Time Series Analysis

Andrea Beccarini

Center for Quantitative Economics

Winter 2013/2014

- Time series are ubiquitous in economics, and very important in macro economics and financial economics
- GDP, inflation rates, unemployment, interest rates, stock prices
- You will learn ...
  - the formal mathematical treatment of time series and stochastic processes
  - what the most important standard models in economics are
  - how to fit models to real world time series

- Descriptive Statistics
- Probability Theory
- Statistical Inference

Class

- Class teacher: Sarah Meyer
- Time: Tu., 12:00-14:00
- Location: CAWM 3
- Start: 22 October 2013

Material

- Course page on *Blackboard*
- Slides and class material are (or will be) downloadable

- Neusser, Klaus (2011), *Zeitreihenanalyse in den Wirtschaftswissenschaften*, 3. Aufl., Teubner, Wiesbaden.
  $\longrightarrow$ available online in the RUB-Netz
- Hamilton, James D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.
- Pfaff, Bernhard (2006), *Analysis of Integrated and Cointegrated Time Series with R*, Springer, New York.
- Schlittgen, Rainer und Streitberg, Bernd (1997), *Zeitreihenanalyse*, 7. Aufl., Oldenbourg, München.

Definition: Time series

A sequence of observations ordered by time is called time series

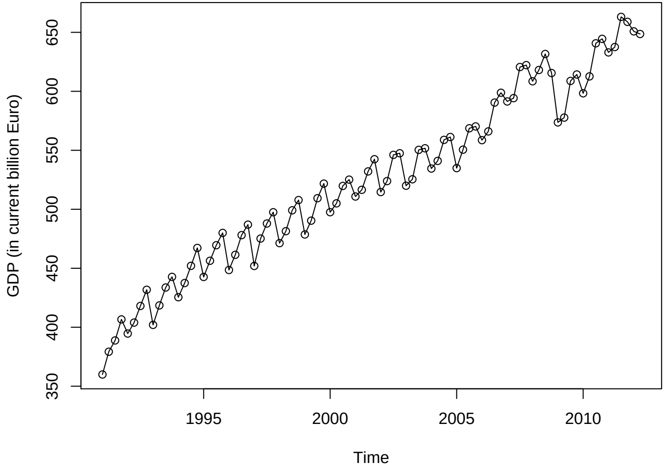
- Time series can be univariate or multivariate
- Time can be discrete or continous
- The states can be discrete or continuous

- Typical notations

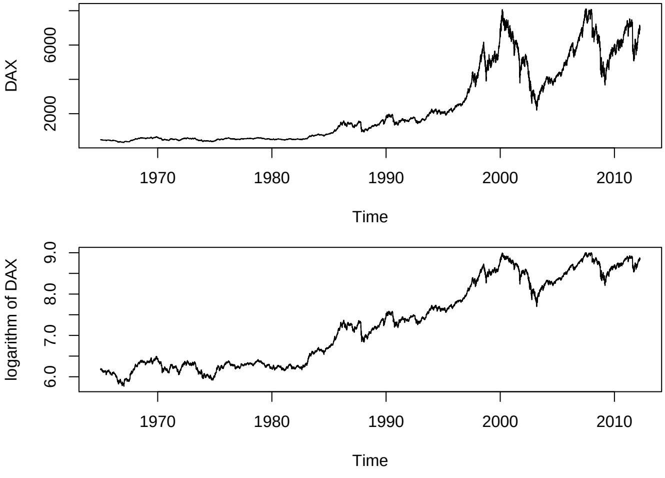
$$\begin{array}{ll} & x_1, x_2, \ldots, x_T \\ \text{or} & x(1), x(2), \ldots, x(T) \\ \text{or} & x_t, t = 1, \ldots, T \\ \text{or} & (x_t)_{t \geq 0} \end{array}$$

- This course is about ...
  - univariate time series
  - in discrete time
  - with continuous states

Quarterly GDP Germany, 1991 I to 2012 II



DAX index and log(DAX), 31.12.1964 to 6.4.2009



Definition: Stochastic process

A sequence $(X_t)_{t\in\mathbb{T}}$ of random variables, all defined on the same probability space $(\Omega, \mathcal{A}, P)$, is called stochastic process with discrete time parameter (usually $\mathbb{T} = \mathbb{N}$ or $\mathbb{T} = \mathbb{Z}$)

- Short version: A stochastic process is a sequence of random variables
- A stochastic process depends on **both** chance and time

- Distinguish four cases: both time and chance can be fixed or variable

	t fixed	t variable
ω fixed	$X_t(\omega)$ is a real number	$X_t(\omega)$ is a sequence of real numbers (path, realization, trajectory)
ω variable	$X_t(\omega)$ is a random variable	$X_t(\omega)$ is a stochastic process

- `process.R`

- **Example 1**: White noise

$$\varepsilon_t \sim NID\left(0, \sigma^2\right)$$

- **Example 2**: Random walk

$$\begin{aligned} X_t &= X_{t-1} + \varepsilon_t \quad \text{and} \quad X_0 = 0 \\ \varepsilon_t &\sim NID(0, \sigma^2) \end{aligned}$$

- **Example 3**: A random constant

$$\begin{aligned} X_t &= Z \\ Z &\sim N(0, \sigma^2) \end{aligned}$$

Definition: Moment functions

The following functions of time are called moment functions:

$$\begin{array}{rcll} \mu(t) & = & E(X_t) & \text{(expectation function)} \\ \sigma^2(t) & = & Var(X_t) & \text{(variance function)} \\ \gamma(s,t) & = & Cov(X_s, X_t) & \text{(covariance function)} \end{array}$$

- Correlation function (autocorrelation function)

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\sigma^2(s)}\sqrt{\sigma^2(t)}}$$

- `moments.R` [1]

- Usually, the moment functions are unknown and have to be estimated
- Problem: Only a **single** path (realization) can be observed

$X_1^{(1)}$	$X_1^{(2)}$	$\dots$	$X_1^{(n)}$
$X_2^{(1)}$	$X_2^{(2)}$	$\dots$	$X_2^{(n)}$
$\vdots$	$\vdots$	$\dots$	$\vdots$
$X_T^{(1)}$	$X_T^{(2)}$	$\dots$	$X_T^{(n)}$

- Can we still estimate the expectation function $\mu(t)$ and the autocovariance function $\gamma(s,t)$? Under which conditions?

$X_1^{(1)}$	$X_1^{(2)}$	$\dots$	$X_1^{(n)}$
$X_2^{(1)}$	$X_2^{(2)}$	$\dots$	$X_2^{(n)}$
$\vdots$	$\vdots$	$\dots$	$\vdots$
$X_T^{(1)}$	$X_T^{(2)}$	$\dots$	$X_T^{(n)}$

Usually, the expectation function $\mu(t)$ should be estimated by **averaging over realizations**,

$$\hat{\mu}(t) = \frac{1}{n} \sum_{i=1}^{n} X_t^{(i)}$$

$X_1^{(1)}$	$X_1^{(2)}$	$\dots$	$X_1^{(n)}$
$X_2^{(1)}$	$X_2^{(2)}$	$\dots$	$X_2^{(n)}$
$\vdots$	$\vdots$	$\dots$	$\vdots$
$X_T^{(1)}$	$X_T^{(2)}$	$\dots$	$X_T^{(n)}$

Under certain conditions, $\mu(t)$ can be estimated by **averaging over time**,

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} X_t{}^{(1)}$$

$X_1^{(1)}$	$X_1^{(2)}$	$\dots$	$X_1^{(n)}$
$X_2^{(1)}$	$X_2^{(2)}$	$\dots$	$X_2^{(n)}$
$\vdots$	$\vdots$	$\dots$	$\vdots$
$X_T^{(1)}$	$X_T^{(2)}$	$\dots$	$X_T^{(n)}$

Usually, the autocovariance $\gamma(t, t+h)$ should be estimated by **averaging over realizations**,

$$\hat{\gamma}(t, t+h) = \frac{1}{n} \sum_{i=1}^{n} (X_t^{(i)} - \hat{\mu}(t))(X_{t+h}^{(i)} - \hat{\mu}(t+h))$$

$X_1^{(1)}$	$X_1^{(2)}$	$\ldots$	$X_1^{(n)}$
$X_2^{(1)}$	$X_2^{(2)}$	$\ldots$	$X_2^{(n)}$
$\vdots$	$\vdots$	$\ldots$	$\vdots$
$X_T^{(1)}$	$X_T^{(2)}$	$\ldots$	$X_T^{(n)}$

Under certain conditions, $\gamma(t, t+h)$ can be estimated by **averaging over time**,

$$\hat{\gamma}(t, t+h) = \frac{1}{T} \sum_{t=1}^{T-h} ({X_t}^{(1)} - \hat{\mu})({X_{t+h}}^{(1)} - \hat{\mu})$$

- Moment functions cannot be estimated without additional assumptions since only **one path** is observed
- There are restrictions which allow to estimate the moment functions
- **Restriction of the time heterogeneity**:
  The distribution of $(X_t(\omega))_{t\in\mathbb{T}}$ must not be completely different for each $t \in \mathbb{T}$
- **Restriction of the memory**:
  If the values of the process are coupled too closely over time, the individual observations do not supply any (or only insufficient) information about the distribution

Definition: Strong stationarity

Let $(X_t)_{t\in\mathbb{T}}$ be a stochastic process, and let $t_1, \ldots, t_n \in \mathbb{T}$ be an arbitrary number of $n \in \mathbb{N}$ arbitrary time points.

$(X_t)_{t\in\mathbb{T}}$ is called strongly stationary if for arbitrary $h \in \mathbb{Z}$

$$P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n) = P(X_{t_1+h} \leq x_1, \ldots, X_{t_n+h} \leq x_n)$$

Implication: all univariate marginal distributions are identical

Definition: Weak stationarity

$(X_t)_{t\in\mathbb{T}}$ is called weakly stationary if

1. the expectation exists and is constant: $E(X_t) = \mu < \infty$ for all $t \in \mathbb{T}$
2. the variance exists and is constant: $Var(X_t) = \sigma^2 < \infty$ for all $t \in \mathbb{T}$
3. for all $t, s, r \in \mathbb{Z}$ (in admissible range)

$$\gamma(t,s) = \gamma(t+r, s+r)$$

Simplified notation for covariance and correlation functions

$$\begin{aligned} \gamma(h) &= \gamma(t, t+h) \\ \rho(h) &= \rho(t, t+h) \end{aligned}$$

- Strong stationarity implies weak stationarity
  (but only if the first two moments exist)
- A stochastic process is called Gaussian if the joint distribution of $X_{t_1}, \ldots, X_{t_n}$ is multivariate normal
- For Gaussian processes, weak and strong stationarity coincide
- Intuition: An observed time series can be regarded as a realization of a stationary process, if a gliding window of „appropriate width“ always displays „qualitatively the same“ picture
- `stationary.R`
- **Examples** [2]

Definition: Ergodicity (I)

Let $(X_t)_{t\in\mathbb{T}}$ be a weakly stationary stochastic process with expectation μ and autocovariance $\gamma(h)$; define

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} X_t$$

$(X_t)_{t\in\mathbb{T}}$ is called (expectation) ergodic, if

$$\lim_{T\to\infty} E\left[(\hat{\mu}_T - \mu)^2\right] = 0$$

Definition: Ergodicity (II)

Let $(X_t)_{t\in\mathbb{T}}$ be a weakly stationary stochastic process with expectation μ and autocovariance $\gamma(h)$; define

$$\hat{\gamma}(h) = \frac{1}{T}\sum_{t=1}^{T-h}(X_t - \mu)(X_{t+h} - \mu)$$

$(X_t)_{t\in\mathbb{T}}$ is called (covariance) ergodic, if for all $h \in \mathbb{Z}$

$$\lim_{T\to\infty} E\left[(\hat{\gamma}(h) - \gamma(h))^2\right] = 0$$

- Ergodicity is consistency (in quadratic mean) of the estimators $\hat{\mu}$ of μ and $\hat{\gamma}(h)$ of $\gamma(h)$ for **dependent** observations
- The process $(X_t)_{t\in\mathbb{T}}$ is expectation ergodic if $(\gamma(h))_{h\in\mathbb{Z}}$ is absolutely summable, i.e.

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$$

- The dependence between far away observations must be sufficiently small

- Ergodicity condition (for autocovariance): A stationary **Gaussian** process $(X_t)_{t\in\mathbb{T}}$ with absolutely summable autocovariance function $\gamma(h)$ is (autocovariance) ergodic
- Under ergodicity, the law of large numbers holds even if the observations are dependent
- If the dependence $\gamma(h)$ does not diminish fast enough, the estimators are no longer consistent
- **Examples** [3]

- Summary of estimators `electricity.R`

$$
\begin{aligned}
\hat{\mu} &= \bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t \\
\hat{\gamma}(h) &= \frac{1}{T}\sum_{t=1}^{T-h}(X_t - \hat{\mu})(X_{t+h} - \hat{\mu}) \\
\hat{\rho}(h) &= \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}
\end{aligned}
$$

- Sometimes, $\hat{\gamma}(h)$ is defined with factor $1/(T-h)$

A closer look at the expectation estimator $\hat{\mu}$

- The estimator $\hat{\mu}$ is unbiased, i.e. $E(\hat{\mu}) = \mu$ [4]
- The variance of $\hat{\mu}$ is [5]

$$Var\left(\hat{\mu}\right) = \frac{\gamma\left(0\right)}{T} + \frac{2}{T}\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right)\gamma\left(h\right)$$

- Under ergodicity, for $T \to \infty$

$$T \cdot Var\left(\hat{\mu}\right) \to \gamma\left(0\right) + 2\sum_{h=1}^{\infty}\gamma\left(h\right) = \sum_{h=-\infty}^{\infty}\gamma(h)$$

- For Gaussian processes, $\hat{\mu}$ is normally distributed

$$\hat{\mu} \sim N(\mu, Var(\hat{\mu}))$$

and asymptotically

$$\sqrt{T}(\hat{\mu} - \mu) \to Z \sim N\left(0, \gamma(0) + 2\sum_{h=1}^{\infty} \gamma(h)\right)$$

- For non-Gaussian processes, $\hat{\mu}$ is (often) asymptotically normal

$$\sqrt{T}(\hat{\mu} - \mu) \to Z \sim N\left(0, \gamma(0) + 2\sum_{h=1}^{\infty} \gamma(h)\right)$$

A closer look at the autocovariance estimators $\hat{\gamma}(h)$

- For Gaussian processes with absolutely summable covariance function,

$$\left(\sqrt{T}\left(\hat{\gamma}(0)-\gamma(0)\right),\ldots,\sqrt{T}\left(\hat{\gamma}(K)-\gamma(K)\right)\right)'$$

is multivariate normal with expectation vector $(0,\ldots,0)'$ and

$$\begin{aligned} & T \cdot Cov\left(\hat{\gamma}(h_1),\hat{\gamma}(h_2)\right) \\ = & \sum_{r=-\infty}^{\infty}\left(\gamma(r)\gamma(r+h_1+h_2)+\gamma(r-h_2)\gamma(r+h_1)\right) \end{aligned}$$

A closer look at the autocorrelation estimators $\hat{\rho}(h)$

- For Gaussian processes with absolutely summable covariance function, the random vector

$$\left(\sqrt{T}\,(\hat{\rho}\,(0)-\rho\,(0))\,,\ldots,\sqrt{T}\,(\hat{\rho}\,(K)-\rho\,(K))\right)'$$

  is multivariate normal with expectation vector $(0,\ldots,0)'$ and a complicated covariance matrix

- Be careful: For small to medium sample sizes the autocovariance and autocorrelation estimators are biased!

- `autocorr.R`

An important special case for autocorrelation estimators:

- Let (ε_t) be a white-noise process with $Var(\varepsilon_t) = \sigma^2 < \infty$, then

$$\begin{aligned} E(\hat{\rho}(h)) &= -T^{-1} + O(T^{-2}) \\ Cov(\hat{\rho}(h_1), \hat{\rho}(h_2)) &= \begin{cases} T^{-1} + O(T^{-2}) & \text{for } h_1 = h_2 \\ O(T^{-2}) & \text{else} \end{cases} \end{aligned}$$

- For white-noise processes and long time series, the empirical autocorrelations are approximately independent normal random variables with expectation $-T^{-1}$ and variance T^{-1}

Mathematical digression (I)

Complex numbers

- Some quadratic equations do not have real solutions, e.g.

$$x^2 + 1 = 0$$

- Still it is possible (and sensible) to define solutions to such equations
- The definition in common notation is

$$i = \sqrt{-1}$$

where i is the number which, when squared, equals -1

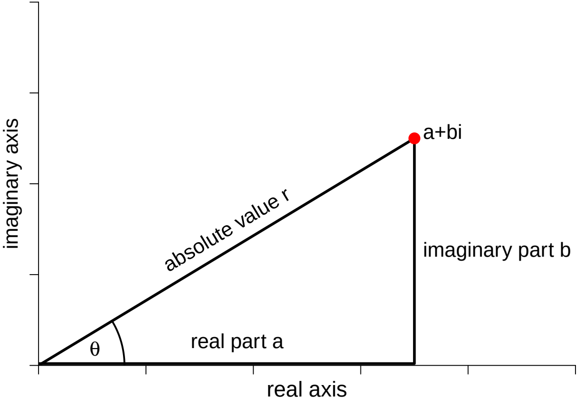
- The number i is called **imaginary** (i.e. not real)

- Other imaginary numbers follow from this definition, e.g.

$$\begin{aligned}\sqrt{-16} &= \sqrt{16}\sqrt{-1} = 4i \\ \sqrt{-5} &= \sqrt{5}\sqrt{-1} = \sqrt{5}i\end{aligned}$$

- Further, it is possible to define numbers that contain both a real part and an imaginary part, e.g. $5 - 8i$ or $a + bi$
- Such numbers are called **complex** and the set of complex numbers is denoted as $\mathbb{C}$
- The pair $a + bi$ and $a - bi$ is called conjugate complex

Geometric interpretation:



Polar coordinates and Cartesian coordinates

$$
\begin{aligned}
z &= a + bi \\
&= r \cdot (\cos\theta + i \sin\theta) \\
&= re^{i\theta} \\
a &= r\cos\theta \\
b &= r\sin\theta \\
r &= \sqrt{a^2 + b^2} \\
\theta &= \arctan\left(\left|\frac{b}{a}\right|\right)
\end{aligned}
$$

Rules of calculus:

- Addition

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

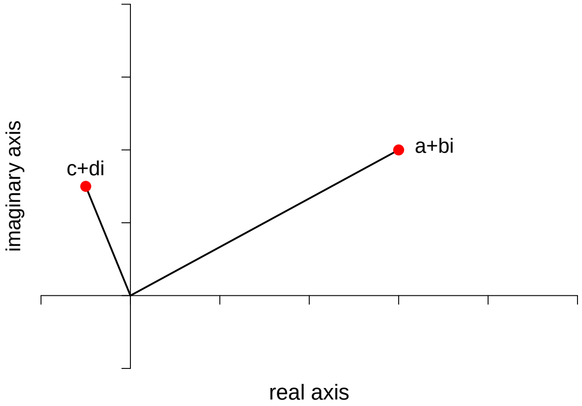
- Multiplication (cartesian coordinates)

$$(a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i$$

- Multiplication (polar coordinates)

$$r_1 e^{i\theta_1} \cdot r_2 e^{i\theta_2} = r_1 r_2 e^{i(\theta_1 + \theta_2)}$$

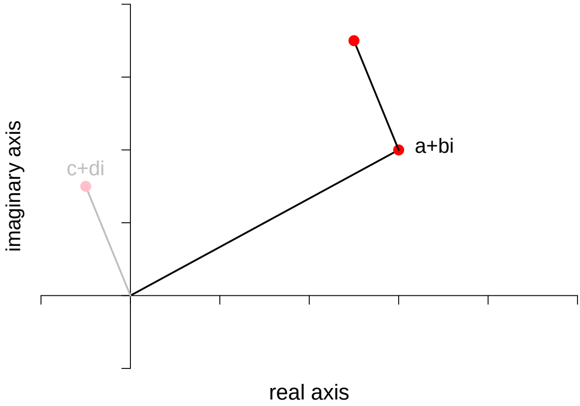
Addition:



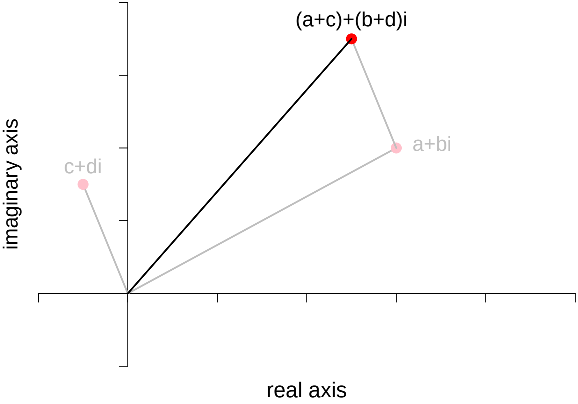
Mathematical digression (I)

Complex numbers

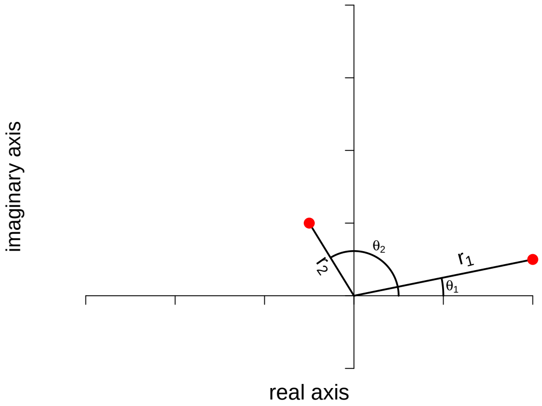
Addition:



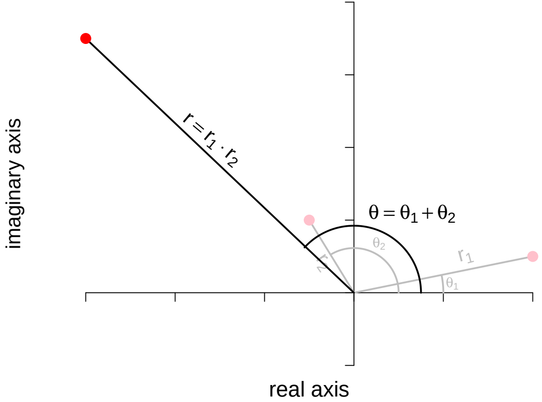
Addition:



Multiplication:



Multiplication:



- The quadratic equation

$$x^2 + px + q = 0$$

has the solutions

$$x = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}$$

- If $\frac{p^2}{4} - q < 0$ the solutions are complex (and conjugate)

Example: The solutions of

$$x^2 - 2x + 5 = 0$$

are

$$x = -\frac{(-2)}{2} + \sqrt{\frac{(-2)^2}{4} - 5} = 1 + 2i$$

and

$$x = -\frac{(-2)}{2} - \sqrt{\frac{(-2)^2}{4} - 5} = 1 - 2i$$

- First order difference equation with initial value x_0:

$$x_t = c + \phi_1 x_{t-1}$$

- p-th order difference equation with initial value x_0:

$$x_t = c + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p}$$

- A sequence $(x_t)_{t=0,1,\ldots}$ that satisfies the difference equation is called a solution of the difference equation
- **Examples** (`diffequation.R`)

- We only consider the homogeneous case, i.e. $c = 0$
- The general solution of the first-order difference equation

$$x_t = \phi_1 x_{t-1}$$

is

$$x_t = A \cdot \phi_1^t$$

with arbitrary constant A since $x_t = A\phi_1^t = \phi_1 A\phi_1^{t-1} = \phi_1 x_{t-1}$

- The constant is definitized by the initial condition, $A = x_0$
- The sequence $x_t = A\phi_1^t$ is convergent if and only if $|\phi_1| < 1$

- Solution of the p-th order difference equation

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p}$$

- Let $x_t = Az^{-t}$, then

$$\begin{aligned} Az^{-t} &= \phi_1 Az^{-(t-1)} + \ldots + \phi_p Az^{-(t-p)} \\ z^{-t} &= \phi_1 z^{-(t-1)} + \ldots + \phi_p z^{-(t-p)} \end{aligned}$$

and thus

$$1 - \phi_1 z^1 - \ldots - \phi_p z^p = 0$$

- Characteristic polynomial, characteristic equation

Mathematical digression (II)

Linear difference equations

- There are p (possibly complex, possibly nondistinct) solutions of the characteristic equation
- Denote the solutions (called roots) by $z_1, \ldots, z_p$
- If all roots are real and distinct, then

$$x_t = A_1 z_1^{-t} + \ldots + A_p z_p^{-t}$$

  is a solution of the homogeneous difference equation
- If there are complex roots the solution is oscillating
- The constants $A_1, \ldots, A_p$ can be definitized with p initial conditions $(x_0, x_{-1}, \ldots, x_{p-1})$

- **Stability condition**: The linear difference equation

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p}$$

  is stable (i.e. convergent) if and only if all roots of the characteristic polynomial

$$1 - \phi_1 z - \ldots - \phi_p z^p = 0$$

  are outside the unit circle, i.e. $|z_i| > 1$ for all $i = 1, \ldots, p$
- In R, the stability condition can be checked easily using the commands `polyroot` (base package) or `ArmaRoots` (`fArma` package)

Definition: *ARMA* process

Let $(\varepsilon_t)_{t \in \mathbb{T}}$ be a white noise process; the stochastic process

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

with $\phi_p, \theta_q \neq 0$ is called $ARMA(p, q)$ process

- **A**uto**R**egressive **M**oving **A**verage process
- *ARMA* processes are important since every stationary process can be approximated by an *ARMA* process

- The lag operator is a convenient notational tool
- The lag operator L shifts the time index of a stochastic process

$$\begin{aligned} L\left(X_t\right)_{t\in\mathbb{T}} &= \left(X_{t-1}\right)_{t\in\mathbb{T}} \\ LX_t &= X_{t-1} \end{aligned}$$

- Rules

$$\begin{aligned} L^2X_t &= L\left(LX_t\right) = X_{t-2} \\ L^nX_t &= X_{t-n} \\ L^{-1} &= X_{t+1} \\ L^0X_t &= X_t \end{aligned}$$

- Lag polynomial

$$A(L) = a_0 + a_1 L + a_2 L^2 + \ldots + a_p L^p$$

- **Example**: Let $A(L) = 1 - 0.5L$ and $B(L) = 1 + 4L^2$, then

$$\begin{aligned} C(L) &= A(L)B(L) \\ &= (1 - 0.5L)\left(1 + 4L^2\right) \\ &= 1 - 0.5L + 4L^2 - 2L^3 \end{aligned}$$

- Lag polynomials can be treated in the same way as ordinary polynomials

- Define the lag polynomials

$$\begin{aligned} \Phi(L) &= 1 - \phi_1 L - \ldots - \phi_p L^p \\ \Theta(L) &= 1 + \theta_1 L + \ldots + \theta_q L^q \end{aligned}$$

- The $ARMA(p, q)$ process can be written compactly as

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

- Important special cases

$$\begin{aligned} MA(q) \text{ process} &: \quad X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \\ AR(1) \text{ process} &: \quad X_t = \phi_1 X_{t-1} + \varepsilon_t \\ AR(p) \text{ process} &: \quad X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t \end{aligned}$$

- The $MA(q)$ process is

$$\begin{aligned} X_t &= \Theta(L)\varepsilon_t \\ X_t &= \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q} \end{aligned}$$

with $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$

- Expectation function

$$\begin{aligned} E(X_t) &= E\left(\varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}\right) \\ &= E(\varepsilon_t) + \theta_1 E(\varepsilon_{t-1}) + \ldots + \theta_q E(\varepsilon_{t-q}) \\ &= 0 \end{aligned}$$

- Autocovariance function

$$
\begin{aligned}
& \gamma(s,t) \\
= \; & E\Big[\left(\varepsilon_s + \theta_1\varepsilon_{s-1} + \ldots + \theta_q\varepsilon_{s-q}\right)\left(\varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}\right)\Big] \\
= \; & E\Big[\varepsilon_s\varepsilon_t + \theta_1\varepsilon_s\varepsilon_{t-1} + \theta_2\varepsilon_s\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_s\varepsilon_{t-q} \\
& +\theta_1\varepsilon_{s-1}\varepsilon_t + \theta_1^2\varepsilon_{s-1}\varepsilon_{t-1} + \theta_1\theta_2\varepsilon_{s-1}\varepsilon_{t-2} + \ldots + \theta_1\theta_q\varepsilon_{s-1}\varepsilon_{t-q} \\
& +\ldots \\
& +\theta_q\varepsilon_{s-q}\varepsilon_t + \theta_1\theta_q\varepsilon_{s-q}\varepsilon_{t-1} + \theta_2\theta_q\varepsilon_{s-q}\varepsilon_{t-2} + \ldots + \theta_q^2\varepsilon_{s-q}\varepsilon_{t-q}\Big]
\end{aligned}
$$

- The expectations of the cross products are

$$
E(\varepsilon_s\varepsilon_t) = \begin{cases} 0 & \text{for } s \neq t \\ \sigma_\varepsilon^2 & \text{for } s = t \end{cases}
$$

- Define $\theta_0 = 1$, then

$$\begin{aligned}
\gamma(t,t) &= \sigma_\varepsilon^2 \sum\nolimits_{i=0}^{q} \theta_i^2 \\
\gamma(t-1,t) &= \sigma_\varepsilon^2 \sum\nolimits_{i=0}^{q-1} \theta_i\theta_{i+1} \\
\gamma(t-2,t) &= \sigma_\varepsilon^2 \sum\nolimits_{i=0}^{q-2} \theta_i\theta_{i+2} \\
\gamma(t-q,t) &= \sigma_\varepsilon^2 \theta_0\theta_q = \sigma_\varepsilon^2\theta_q \\
\gamma(s,t) &= 0 \text{ for } s < t-q
\end{aligned}$$

- Hence, $MA(q)$ processes are always stationary
- Simulation of $MA(q)$ processes (`maqsim.R`)

- The $AR(1)$ process is

$$\begin{aligned}\Phi(L)X_t &= \varepsilon_t \\ (1-\phi_1 L)X_t &= \varepsilon_t \\ X_t &= \phi_1 X_{t-1} + \varepsilon_t\end{aligned}$$

  with $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$
- Expectation and variance function [6]
- Stability condition: $AR(1)$ processes are stable if $|\phi_1| < 1$

- Stationarity: Stable $AR(1)$ processes are weakly stationary if [7]

$$
\begin{aligned}
E(X_0) &= 0 \\
Var(X_0) &= \frac{\sigma_\varepsilon^2}{1-\phi_1^2}
\end{aligned}
$$

- Nonstationary stable processes converge towards stationarity [8]
- It is common parlance to call stable processes stationary
- Covariance function of stationary $AR(1)$ process [9]

- The $AR(p)$ process is

$$\begin{aligned} \Phi(L)X_t &= \varepsilon_t \\ X_t &= \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t \end{aligned}$$

  with $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$
- Assumption: ε_t is independent from X_{t-1}, X_{t-2}, ... (innovations)
- Expectation function [10]
- The covariance function is complicated (`ar2autocov.R`)

- $AR(p)$ processes are stable if all roots of the characteristic equation

$$\Phi(z) = 0$$

  are larger than 1 in absolute value, $|z_i| > 1$ for $i = 1, \ldots, p$
- An $AR(p)$ process is weakly stationary if the joint distribution of the p initial values $(X_0, X_{-1}, \ldots, X_{-(p-1)})$ is „appropriate“
- Stable $AR(p)$ processes converge towards stationarity; they are often called stationary
- Simulation of $AR(p)$ processes (`arpsim.R`)

- AR and MA processes can be inverted (into each other)
- **Example**: Consider the stable $AR(1)$ process with $|\phi_1| < 1$

$$
\begin{aligned}
X_t &= \phi_1 X_{t-1} + \varepsilon_t \\
&= \phi_1(\phi_1 X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \phi_1^2 X_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
&= \phi_1^n X_{t-n} + \phi_1^{n-1} \varepsilon_{t-(n-1)} + \ldots + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t
\end{aligned}
$$

- Since $|\phi_1| < 1$

$$
\begin{aligned}
X_t &= \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} \\
&= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots
\end{aligned}
$$

  with $\theta_i = \phi_1^i$
- A stable $AR(1)$ process can be written as an $MA(\infty)$ process (the same is true for stable $AR(p)$ processes)

- Using lag polynomials this can be written as

$$
\begin{aligned}
(1-\phi_1 L)X_t &= \varepsilon_t \\
X_t &= (1-\phi_1 L)^{-1}\varepsilon_t \\
X_t &= \sum_{i=0}^{\infty}(\phi_1 L)^i \varepsilon_t
\end{aligned}
$$

- General compact and elegant notation

$$
\begin{aligned}
\Phi(L)X_t &= \varepsilon_t \\
X_t &= (\Phi(L))^{-1}\varepsilon_t \\
&= \Theta(L)\varepsilon_t
\end{aligned}
$$

- $MA(q)$ can be written as $AR(\infty)$ if all roots of $\Theta(z) = 0$ are larger than 1 in absolute value (invertability condition)
- **Example**: $MA(1)$ with $|\theta_1| < 1$; from

$$\begin{aligned} X_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} \\ \theta_1 X_{t-1} &= \theta_1 \varepsilon_{t-1} + \theta_1^2 \varepsilon_{t-2} \end{aligned}$$

  we find $X_t = \theta_1 X_{t-1} + \varepsilon_t - \theta_1^2 \varepsilon_{t-2}$
- Repeated substitution of the ε_{t-i} terms yields

$$X_t = \sum_{i=1}^{\infty} \phi_i X_{t-i} + \varepsilon_t \qquad \text{with } \phi_i = (-1)^{i+1}\,\theta_1^i$$

Summary

- $ARMA(p,q)$ processes are stable if all roots of

$$\Phi(z) = 0$$

  are larger than 1 in absolute value
- $ARMA(p,q)$ processes are invertible if all roots of

$$\Theta(z) = 0$$

  are larger than 1 in absolute value

- Sometimes (e.g. for proofs), it is useful to write an $ARMA(p,q)$ process either as $AR(\infty)$ or as $MA(\infty)$
- $ARMA(p,q)$ can be written as $AR(\infty)$ or $MA(\infty)$

$$
\begin{aligned}
\Phi(L)X_t &= \Theta(L)\varepsilon_t \\
X_t &= (\Phi(L))^{-1}\,\Theta(L)\varepsilon_t \\
(\Theta(L))^{-1}\,\Phi(L)X_t &= \varepsilon_t
\end{aligned}
$$

- Until now we only considered processes with zero expectation
- Many processes have both a zero-expectation stochastic component (Y_t) and a non-zero deterministic component (D_t)
- **Examples**:
  - linear trend $D_t = a + bt$
  - exponential trend $D_t = ab^t$
  - saisonal patterns
- Let $(X_t)_{t\in\mathbb{Z}}$ be a stochastic process with deterministic component D_t and define $Y_t = X_t - D_t$

- Then $E(Y_t) = 0$ and

$$\begin{aligned}
Cov\left(Y_t, Y_s\right) &= E\left[\left(Y_t - E\left(Y_t\right)\right)\left(Y_s - E\left(Y_s\right)\right)\right] \\
&= E[(X_t - D_t - E(X_t - D_t))(X_s - D_s - E(X_s - D_s))] \\
&= E\left[\left(X_t - E\left(X_t\right)\right)\left(X_s - E\left(X_s\right)\right)\right] \\
&= Cov\left(X_t, X_s\right)
\end{aligned}$$

- The covariance function does not depend on the deterministic component
- To derive the covariance function of a stochastic process, simply drop the deterministic component

- Special case: $D_t = \mu_t = \mu$
- $ARMA(p,q)$ process with constant (non-zero) expectation

$$\begin{aligned} X_t - \mu &= \phi_1(X_{t-1} - \mu) + \ldots + \phi_p(X_{t-p} - \mu) \\ &\quad + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \end{aligned}$$

- The process can also be written as

$$X_t = c + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

where $c = \mu\,(1 - \phi_1 - \ldots - \phi_p)$

- Wold's representation theorem: Every stationary stochastic process $(X_t)_{t\in\mathbb{T}}$ can be represented as

$$X_t = \sum_{h=0}^{\infty} \psi_h \varepsilon_{t-h} + D_t$$

with $\psi_0 = 1$, $\sum_{h=0}^{\infty} \psi_j^2 < \infty$ and ε_t white noise with variance $\sigma^2 > 0$

- Stationary stochastic processes can be written as a sum of a deterministic process and an $MA(\infty)$ process
- Often, low order $ARMA(p, q)$ processes can approximate $MA(\infty)$ processes well

Definition: Linear process

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a white noise process; a stochastic process $(X_t)_{t\in\mathbb{Z}}$ is called linear if it can be written as

$$\begin{aligned} X_t &= \sum_{h=-\infty}^{\infty} \psi_h \varepsilon_{t-h} \\ &= \Psi(L)\varepsilon_t \end{aligned}$$

where the coefficients are absolutely summable, i.e. $\sum_{h=-\infty}^{\infty} |\psi_h| < \infty$.

The lag polynomial $\Psi(L)$ is called (linear) filter

Some special filters

- Change from previous period (difference filter)

$$\Psi(L) = 1 - L$$

- Change from last year (for quarterly or monthly data)

$$\begin{aligned} \Psi(L) &= 1 - L^4 \\ \Psi(L) &= 1 - L^{12} \end{aligned}$$

- Elimination of saisonal influences (quarterly data)

$$\begin{aligned} \Psi(L) &= \left(1 + L + L^2 + L^3\right)/4 \\ \Psi(L) &= 0.125L^2 + 0.25L + 0.25 + 0.25L^{-1} + 0.125L^{-2} \end{aligned}$$

Hodrick-Prescott filter (important tool in empirical macro economics)

- Decompose a time series (X_t) into a long-term growth component (G_t) and a short-term cyclical component (C_t)

$$X_t = G_t + C_t$$

- Trade-off between goodness-of-fit and smoothness of G_t
- Minimize the criterion function

$$\sum_{t=1}^{T} (X_t - G_t)^2 + \lambda \sum_{t=2}^{T-1} [(G_{t+1} - G_t) - (G_t - G_{t-1})]^2$$

with respect to G_t for given smoothness parameter λ

- The FOCs of the minimization problem are

$$\begin{pmatrix} G_1 \\ \vdots \\ G_T \end{pmatrix} = A \begin{pmatrix} X_1 \\ \vdots \\ X_T \end{pmatrix}$$

where $A = (I + \lambda K'K)^{-1}$ with

$$K = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix}$$

- The HP filter is a linear filter
- Typical values for smoothing parameter λ

$$\begin{array}{ll} \lambda = 10 & \text{annual data} \\ \lambda = 1600 & \text{quarterly data} \\ \lambda = 14400 & \text{monthly data} \end{array}$$

- Implementation in R (code by Olaf Posch)
- **Empirical examples** (`hpfilter.R`)

- Problem: The parameters $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_\varepsilon^2$ of an $ARMA(p, q)$ process are usually unknown
- They have to be estimated from an observed time series $X_1, \ldots, X_T$
- Standard estimation methods:
    - Least squares (OLS)
    - Maximum likelihood (ML)
- Assumption: the lag orders p and q are known

- The $AR(p)$ model with non-zero constant expectation

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

can be writte in matrix notation

$$\begin{bmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_T \end{bmatrix} = \begin{bmatrix} 1 & X_p & X_{p-1} & \ldots & X_1 \\ 1 & X_{p+1} & X_p & \ldots & X_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{T-1} & X_{T-2} & \ldots & X_{T-p} \end{bmatrix} \begin{bmatrix} c \\ \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} + \begin{bmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

- Compact notation: $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$

- The standard least squares estimator is

$$\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

- The matrix of exogenous variables $\mathbf{X}$ is stochastic
  $\longrightarrow$ usual results for OLS regression do not hold
- But: There is no contemporaneous correlation between the error term and the exogenous variables
- Hence, the OLS estimators are consistent and asymptotically efficient

- Solve the *ARMA* equation

$$X_t = c + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

for ε_t,

$$\varepsilon_t = X_t - c - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

- Define the residuals as functions of the unknown parameters

$$\begin{aligned} \hat{\varepsilon}_t \left(d, f_1, \ldots, f_p, g_1, \ldots, g_q\right) &= X_t - d - f_1 X_{t-1} - \ldots - f_p X_{t-p} \\ &\quad -g_1 \hat{\varepsilon}_{t-1} - \ldots - g_q \hat{\varepsilon}_{t-q} \end{aligned}$$

- Define the sum of squared residuals

$$S\left(d, f_1, \ldots, f_p, g_1, \ldots, g_q\right)=\sum_{t=1}^{T}\left(\hat{\varepsilon}_t\left(d, f_1, \ldots, f_p, g_1, \ldots, g_q\right)\right)^2$$

- The least squares estimators are

$$(\hat{c}, \hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q)=\arg\min S\left(d, f_1, \ldots, f_p, g_1, \ldots, g_q\right)$$

- Since the residuals are defined recursively one needs starting values $\hat{\varepsilon}_0, \ldots, \hat{\varepsilon}_{-q+1}$ and $X_0, \ldots, X_{-p+1}$ to calculate $\hat{\varepsilon}_1$
- Easiest way: Set all starting values to zero („conditional estimation“)

- The first order conditions are a nonlinear equation system which cannot be solved easily
- Minimization by standard numerical methods (implemented in all usual statistical packages)
- Either solve the nonlinear first order conditions equation system or minimize S
- Simple special case: $ARMA(1,1)$
- `arma11.R`

- Additional assumption: The innovations ε_t are normally distributed
- Implication: *ARMA* processes are Gaussian
- The joint distribution of $X_1, \ldots, X_T$ is multivariat normal

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_T \end{pmatrix} \sim N\left(\mu, \mathbf{\Sigma}\right)$$

- Expectation vector

$$\mu = E\left(\begin{bmatrix} X_1 \\ \vdots \\ X_T \end{bmatrix}\right) = \begin{pmatrix} c/\left(1-\phi_1-\ldots-\phi_p\right) \\ \vdots \\ c/\left(1-\phi_1-\ldots-\phi_p\right) \end{pmatrix}$$

- Covariance matrix

$$\boldsymbol{\Sigma} = Cov\left(\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{bmatrix}\right) = \begin{pmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(T-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma\left(T-2\right) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(T-1) & \gamma\left(T-2\right) & \ldots & \gamma(0) \end{pmatrix}$$

- The expectation vector and the covariance matrix contain all unknown parameters $\psi = \left(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, c, \sigma_\varepsilon^2\right)$
- The likelihood function is

$$L(\psi; \mathbf{X}) = (2\pi)^{-T/2} (\det \mathbf{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{X}-\mu)' \mathbf{\Sigma}^{-1} (\mathbf{X}-\mu)\right)$$

and the loglikelihood function is

$$\ln L(\psi; \mathbf{X}) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln(\det \mathbf{\Sigma}) - \frac{1}{2}(\mathbf{X}-\mu)' \mathbf{\Sigma}^{-1} (\mathbf{X}-\mu)$$

- The ML estimators are $\hat{\psi} = \arg\max \ln L(\psi; \mathbf{X})$

- The loglikelihood function has to be maximized by numerical methods
- Standard properties of ML estimators:
    1. consistency
    2. asymptotic efficiency
    3. asymptotically jointly normally distributed
    4. the covariance matrix of the estimators can be consistently estimated
- **Example**: ML estimation of an $ARMA(3,3)$ model for the interest rate spread (`arma33.R`)

- Since the estimation method is maximum likelihood, the classical tests (Wald, LR, LM) are applicable
- General null and alternative hypotheses

$$\begin{aligned} H_0 &: \ g(\psi) = 0 \\ H_1 &: \ \text{not } H_0 \end{aligned}$$

  where $g(\psi)$ is an m-valued function of the parameters
- **Example**: If $H_0 : \phi_1 = 0$ then $m = 1$ and $g(\psi) = \phi_1$

- Likelihood ratio test statistic

$$LR = 2(\ln L(\hat{\theta}_{ML}) - \ln L(\hat{\theta}_R))$$

where $\hat{\theta}_{ML}$ and $\hat{\theta}_R$ are the unrestricted and restricted estimators

- Under the null hypothesis

$$LR \xrightarrow{d} U \sim \chi^2_m$$

and H_0 is rejected at significance level α if $LR > \chi^2_{m;1-\alpha}$

- Disadvantage: Two models must be estimated

- For the Wald test we only consider $g(\psi) = \psi - \psi_0$, i.e.

$$\begin{aligned} H_0 &: \psi = \psi_0 \\ H_1 &: \text{not } H_0 \end{aligned}$$

- Test statistic

$$W = (\hat{\psi} - \psi_0)' \widehat{Cov}(\hat{\psi})(\hat{\psi} - \psi_0)$$

- If the null hypothesis is true then $W \xrightarrow{d} U \sim \chi^2_m$
- The asymptotic covariance matrix can be estimated consistently as $\widehat{Cov}(\hat{\psi}) = H^{-1}$ where H is the Hessian matrix returned by the maximization procedure

- Test **example 1**:

$$\begin{aligned} H_0 &: \phi_1 = 0 \\ H_1 &: \phi_1 \neq 0 \end{aligned}$$

- Test **example 2**

$$\begin{aligned} H_0 &: \psi = \psi_0 \\ H_1 &: \text{not } H_0 \end{aligned}$$

- Illustration (`arma33.R`)

- Usually, the lag orders p and q of an $ARMA$ model are unknown
- Trade-off: Goodness-of-fit against parsimony
- Akaike's information criterion for the model with non-zero expectation

$$AIC = \underbrace{\ln \hat{\sigma}^2}_{\text{goodness-of-fit}} + \underbrace{2\left(p+q+1\right)/T}_{\text{penalty}}$$

- Choose the model with the smallest AIC

- Bayesian information criterion BIC (Schwarz information criterion)

$$BIC = \ln \hat{\sigma}^2 + (p + q + 1) \cdot \ln T / T$$

- Hannan-Quinn information criterion

$$HQ = \ln \hat{\sigma}^2 + 2\,(p + q + 1) \cdot \ln (\ln T) / T$$

- Both BIC and HQ are consistent while the AIC tends to overfit
- Illustration (`arma33.R`)

Estimation of ARMA models

Model selection

Another illustration: The true model is $ARMA(2,1)$ with $X_t = 0.5X_{t-1} + 0.3X_{t-2} + \varepsilon_t + 0.7\varepsilon_{t-1}$; 1000 samples of size $n = 500$ were generated; the table shows the model's orders p and q as selected by AIC and BIC

	# orders selected by AIC						# orders selected by BIC					
	q						q					
p	0	1	2	3	4	5	0	1	2	3	4	5
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	18	64	23	14	6	0	310	167	4	0	0
2	0	171	21	16	5	7	0	503	3	1	0	0
3	0	7	35	58	80	45	1	0	2	1	0	0
4	9	2	12	139	37	44	6	1	0	0	0	0
5	11	6	12	56	46	56	1	0	0	0	0	0

- Define the difference operator

$$\Delta = 1 - L,$$

then

$$\Delta X_t = X_t - X_{t-1}$$

- Second order differences

$$\Delta^2 = \Delta(\Delta) = (1 - L)^2 = 1 - 2L + L^2$$

- Higher orders Δ^n are defined in the same way; note that $\Delta^n \neq 1 - L^n$

Definition: Integrated process

A stochastic process is called integrated of order 1 if

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

where ε_t is white noise, $\Psi(1) \neq 0$, and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$

- Common notation: $X_t \sim I(1)$
- $I(1)$ processes are also called difference stationary or unit root processes
- Stochastic and deterministic trends
- Trend stationary processes are not $I(1)$ (since $\Psi(1) = 0$)

- Stationary processes are sometimes called $I(0)$
- Higher order integrations are possible, e.g.

$$
\begin{aligned}
X_t &\sim I(2) \\
\Delta^2 X_t &\sim I(0)
\end{aligned}
$$

- In general, $X_t \sim I(d)$ means that $\Delta^d X_t \sim I(0)$
- Most economic time series are either $I(0)$ or $I(1)$
- Some economic time series may be $I(2)$

Example 1: The random walk with drift, $X_t = b + X_{t-1} + \varepsilon_t$, is $I(1)$ because

$$\begin{aligned} \Delta X_t &= X_t - X_{t-1} \\ &= b + \varepsilon_t \\ &= b + \Psi(L)\varepsilon_t \end{aligned}$$

where $\psi_0 = 1$ and $\psi_j = 0$ for $j \neq 0$

Example 2: The trend stationary process, $X_t = a + bt + \varepsilon_t$, is not $I(1)$ because

$$\begin{aligned} \Delta X_t &= b + \varepsilon_t - \varepsilon_{t-1} \\ &= \Psi(L)\varepsilon_t \end{aligned}$$

with $\psi_0 = 1$, $\psi_1 = -1$ and $\psi_j = 0$ for all other j

Example 3: The „$AR(2)$ process“

$$\begin{aligned} X_t &= b + (1+\phi)\, X_{t-1} - \phi X_{t-2} + \varepsilon_t \\ (1-\phi L)\,(1-L)\, X_t &= b + \varepsilon_t \end{aligned}$$

is $I(1)$ if $|\phi| < 1$ because $\Delta X_t = \Psi(L)\,(b + \varepsilon_t)$ with

$$\Psi(L) = (1-\phi L)^{-1} = 1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + \phi^4 L^4 + \dots$$

and thus $\Psi(1) = \sum_{i=0}^{\infty} \phi^i = \frac{1}{1-\phi} \neq 0$. The roots of the characteristic equation are $z = 1$ and $z = 1/\phi$

Example 4: The process

$$X_t = 0.5X_{t-1} - 0.4X_{t-2} + \varepsilon_t$$

is a stationary (stable) zero expectation $AR(2)$ process; the process

$$Y_t = a + bt + X_t$$

is trend stationary and $I(0)$ since

$$\Delta Y_t = b + \Delta X_t$$

with $\Delta X_t = \Psi(L)\varepsilon_t = (1 - L)\left(1 - 0.5L + 0.4L^2\right)^{-1}\varepsilon_t$
and therefore $\Psi(1) = 0$ (`i0andi1.R`)

Definition: *ARIMA* process

Let $(\varepsilon_t)_{t\in\mathbb{T}}$ be a white noise process; the stochastic process $(X_t)_{t\in\mathbb{Z}}$ is called integrated autoregressive moving-average process of the orders p, d and q, or $ARIMA(p, d, q)$, if $\Delta^d X_t$ is an $ARMA(p, q)$ process

$$\Phi(L)\Delta^d X_t = \Theta(L)\varepsilon_t$$

- For $d > 0$ the process is nonstationary ($I(d)$) even if all roots of $\Phi(z) = 0$ are outside the unit circle
- Simulation of an $ARIMA(p, d, q)$ process (`arimapdqsim.R`)

Why is it important to distinguish deterministic and stochastic trends?

- **Reason 1**: Long-term forecasts and forecasting errors
- Deterministic trend: The forecasting error variance is bounded
- Stochastic trend: The forecasting error variance is unbounded
- Illustrations
- `i0andi1.R`

Why is it important to distinguish deterministic and stochastic trends?

- **Reason 2**: Spurious regression
- OLS regressions will show spurious relationships between time series with (deterministic or stochastic) trends
- Detrending works if the series have deterministic trends, but it does not help if the series are integrated
- Illustrations
- `spurious1.R`

- OLS estimators (and ML estimators) are consistent and asymptotically normal for stationary processes
- The asymptotic normality is lost if the processes are integrated
- We only look at the very special case

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

  with $\varepsilon_t \sim NID(0,1)$ and $X_0 = 0$
- The $AR(1)$ process is stationary if $|\phi_1| < 1$ and has a unit root if $|\phi_1| = 1$

- The usual OLS estimator of ϕ_1 is

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{T} X_t X_{t-1}}{\sum_{t=1}^{T} X_{t-1}^2}$$

- How does the distribution of $\hat{\phi}$ look like?
- Influence of ϕ and T
- Consistency?
- Asymptotic normality?
- Illustration (`phihat.R`)

- Consistency and asymptotic normality for $I(0)$ processes ($|\phi_1| < 1$)

$$\text{plim } \hat{\phi}_1 = \phi_1$$

$$\sqrt{T}\left(\hat{\phi}_1 - \phi_1\right) \xrightarrow{d} Z \sim N\left(0, 1 - \phi_1^2\right)$$

- Consistency and asymptotic normality for $I(1)$ processes ($\phi_1 = 1$)

$$\text{plim } \hat{\phi}_1 = 1$$

$$T\left(\hat{\phi}_1 - 1\right) \xrightarrow{d} V$$

where V is a nondegenerate, nonnormal random variable

- Root-T-consistency and superconsistency

- Importance to distinguish between trend stationarity and difference stationarity
- Test of hypothesis that a process has a unit root (i.e. is $I(1)$)
- Classical approaches: (Augmented) Dickey-Fuller-Test, Phillips-Perron-Test
- Basic tool: Linear regression

$$\begin{aligned} X_t &= \text{deterministics} + \phi X_{t-1} + \varepsilon_t \\ \Delta X_t &= \text{deterministics} + \underbrace{(\phi - 1)}_{=:\beta} X_{t-1} + \varepsilon_t \end{aligned}$$

- Null and alternative hypothesis

$$\begin{aligned} H_0 &: \phi = 1 \quad \text{(unit root)} \\ H_1 &: |\phi| < 1 \quad \text{(no unit root)} \end{aligned}$$

or, equivalently,

$$\begin{aligned} H_0 &: \beta = 0 \quad \text{(unit root)} \\ H_1 &: \beta < 0 \quad \text{(no unit root)} \end{aligned}$$

- Unit root tests are one-sided; explosive process are ruled out
- Rejecting the null hypothesis is evidence in favour of stationarity
- If the null hypothesis is not rejected, there could be a unit root

Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) tests

- Possible regressions

$$\begin{array}{ll} X_t = \phi X_{t-1} + \varepsilon_t & \text{or } \Delta X_t = \beta X_{t-1} + \varepsilon_t \\ X_t = a + \phi X_{t-1} + \varepsilon_t & \text{or } \Delta X_t = a + \beta X_{t-1} + \varepsilon_t \\ X_t = a + bt + \phi X_{t-1} + \varepsilon_t & \text{or } \Delta X_t = a + bt + \beta X_{t-1} + \varepsilon_t \end{array}$$

- Assumption for Dickey-Fuller test: no autocorrelation in ε_t
- If there is autocorrelation in ε_t, use the augmented DF test

- Dickey-Fuller regression, case 1: no constant, no trend

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

- Null and alternative hypotheses

$$\begin{aligned} H_0 &: \beta = 0 \\ H_1 &: \beta < 0 \end{aligned}$$

- Null hypothesis: stochastic trend without drift
- Alternative hypothesis: stationary process around zero

- Dickey-Fuller regression, case 2: constant, no trend

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

- Null and alternative hypotheses

$$\begin{aligned} H_0 &: \beta = 0 \quad \text{or} \quad H_0 : \beta = 0, a = 0 \\ H_1 &: \beta < 0 \quad \text{or} \quad H_0 : \beta < 0, a \neq 0 \end{aligned}$$

- Null hypothesis: stochastic trend without drift
- Alternative hypothesis: stationary process around a constant

- Dickey-Fuller regression, case 3: constant and trend

$$\Delta X_t = a + bt + \beta X_{t-1} + \varepsilon_t$$

- Null and alternative hypotheses

$$\begin{aligned} H_0 &: \quad \beta = 0 \quad \text{or} \quad \beta = 0, b = 0 \\ H_1 &: \quad \beta < 0 \quad \text{or} \quad \beta < 0, b \neq 0 \end{aligned}$$

- Null hypothesis: stochastic trend with drift
- Alternative hypothesis: trend stationary process

- Dickey-Fuller test statistics for single hypotheses

$$
\begin{aligned}
\text{“}\rho\text{-test”} &\quad : \qquad T \cdot \hat{\beta} \\
\text{“}\tau\text{-test”} &\quad : \qquad \hat{\beta}/\hat{\sigma}_{\hat{\phi}}
\end{aligned}
$$

- The τ-test statistic is computed in the same way as the usual t-test statistic
- Reject the null hypothesis if the test statistics are too small
- The critical values are *not* the quantiles of the t-distribution
- There are tables with the correct critical values (e.g. Hamilton, table B.6)

- The Dickey-Fuller test statistics for the joint hypotheses are computed in the same way as the usual F-test statistics
- Reject the null hypothesis if the test statistic is too large
- The critical values are *not* the quantiles of the F-distribution
- There are tables with the correct critical values (e.g. Hamilton, table B.7)
- Illustrations (`dftest.R`)

- If there is autocorrelation in ε_t the DF test does not work (`dftest.R`)
- Augmented Dickey-Fuller test (ADF test) regressions:

$$\begin{aligned}
\Delta X_t &= \gamma_1 \Delta X_{t-1} + \ldots + \gamma_p \Delta X_{t-p} + \beta X_{t-1} + \varepsilon_t \\
\Delta X_t &= a + \gamma_1 \Delta X_{t-1} + \ldots + \gamma_p \Delta X_{t-p} + \beta X_{t-1} + \varepsilon_t \\
\Delta X_t &= a + bt + \gamma_1 \Delta X_{t-1} + \ldots + \gamma_p \Delta X_{t-p} + \beta X_{t-1} + \varepsilon_t
\end{aligned}$$

- The added lagged differences capture the autocorrelation
- The number of lags p must be large enough to make ε_t white noise
- The critical values remain the same as in the no-correlation case

Further interesting topics (but we skip these)

- Phillips-Perron test
- Structural breaks and unit roots
- KPSS test of stationarity

$$
\begin{aligned}
H_0 &: X_t \sim I(0) \\
H_1 &: X_t \sim I(1)
\end{aligned}
$$

- Spurious regression: If X_t and Y_t are independent but both $I(1)$ then the regression
$$Y_t = \alpha + \beta X_t + u_t$$
will result in an estimated coefficient $\hat{\beta}$ that is significantly different from 0 with probability 1 as $T \to \infty$
- BUT: The regression
$$Y_t = \alpha + \beta X_t + u_t$$
may be sensible even though X_t and Y_t are $I(1)$
- Cointegration

Definition: Cointegration

Two stochastic processes $(X_t)_{t\in\mathbb{T}}$ and $(Y_t)_{t\in\mathbb{T}}$ are cointegrated if both processes are $I(1)$ and there is a constant β such that the process $(Y_t - \beta X_t)$ is $I(0)$

- If β is known, cointegration can be tested using a standard unit root test on the process $(Y_t - \beta X_t)$
- If β is unknown, it can be estimated from the linear regression

$$Y_t = \alpha + \beta X_t + u_t$$

and cointegration is tested using a modified unit root test on the residual process $(\hat{u}_t)_{t=1,\dots,T}$

- Let (X, Y) be a bivariate random variable with a joint density function, then

$$E(X|Y = y) = \int_{-\infty}^{\infty} x\, f_{X|Y=y}(x) dx$$

  is the conditional expectation of X given $Y = y$
- $E(X|Y)$ denotes a random variable with realization $E(X|Y = y)$ if the random variable Y realizes as y
- Both $E(X|Y)$ and $E(X|Y = y)$ are called conditional expectation

- Let (X, Y) be a bivariate random variable with a joint density function, then

$$Var(X|Y=y) = \int_{-\infty}^{\infty} (x - E(X|Y=y))^2 f_{X|Y=y}(x) dx$$

  is the conditional variance of X given $Y = y$
- $Var(X|Y)$ denotes a random variable with realization $Var(X|Y=y)$ if the random variable Y realizes as y
- Both $Var(X|Y=y)$ and $Var(X|Y)$ are called conditional variance

1. Law of iterated expectations: $E\left(E\left(X|Y\right)\right) = E\left(X\right)$
2. If X and Y are independent, then $E\left(X|Y\right) = E\left(X\right)$
3. The condition can be treated like a constant, $E\left(XY|Y\right) = Y \cdot E\left(X|Y\right)$
4. The conditional expecation is a linear operator. For $a_1, \ldots, a_n \in \mathbb{R}$

$$E\left(\sum_{i=1}^{n} a_i X_i | Y\right) = \sum_{i=1}^{n} a_i E(X_i|Y)$$

- Some economic time series show volatility clusters, e.g. stock returns, commodity price changes, inflation rates, ...
- Simple autoregressive models cannot capture volatility clusters since their conditional variance is constant
- **Example**: Stationary $AR(1)$-process, $X_t = \alpha X_{t-1} + \varepsilon_t$ with $|\alpha| < 1$; then

$$Var(X_t) = \sigma_X^2 = \frac{\sigma_\varepsilon^2}{1-\alpha^2},$$

and the conditional variance is

$$Var(X_t|X_{t-1}) = \sigma_\varepsilon^2$$

- In the following, we will focus on stock returns
- Empirical fact: squared (or absolute) returns are positively autocorrelated
- Implication: Returns are not independent over time
- The dependence is nonlinear
- How can we model this kind of dependence?

Definition: ARCH(1)-process

The stochastic process $(X_t)_{t\in\mathbb{Z}}$ is called $ARCH(1)$-process if

$$\begin{aligned} E(X_t|X_{t-1}) &= 0 \\ Var(X_t|X_{t-1}) &= \sigma_t^2 \\ &= \alpha_0 + \alpha_1 X_{t-1}^2 \end{aligned}$$

for all $t \in \mathbb{Z}$, with $\alpha_0, \alpha_1 > 0$

Often, an additional assumption is

$$X_t|\left(X_{t-1} = x_{t-1}\right) \sim N(0, \alpha_0 + \alpha_1 x_{t-1}^2)$$

- The unconditional distribution of X_t is a non-normal distribution
- Leptokurtosis: The tails are heavier than the tails of the normal distribution
- **Example** of an $ARCH(1)$-process

$$X_t = \varepsilon_t \sigma_t$$

where $(\varepsilon_t)_{t \in \mathbb{Z}}$ is white noise with $\sigma_\varepsilon^2 = 1$ and

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2}$$

- One can show that [11]

$$
\begin{aligned}
E(X_t|X_{t-1}) &= 0 \\
E(X_t) &= 0 \\
Var(X_t|X_{t-1}) &= \alpha_0 + \alpha_1 X_{t-1}^2 \\
Var(X_t) &= \alpha_0 / (1 - \alpha_1) \\
Cov(X_t, X_{t-i}) &= 0 \quad \text{for } i > 0
\end{aligned}
$$

- Stationarity condition: $0 < \alpha_1 < 1$
- The unconditional kurtosis is $3(1-\alpha_1^2)/(1-3\alpha_1^2)$ if $\varepsilon_t \sim N(0,1)$. If $\alpha_1 > \sqrt{1/3} = 0.57735$, the kurtosis does not exist. [12]

- Squared returns follow [13]

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + v_t$$

  with $v_t = \sigma_t^2(\varepsilon_t^2 - 1)$
- Thus, squared returns of $ARCH(1)$ are $AR(1)$
- The process $(v_t)_{t\in\mathbb{Z}}$ is white noise

$$\begin{aligned} E(v_t) &= 0 \\ Var(v_t) &= E(v_t^2) = const. \\ Cov(v_t, v_{t-i}) &= 0 \quad (i = 1, 2, \ldots) \end{aligned}$$

- Simulation of an $ARCH(1)$-process for $t = 1, \ldots, 2500$
- Parameters: $\alpha_0 = 0.05$, $\alpha_1 = 0.95$, start value $X_0 = 0$
- Conditional distribution: $\varepsilon_t \sim N(0, 1)$
- `archsim.R`
- Check whether the simulated time series shows the typical stylized facts of return distributions

- Of course, we do not know the true values of the model parameters α_0 and α_1
- How can we estimate the unknown parameters α_0 and α_1 ?
- Observations $X_1, \ldots, X_T$
- Because of

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + v_t$$

  a possible estimation method is OLS

- OLS estimator of α_1

$$\hat{\alpha}_1 = \frac{\sum_{t=2}^{T} \left(X_t^2 - \overline{X_t^2}\right)\left(X_{t-1}^2 - \overline{X_{t-1}^2}\right)}{\sum_{t=2}^{T} \left(X_{t-1}^2 - \overline{X_{t-1}^2}\right)^2} \approx \hat{\rho}(X_t^2, X_{t-1}^2)$$

- Careful: These estimators are only consistent if the kurtosis exists (i.e. if $\alpha_1 < \sqrt{1/3}$)
- Test of *ARCH*-effects

$$\begin{aligned} H_0 &: \alpha_1 = 0 \\ H_1 &: \alpha_1 > 0 \end{aligned}$$

- For T large, under H_0

$$\sqrt{T}\hat{\alpha}_1 \sim N(0,1)$$

- Reject H_0 if $\sqrt{T}\hat{\alpha}_1 > \Phi^{-1}(1-\alpha)$
- Second version of this test: Consider the R^2 of the regression

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + v_t,$$

then under H_0

$$T\hat{\alpha}_1^2 \approx TR^2 \overset{appr}{\sim} \chi_1^2$$

- Reject H_0 if $TR^2 > F_{\chi_1^2}^{-1}(1-\alpha)$

Definition: ARCH(p)-process

The stochastic process $(X_t)_{t\in\mathbb{Z}}$ is called $ARCH(p)$-process if

$$\begin{aligned} E(X_t|X_{t-1},\ldots X_{t-p}) &= 0 \\ Var(X_t|X_{t-1},\ldots,X_{t-p}) &= \sigma_t^2 \\ &= \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2 \end{aligned}$$

for $t \in \mathbb{Z}$, where $\alpha_i \geq 0$ for $i = 0, 1, \ldots, p-1$ and $\alpha_p > 0$

Often, an additional assumption is that

$$X_t|(X_{t-1} = x_{t-1}, \ldots, X_{t-p} = x_{t-p}) \sim N(0, \sigma_t^2)$$

- **Example** of an $ARCH(p)$-process

$$X_t = \varepsilon_t \sigma_t$$

where$(\varepsilon_t)_{t\in\mathbb{Z}}$ is white noise with $\sigma_\varepsilon^2 = 1$ and

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2}$$

- An $ARCH(p)$ process is weakly stationary if all roots of $1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_p z^p = 0$ are outside the unit circle
- Then, for all $t \in \mathbb{Z}$, $E(X_t) = 0$ and

$$Var(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i}$$

- If $(X_t)_{t\in\mathbb{Z}}$ is a stationary $ARCH(p)$ process, then $(X_t^2)_{t\in\mathbb{Z}}$ is a stationary $AR(p)$ process

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2 + v_t$$

- As to the error term,

$$\begin{aligned} E(v_t) &= 0 \\ Var(v_t) &= const. \\ Cov(v_t, v_{t-i}) &= 0 \qquad \text{for } i = 1, 2, \ldots \end{aligned}$$

- Simulating an $ARCH(p)$ is easy

- OLS estimation of

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2 + v_t$$

- Test of *ARCH*-effects

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0 \quad \text{vs} \quad H_1 : \text{not } H_0$$

- Let R^2 denote the coefficient of determination of the regression
- Under H_0, the test statistic $TR^2 \sim \chi_p^2$;
  thus reject H_0 if $TR^2 > F_{\chi_p^2}^{-1}(1-\alpha)$

- Basic idea of the maximum likelihood estimation method:
  Choose parameters such that the joint density of the observations

$$f_{X_1,\ldots,X_T}(x_1,\ldots,x_T)$$

  is maximized
- Let $X_1,\ldots,X_T$ denote a random sample from X
- The density $f_X(x;\theta)$ depends on R unknown parameters $\theta = (\theta_1,\ldots,\theta_R)$

- ML estimation of θ: Maximize the (log)likelihood function

$$
\begin{aligned}
L(\theta) &= f_{X_1,\ldots X_T}(x_1,\ldots,x_T;\theta) \\
&= \prod_{t=1}^{T} f_X(x_t;\theta) \\
\ln L(\theta) &= \sum_{t=1}^{T} \ln f_X(x_t;\theta)
\end{aligned}
$$

- ML estimate

$$\hat{\theta} = \operatorname{argmax}\left[\ln L(\theta)\right]$$

- Since observations are independent in random samples

$$f_{X_1,\ldots,X_T}(x_1,\ldots,x_T) = \prod_{t=1}^{T} f_{X_t}(x_t)$$

or

$$\begin{aligned}\ln f_{X_1,\ldots,X_T}(x_1,\ldots,x_T) &= \sum_{t=1}^{T} \ln f_{X_t}(x_t) \\ &= \sum_{t=1}^{T} \ln f_X(x_t)\end{aligned}$$

- But: *ARCH*-returns are not independent!

- Factorization with dependent observations

$$f_{X_1,\dots,X_T}(x_1,\dots,x_T) = \prod_{t=1}^{T} f_{X_t|X_{t-1},\dots,X_1}(x_t|x_{t-1},\dots,x_1)$$

or

$$\ln f_{X_1,\dots,X_T}(x_1,\dots,x_T) = \sum_{t=1}^{T} \ln f_{X_t|X_{t-1},\dots,X_1}(x_t|x_{t-1},\dots,x_1)$$

- Hence, for an $ARCH(1)$-process

$$f_{X_1,\dots,X_T}(x_1,\dots,x_T) = f_{X_1}(x_1) \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi}\sqrt{\sigma_t^2}} \exp\left(-\frac{1}{2}\left(\frac{x_t}{\sigma_t}\right)^2\right)$$

- The marginal density of X_1 is complicated but becomes negligible for large T and, therefore, will be dropped from now on
- Log-likelihood function (without initial marginal density)

$$
\begin{aligned}
& \ln L(\alpha_0, \alpha_1 | x_1, \ldots, x_T) \\
= \quad & -\frac{T}{2} \ln 2\pi - \frac{1}{2} \sum_{t=2}^{T} \ln \sigma_t^2 - \frac{1}{2} \sum_{t=2}^{T} \left(\frac{x_t}{\sigma_t} \right)^2
\end{aligned}
$$

  where $\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2$
- ML-estimation of α_0 and α_1 by numerical maximization of $\ln L(\alpha_0, \alpha_1)$ with respect to α_0 and α_1

Definition: GARCH(p,q)-process

The stochastic process $(X_t)_{t\in\mathbb{Z}}$ is called *GARCH*(p,q)-process if

$$\begin{aligned} E(X_t|X_{t-1},X_{t-2},\ldots) &= 0 \\ Var(X_t|X_{t-1},X_{t-2},\ldots) &= \sigma_t^2 \\ &= \alpha_0+\alpha_1 X_{t-1}^2+\ldots+\alpha_p X_{t-p}^2 \\ &\quad +\beta_1\sigma_{t-1}^2+\ldots+\beta_q\sigma_{t-q}^2 \end{aligned}$$

for $t\in\mathbb{Z}$ with $\alpha_i,\beta_i\geq 0$

Often, an additional assumption is that

$$(X_t|X_{t-1}=x_{t-1},X_{t-2}=x_{t-2},\ldots)\sim N(0,\sigma_t^2)$$

- Conditional variance of $GARCH(1,1)$

$$
\begin{aligned}
Var(X_t|X_{t-1},X_{t-2},\ldots) &= \sigma_t^2 \\
&= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1\sigma_{t-1}^2 \\
&= \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{i=1}^{\infty} \beta_1^{i-1} X_{t-i}^2
\end{aligned}
$$

- Unconditional variance

$$
Var(X_t) = \frac{\alpha_0}{1-\sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j}
$$

- Necessary condition for weak stationarity

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$$

- $(X_t)_{t \in \mathbb{Z}}$ has no autocorrelation
- *GARCH*-processes can be written as *ARMA*$(\max(p,q), q)$-processes in the squared returns
- **Example**: *GARCH*$(1,1)$-process with $X_t = \varepsilon_t \sigma_t$ and $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$

- Estimation of the $ARMA(\max(p,q), q)$-process in the squared returns
- Alternative (and better) method: Maximum likelihood
- For a $GARCH(1,1)$-process

$$
\begin{aligned}
& f_{X_1,\ldots,X_T}(x_1,\ldots,x_T) \\
= \quad & f_{X_1}(x_1) \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi}\sqrt{\sigma_t^2}} \exp\left(-\frac{1}{2} \left(\frac{x_t}{\sigma_t} \right)^2 \right)
\end{aligned}
$$

- Again, the density of X_1 can be neglected
- Log-Likelihood function

$$
\begin{aligned}
& \ln L(\alpha_0, \alpha_1, \beta_1 | x_1, \ldots, x_T) \\
= \quad & -\frac{T}{2} \ln 2\pi - \frac{1}{2} \sum_{t=2}^{T} \ln \sigma_t^2 - \frac{1}{2} \sum_{t=2}^{T} \left(\frac{x_t}{\sigma_t} \right)^2
\end{aligned}
$$

with $\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ and $\sigma_1^2 = 0$

- ML-estimation of α_0, α_1 and β_1 by numerical maximization

- Conditional h-step forecast of the volatility σ^2_{t+h} in a $GARCH(1,1)$ model

$$
\begin{aligned}
E\left(\sigma^2_{t+h}|X_t, X_{t-1}, \ldots\right) &= (\alpha_1+\beta_1)^h \left(\sigma^2_t - \frac{\alpha_0}{1-\alpha_1-\beta_1}\right) \\
&\quad + \frac{\alpha_0}{1-\alpha_1-\beta_1}
\end{aligned}
$$

- If the process is stationary

$$
\lim_{h\to\infty} E(\sigma^2_{t+h}|X_t, X_{t-1}, \ldots) = \frac{\alpha_0}{1-\alpha_1-\beta_1}
$$

- Simulation of *GARCH*-processes is easy; the estimation can be computer intensive

- Careful: Residuals are slightly different from what you know from OLS regressions
- Estimates: $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\mu}$
- From $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ and $X_t = \mu + \sigma_t \varepsilon_t$ we calculate the standardized residuals

$$\hat{\varepsilon}_t = \frac{X_t - \hat{\mu}}{\hat{\sigma}_t} = \frac{X_t - \hat{\mu}}{\sqrt{\hat{\alpha}_0 + \hat{\alpha}_1 X_{t-1}^2 + \hat{\beta}_1 \sigma_{t-1}^2}}$$

- Histogram of the standardized residuals

- Definition: $(X_t)_{t\in\mathbb{Z}}$ is called *AR(p)-ARCH(q)*-process if

$$\begin{aligned} X_t &= \mu + \phi_1 X_{t-1} + \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \end{aligned}$$

  where $\varepsilon_t \sim N(0, \sigma_t^2)$
- mean equation / variance equation
- Maximum likelihood estimation

There are a number of possible extensions to the GARCH model:

- Empirical fact: Negative shocks have a larger impact on volatility than positive shocks (leverage effect)
- News impact curve
- Nonnormal innovations, e.g. $\varepsilon_t \sim t_\nu$