsilon_t$$

we get $A_1 = [1]$ and $A_2 = [-\alpha_1]$

$$|S_a| = 1 \cdot 1' - (-\alpha_1) \cdot (-\alpha_1)' = 1 - \alpha_1^2 > 0 \Longleftrightarrow |\alpha_1| < 1$$

Stationarity of AR(2) processes: **An Alternative Schur Criterion**

For

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t$$

we get

$$A_1 = \begin{bmatrix} 1 & 0 \\ -\alpha_1 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} -\alpha_2 & 0 \\ -\alpha_1 & -\alpha_2 \end{bmatrix}$$

$$S_a = \begin{bmatrix} 1 - \alpha_2^2 & -\alpha_1 - \alpha_2\alpha_1 \\ -\alpha_1 - \alpha_2\alpha_1 & 1 - \alpha_2^2 \end{bmatrix}$$

Stationarity of an AR(2) Process

For an AR(2) process covariance stationarity requires that the AR coefficients satisfy

$$\begin{aligned} |\alpha_2| &< 1, \\ \alpha_2 + \alpha_1 &< 1, \\ \alpha_2 - \alpha_1 &< 1. \end{aligned} \tag{45}$$

Stationarity of MA(q) Processes

Pure MA processes are always stationary, because it has no autoregressive roots.

Stationarity of ARMA(p, q) Processes

The stationarity property of the mixed ARMA process

$$a(L)y_t = b(L)\epsilon_t \tag{46}$$

does not dependent on the values of the MA parameters. Stationarity is a property that depends solely on the AR parameters.

Stationarity: Examples

	α_1	α_2	Stationary?
AR(1)	0.5		
AR(1)	-0.99		
AR(1)	1		
AR(1)	1.5		
AR(2)	0.5	0.4	
AR(2)	0.2	-0.9	
AR(2)	1.5	-0.5	

$\Rightarrow$ Same conclusions for ARMA models with q MA lags with arbitrary parameters (β_i).

Examples

Are the following process stationary? Employ the Schur-Criterion:

$$y_t = 0.5y_{t-1} + \epsilon_t \tag{47}$$
$$y_t = 0.5 + 0.5y_{t-1} + \epsilon_t \tag{48}$$
$$y_t = 0.5 - 0.5y_{t-1} + \epsilon_t \tag{49}$$
$$y_t = 0.5 + 0.5y_{t-1} + 0.5y_{t-2} + \epsilon_t \tag{50}$$
$$y_t = 0.5 + 0.5y_{t-1} + 0.5y_{t-2} - 0.8y_{t-3} + 0.5\epsilon_{t-1} + \epsilon_t \tag{51}$$

Autocovariance and Autocorrelation Functions

How to determine the order of an ARMA(p, q) process?

- Useful tools are the
  - *sample autocovariance function* (SACovF)
  - and its scaled counterpart *sample autocorrelation function* (SACF)

Deriving the ACovF and ACF for an AR(1) Process

Derive the Autocovariance Function for an AR(1) process.

$$y_t = ay_{t-1} + \epsilon_t, \tag{52}$$

where ϵ_t is the usual white–noise process with $\mathrm{E}(\epsilon_t^2) = \sigma^2$.

Deriving the ACovF and ACF for an AR(1) Process

Consider the stationary AR(1) process

$$y_t = a y_{t-1} + \epsilon_t, \tag{53}$$

where ϵ_t is the usual white–noise process with $\mathrm{E}(\epsilon_t^2) = \sigma^2$.
To obtain the variance $\gamma_0 = \mathrm{E}(y_t^2)$, multiply both sides of (52) by y_t,

$$y_t^2 = a y_t y_{t-1} + y_t \epsilon_t,$$

and take expectations, i.e.,

$$\mathrm{E}(y_t^2) = a\mathrm{E}(y_t y_{t-1}) + \mathrm{E}(y_t \epsilon_t)$$

or

$$\gamma_0 = a\gamma_1 + \mathrm{E}(y_t \epsilon_t).$$

Deriving the ACovF and ACF for an AR(1) Process

Thus, to specify γ_0, we have to determine γ_1 and $\mathrm{E}(y_t\epsilon_t)$. To obtain the latter quantity, substitute the RHS of (52) for y_t,

$$\begin{aligned} \mathrm{E}(y_t\epsilon_t) &= \mathrm{E}[(ay_{t-1}+\epsilon_t)\epsilon_t] \\ &= a\mathrm{E}(y_{t-1}\epsilon_t)+\mathrm{E}(\epsilon_t^2). \end{aligned}$$

Since y_{t-1} is independent of the future disturbances ϵ_{t+i}, $i=0,1,\ldots$, $\mathrm{E}(y_{t-1}\epsilon_t)=0$ and $\mathrm{E}(\epsilon_t^2)=\sigma^2$,

$$\mathrm{E}(\epsilon_t y_t)=\sigma^2.$$

Therefore,

$$\gamma_0 = a\gamma_1+\sigma^2. \tag{54}$$

Deriving the ACovF and ACF for an AR(1) Process

To determine $\gamma_1 = E(y_t y_{t-1})$, we basically repeat the above procedure. Multiplying (52) by y_{t-1} and taking expectations on both sides gives

$$\mathrm{E}(y_t y_{t-1}) = a\mathrm{E}(y_{t-1}^2) + \mathrm{E}(y_{t-1}\epsilon_t).$$

Using $\mathrm{E}(y_{t-1}\epsilon_t) = 0$ and the fact that stationarity implies that $\mathrm{E}(y_{t-1}^2) = \mathrm{E}(y_t^2) = \gamma_0$, we have

$$\gamma_1 = a\gamma_0. \tag{55}$$

Deriving the ACovF and ACF for an AR(1) Process

Substituting (55) into (54) and solving for γ_0 gives the expression for the theoretical variance of an AR(1) process, which we derived in the previous section,

$$\gamma_0 = \frac{\sigma^2}{1 - a^2}. \tag{56}$$

It follows from (55) that

$$\gamma_1 = a\frac{\sigma^2}{1 - a^2}. \tag{57}$$

Deriving the ACovF and ACF for an AR(1) Process

In fact, since

$$\mathrm{E}(y_t y_{t-k}) = a\mathrm{E}(y_{t-1} y_{t-k}) + \mathrm{E}(\epsilon_t y_{t-k}), \qquad k = 1, 2, \ldots,$$

and $\mathrm{E}(\epsilon_t y_{t-k}) = 0$, for $k = 1, 2, \ldots$, first and higher-order autocovariances are derived recursively by

$$\gamma_k = a\gamma_{k-1}, \qquad k = 1, 2, \ldots . \tag{58}$$

It is obvious that the recursive relationship (58) holds also for the autocorrelation function, $\rho_k = \gamma_k/\gamma_0$, of the AR(1) process, i.e., $\rho_k = a\rho_{k-1}$, for $k = 1, 2, \ldots$.

Deriving the ACov and ACF for an ARMA(1,1) Process

Consider the stationary, zero-mean ARMA(1,1) process

$$y_t = ay_{t-1} + \epsilon_t + b\epsilon_{t-1}, \tag{59}$$

where ϵ_t is again an white–noise process with variance σ^2.

Deriving the ACov and ACF for an ARMA(1,1) Process

As in the previous example, multiplying (59) by y_t and taking expectations yields

$$\gamma_0 = a\gamma_1 + \mathrm{E}[y_t(\epsilon_t + b\epsilon_{t-1})]. \tag{60}$$

To determine $\mathrm{E}[y_t(\epsilon_t + b\epsilon_{t-1})]$, replace y_t by the right hand side of (59), i.e.,

$$\begin{aligned} \mathrm{E}[y_t(\epsilon_t + b\epsilon_{t-1})] &= \mathrm{E}[(ay_{t-1} + \epsilon_t + b\epsilon_{t-1})(\epsilon_t + b\epsilon_{t-1})] \\ &= \mathrm{E}(ay_{t-1}\epsilon_t + \epsilon_t^2 + b\epsilon_{t-1}\epsilon_t + aby_{t-1}\epsilon_{t-1} \\ &\quad + b\epsilon_t\epsilon_{t-1} + b^2\epsilon_{t-1}^2) \\ &= \sigma^2 + ab\sigma^2 + b^2\sigma^2. \end{aligned}$$

Deriving the ACov and ACF for an ARMA(1,1) Process

Taking the expectation operator inside the parentheses and noting the fact that $\mathrm{E}(y_{t-1}\epsilon_t) = \mathrm{E}(\epsilon_{t-1}\epsilon_t) = 0$ and $\mathrm{E}(y_{t-1}\epsilon_{t-1}) = \sigma^2$, we have

$$\mathrm{E}[y_t(\epsilon_t + b\epsilon_{t-1})] = (1 + ab + b^2)\sigma^2. \tag{61}$$

Multiplying (59) by y_{t-1} and taking expectations gives

$$\begin{aligned}\gamma_1 &= \mathrm{E}[ay_{t-1}^2 + y_{t-1}(\epsilon_t + b\epsilon_{t-1})] \\ &= a\gamma_0 + b\sigma^2.\end{aligned}$$

Deriving the ACov and ACF for an ARMA(1,1) Process

Combining (60)–(62) and solving for γ_0 gives us the formula for the variance of an ARMA(1,1) process

$$\gamma_0 = \frac{1 + 2ab + b^2}{1 - a^2}\sigma^2. \tag{62}$$

For the first order autocovariance we obtain from (59) and (60)

$$\begin{aligned} \gamma_1 &= \left(\frac{a(1 + 2ab + b^2)}{1 - a^2} + b^2 \right) \sigma^2 \\ &= \frac{(1 + ab)(a + b)}{1 - a^2}\sigma^2. \end{aligned} \tag{63}$$

Deriving the ACov and ACF for an ARMA(1,1) Process

Higher–order autocovariances can be computed recursively by

$$\gamma_k = a\gamma_{k-1}, \qquad k = 2, 3, \ldots \ . \tag{64}$$

Excursion: The AcovF for a general ARMA(p, q) process

Let y_t be generated by the stationary ARMA (p, q) process

$$a(L)y_t = b(L)\epsilon_t, \tag{65}$$

where ϵ_t is the usual white–noise process with $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{E}(\epsilon_t^2) = \sigma^2$; and $a(L)$ and $b(L)$ are polynomials defined by $a(L) = 1 - \alpha_1 L - \ldots - \alpha_r L^r$ and $b(L) = \beta_0 + \beta_1 L + \ldots + \beta_r L^r$, with $r = \max(p, q)$ and $\alpha_i = 0$ for $i = p+1, p+2, \ldots, r$, if $r > p$ or $\beta_i = 0$ for $i = q+1, q+2, \ldots, r$, if $r > q$.

Excursion: The AcovF for a general ARMA(p, q) process

From the definition of the autocovariance, $\gamma_k = \mathrm{E}(y_t y_{t-k})$, it follows that

$$\begin{aligned} \gamma_k &= \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \ldots + \alpha_r \gamma_{k-r} \\ &\quad + \mathrm{E}(\beta_0 \epsilon_t y_{t-k} + \beta_1 \epsilon_{t-1} y_{t-k} + \ldots + \beta_r \epsilon_{t-r} y_{t-k}), \qquad k = 0, 1, \ldots \end{aligned} \quad ($$

Replacing y_{t-k} by its moving average representation, $y_{t-k} = b(L)/a(L)\epsilon_{t-k} = c(L)\epsilon_{t-k}$, where $c(L) = c_0 + c_1 L + c_2 L^2 \ldots$, we obtain

$$\mathrm{E}(\epsilon_{t-i} y_{t-k}) = \begin{cases} c_{i-k}\sigma^2, & \text{if } i = k, k+1, \ldots, r, \\ 0, & \textit{otherwise}. \end{cases}$$

Excursion: The AcovF for a general ARMA(p, q) process

Defining $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_r)'$, $c = (c_0, c_1, \ldots, c_r)'$ and using the fact that $\gamma_{k-i} = \gamma_{i-k}$, expression (66) can be rewritten in matrix terms as

$$\gamma = M_a \gamma + N_b c \sigma^2. \tag{67}$$

The $(r+1) \times (r+1)$ matrix M_a is the sum of two matrices, $M_a = T_a + H_a$, with T_a denoting the lower-triangular Toeplitz matrix

$$T_a = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ \alpha_1 & 0 & & & 0 \\ \alpha_2 & \alpha_1 & \ddots & & \vdots \\ \vdots & & & & 0 \\ \alpha_r & \alpha_{r-1} & \cdots & \alpha_1 & 0 \end{bmatrix},$$

Excursion: The AcovF for a general ARMA(p, q) process

and H_a is “almost" a Hankel matrix and given by

$$H_a = \begin{bmatrix} 0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{r-1} & \alpha_r \\ 0 & \alpha_2 & \alpha_3 & \cdots & \alpha_r & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & \alpha_{r-1} & \alpha_r & & & 0 \\ 0 & \alpha_r & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Excursion: The AcovF for a general ARMA(p, q) process

Note that matrix H_a is not exactly Hankel due to the zeros in the first column. Finally, the Hankel matrix N_b is defined by

$$N_b = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_{r-1} & \beta_r \\ \beta_1 & \beta_2 & \cdots & \beta_r & 0 \\ \vdots & & & & \vdots \\ \beta_{r-1} & \beta_r & & & 0 \\ \beta_r & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Excursion: The AcovF for a general ARMA(p, q) process

The initial autocovariances can be computed by

$$\gamma = (I - M_a)^{-1} N_b c \sigma^2. \tag{68}$$

Since $c = (I - T_a)^{-1} b$, a closed-form expression, relating the autocovariances of an ARMA process to its parameters α_i, β_i, and σ^2 is given by

$$\gamma = (I - M_a)^{-1} N_b (I - T_a)^{-1} b \sigma^2. \tag{69}$$

Excursion: The AcovF for a general ARMA(p, q) process

Note that $(I - T_a)^{-1}$ always exists, since $| I - T_a |= 1$, and that

$$N_b(I - T_a)^{-1} = [(I - T_a)^{-1}]' N_b,$$

since N_b is Hankel with zeros below the main counterdiagonal and $(I - T_a)^{-1}$ is a lower-triangular Toeplitz matrix. Hence, (69) can finally be rewritten as

$$\gamma = [(I - T_a')(I - M_a)]^{-1} N_b b \sigma^2. \tag{70}$$

Excursion: The AcovF for a general ARMA(p, q) process

Note that for $p < q = r$ only $p + 1$ equations have to be solved simultaneously. The corresponding system of equations is obtained by eliminating the last $p - q$ rows in (67); and higher–order autocovariances can be derived recursively by

$$\gamma_k = \begin{cases} \sum_{i=1}^{p} \alpha_i \gamma_{k-i} + \sigma^2 \sum_{j=k}^{q} \beta_j c_{j-k}, & \text{if } k = p+1, p+2, \ldots, q, \\ \sum_{i=1}^{p} \alpha_i \gamma_{k-i}, & \text{if } k = q+1, q+2, \ldots . \end{cases} \tag{71}$$

Excursion: The AcovF for a general ARMA(p, q) process

For pure autoregressive processes expression (70) reduces to

$$\gamma = [(I - T_a')(I - M_a)]^{-1} s, \tag{72}$$

where the $(r+1) \times 1$ vector s is defined by $s = \sigma^2(\beta_0, 0, \ldots, 0)^T$. Thus, vector γ is given by the first column of $[(I - T_a')(I - M_a)]^{-1}$ multiplied by $\sigma^2 \beta_0$.

Excursion: The AcovF for a general ARMA(p, q) process

In the case of a pure MA process, (70) simplifies to

$$\gamma = N_b b \sigma^2, \tag{73}$$

or

$$\gamma_k = \begin{cases} \sigma^2 \sum_{i=k}^{q} \beta_i \beta_{i-k}, & \text{if } k = 0, 1, \ldots, q, \\ 0, & \text{if } k > q. \end{cases} \tag{74}$$

The AcovF of an ARMA(1,1) reconsidered

Consider again the ARMA(1,1) process $y_t = \alpha_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1}$ from Example 3.4.2. To compute $\gamma = (\gamma_0, \gamma_1)'$, we now apply formula (70). Matrices T_a, H_a, N_b and vector b become:

$$T_a = \begin{bmatrix} 0 & 0 \\ \alpha_1 & 0 \end{bmatrix}, \quad H_a = \begin{bmatrix} 0 & \alpha_1 \\ 0 & 0 \end{bmatrix}, \quad N_b = \begin{bmatrix} 1 & \beta_1 \\ \beta_1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ \beta_1 \end{bmatrix}.$$

The AcovF of an ARMA(1,1) reconsidered

Simple matrix manipulations produce the desired result:

$$
\begin{aligned}
\gamma &= [(I - T_a')(I - M_a)]^{-1} N_b b \sigma^2 \\
&= \begin{bmatrix} 1+\alpha_1^2 & -2\alpha_1 \\ -\alpha_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & \beta_1 \\ \beta_1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \beta_1 \end{bmatrix} \sigma^2 \\
&= \frac{1}{1-\alpha_1^2} \begin{bmatrix} 1 & 2\alpha_1 \\ \alpha_1 & 1+\alpha_1^2 \end{bmatrix} \begin{bmatrix} 1+\beta_1^2 \\ \beta_1 \end{bmatrix} \sigma^2 \\
&= \frac{\sigma^2}{1-a^2} \begin{bmatrix} 1+\beta_1^2+2\alpha_1\beta_1 \\ \alpha_1(1+\beta_1^2)+\beta_1(1+\alpha_1^2) \end{bmatrix},
\end{aligned}
$$

which coincides with results (62) and (63) in the previous example.

An Example

Derive γ_0 and γ_1 using the stated procedure for the following process

$$y_t = 0.5y_{t-1} + \epsilon_t \tag{75}$$

with $\epsilon_t \sim N(0, 1)$.

An Example

Find γ_i for $i = 0, \ldots 3$ for the following process:

$$y_t = 0.5y_{t-1} + 0.5\epsilon_{t-1} + \epsilon_t \tag{76}$$

with $\epsilon_t \sim N(0, 1)$.

The Yule-Walker Equations

Consider the AR(p) process

$$y_t = \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \epsilon_t$$

Multiplying both sides with y_{t-j} and taking expectations yields

$$\mathrm{E}(y_t y_{t-j}) = \alpha_1 \mathrm{E}(y_{t-1} y_{t-j}) + \ldots + \alpha_p \mathrm{E}(y_{t-p} y_{t-j})$$

which gives rise to the following equation system

$$\begin{aligned}
\gamma_1 &= \alpha_1\gamma_0 + \alpha_2\gamma_1 + \ldots + \alpha_p\gamma_{p-1} \\
\gamma_2 &= \alpha_1\gamma_1 + \alpha_2\gamma_0 + \ldots + \alpha_p\gamma_{p-2} \\
& \ldots \\
\gamma_p &= \alpha_1\gamma_{p-1} + \alpha_2\gamma_{p-2} + \ldots + \alpha_p\gamma_0
\end{aligned}$$

The Yule-Walker Equations

Or in matrix notation

$$\gamma = a\Gamma$$

with

$$\Gamma = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots \gamma_{p-2} \\ \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots \gamma_0 \end{bmatrix}$$

We obtain a similar structure for the autocorrelation function by dividing by γ_0.

Partial Autocorrelation Function

The *partial autocorrelation function* (PACF) represents an additional tool for portraying the properties of an ARMA process. The definition of a *partial correlation coefficient* eludes to the difference between the PACF and the ACF. The ACF $\rho_k, k = 0, \pm 1, \pm 2, \ldots$, represents the *unconditional correlation* between y_t and y_{t-k}. By *unconditional correlation* we mean the correlation between y_t and y_{t-k} without taking the influence of the intervening variables $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$ into account.

Partial Autocorrelation Function

The PACF, denoted by $\alpha_{kk}, k = 1, 2, \ldots$, reflects the net association between y_t and y_{t-k} over and above the association of y_t and y_{t-k} which is due to their common relationship with the intervening variables $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$.

The PACF for an AR(1)

Consider the stationary AR(1) process

$$y_t = \alpha_1 y_{t-1} + \epsilon_t$$

Given that y_t and y_{t-2} are both correlated with y_{t-1}, we would like to know whether or not there is an additional association between y_t and y_{t-2} which goes beyond their common association with y_{t-1}.

The PACF for an AR(1)

Let ρ_{12}=Corr(y_t, y_{t-1}), ρ_{13}=Corr(y_t, y_{t-2}) and ρ_{23}=Corr(y_{t-1}, y_{t-2}). The partial correlation between y_t and y_{t-2} conditional on y_{t-1}, denoted by $\rho_{13,2}$, is

$$\rho_{13,2} = \frac{\rho_{13} - \rho_{12}\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}}.$$

The PACF for an AR(1)

Considering an AR(1) process, we know that $\rho_{12} = \rho_{23} = \alpha_1$ and $\rho_{13} = \rho_2 = \alpha_1^2$. Hence, the partial autocorrelation between y_t and y_{t-2}, $\rho_{13,2}$, is zero. Denoting the partial autocorrelation between y_t and y_{t-k} by α_{kk}, it can be easily verified that for any AR(1) process $\alpha_{kk} = 0$, for $k = 2, 3, \ldots$. Since there are no intervening variables between y_t and y_{t-1}, the first-order partial autocorrelation coefficient is equivalent to the first order autocorrelation coefficient, i.e., $\alpha_{11} = \rho_1$. In particular for an AR(1) process we have $\alpha_{11} = \alpha_1$.

The PACF for a general AR process

Another way of interpreting the PACF is to view it as the sequence of the k-th autoregressive coefficients in a k-th order autoregression. Letting $\alpha_{k\ell}$ denote the ℓ-th autoregressive coefficient of an AR(k) process, the **Yule–Walker equations**

$$\rho_\ell = \alpha_{k1}\rho_{\ell-1} + \cdots + \alpha_{k(k-1)}\rho_{\ell-k+1} + \alpha_{kk}\rho_{\ell-k}, \qquad \ell = 1, 2, \ldots, k, \tag{77}$$

The PACF for a general AR process

$$\rho_\ell = \alpha_{k1}\rho_{\ell-1} + \cdots + \alpha_{k(k-1)}\rho_{\ell-k+1} + \alpha_{kk}\rho_{\ell-k}, \qquad \ell = 1, 2, \ldots, k, \tag{78}$$

give rise to the system of linear equations

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & & \rho_{k-2} \\ \rho_2 & \rho_1 & & \rho_{k-3} \\ \vdots & & & \vdots \\ \rho_{k-2} & & & \rho_1 \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_{k1} \\ \alpha_{k2} \\ \alpha_{k3} \\ \vdots \\ \alpha_{k(k-1)} \\ \alpha_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_{k-1} \\ \rho_k \end{bmatrix}$$

or, in short,

$$P_k \alpha_k = \underline{\rho}_k, \qquad k = 1, 2, \ldots \,. \tag{79}$$

The PACF for a general AR process

Using Cramér's rule, to successively solve (79) for α_{kk}, $k = 1, 2 \ldots$, we have

$$\alpha_{kk} = \frac{|\, P_k^* \,|}{|\, P_k \,|}, \qquad k = 1, 2, \ldots, \tag{80}$$

where matrix P_k^* is obtained by replacing the last column of matrix P_k by vector $\underline{\rho}_k = (\rho_1, \rho_2, \ldots, \rho_k)'$, i.e.,

$$P_k^* = \left[\begin{array}{cccc|c} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & & \rho_{k-3} & \rho_2 \\ \rho_2 & \rho_1 & & \rho_{k-4} & \rho_3 \\ \vdots & & & \vdots & \vdots \\ \rho_{k-2} & & & 1 & \rho_{k-1} \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{array}\right]$$

The PACF for a general AR process

Applying (80), the first three terms of the PACF are given by

$$\alpha_{11} = \frac{|\rho_1|}{|1|} = \rho_1,$$

$$\alpha_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2},$$

The PACF for a general AR process

$$\alpha_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\rho_3 + \rho_1\rho_2(\rho_2 - 2) - \rho_1^2(\rho_3 - \rho_1)}{(1-\rho_2) - (1 - \rho_2 - 2\rho_1^2)}.$$

The PACF for a general AR process

From the Yule–Walker equations it is evident that $| P_k^* | = 0$ for an AR process whose order is less than k, since the last column of matrix P_k^* can always be obtained from a linear combination of the first $k - 1$ (or less) columns of P_k^*. Hence, the theoretical PACF of an AR(p) will generally be different from zero for the first p terms and exactly zero for terms of higher order. This property allows us to identify the order of a pure AR process from its PACF.

The PACF for a MA(1) process

Consider the MA(1) process $y_t = \epsilon_t + \beta_1 \epsilon_{t-1}$. Its ACF is given by

$$\rho_k = \begin{cases} \frac{\beta_1}{1+\beta_1}, & \text{if k=1}, \\ 0, & \text{if k=2,3,\ldots} \ . \end{cases}$$

Applying (80), the first 4 terms of the PACF are:

$$\begin{aligned} \alpha_{11} &= \rho_1, \ \alpha_{22} = -\frac{\rho_1^2}{1-\rho_1^2}, \\ \alpha_{33} &= \frac{\rho_1^3}{1-2\rho_1^2}, \ \alpha_{44} = -\frac{\rho_1^4}{1-3\rho_1^2+\rho_1^4}. \end{aligned} \tag{81}$$

The PACF for a MA(1) process

In fact, the general expression for the PACF of an MA(1) process in terms of the MA coefficient β_1 is

$$\alpha_{kk} = -\frac{(-\beta_1)^k(1-\beta_1^2)}{1-\beta_1^{2(k+1)}}.$$

$\Rightarrow$ PACF gradually dies out, in contrast to an AR process
$\Rightarrow$ this allows us to identify processes by looking at its corresponding ACF and PACF

Characteristics of specific processes

Identification Functions:

1. **autocorrelation function (ACF)**, ρ_k,
2. **partial autocorrelation function (PACF)**, α_{kk},

Characteristics of AR processes

- ACF*:* The Yule–Walker equations

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \ldots + \alpha_p \rho_{k-p}, \qquad k = 1, 2, \ldots$$

  imply that the ACF of a stationary AR process is generally different from zero but gradually dies out as k approaches infinity.

- PACF: The first p terms are generally different from zero; higher-order terms are identically zero.

Characteristics of MA Processes

- ACF: We know that the ACF of an MA(q) process is given by

$$\gamma_k = \begin{cases} \sigma^2 \sum_{i=k}^{q} \beta_i \beta_{i-k}, & \text{if } k = 0, 1, \dots, q \\ 0, & \text{if } k > q, \end{cases}$$

  which implies that the ACF is generally different from zero up to lag q and equal to zero thereafter.

- PACF: The PACF is computed successively by

$$\alpha_{kk} = \frac{| P_k^* |}{| P_k |}, \qquad k = 1, 2, \dots,$$

  with matrices P_k^* and P_k defined in the Section before. Example 3.6.2 demonstrated the pattern of the PACF of an MA(1) process.

ACF and PACF

	Model		
	AR(p)	MA(q)	ARMA(p, q)
ACF	tails off	cuts off after q	tails off
PACF	cuts off after p	tails off	tails off

Table: Patterns for Identifying ARMA Processes