Advanced Time Series Analysis

Lecture Notes

Prof. Dr. Joachim Grammig

Department of Econometrics, Statistics and Empirical Economics

Winter Term 2007/08

Table of contents

IV. Univariate GARCH Models (Basics)

V. Vector Autoregressions (Basics)

VI. Cointegration and Error Correction Models (Basics)

VI.1 Application of Cointegration Methods in Finance

VII. Structural Vector Autoregressive Models (Advanced)

I. Introduction to Stochastic Processes

Time Series Analysis:

We observe (economic) variables over time, hence a time series is a collection of observations indexed by the date of each observation.

Examples:

- macroeconomic variables as income, consumption, interest rates, unemployment rates,...
- financial data as stock returns, exchange rates,...

Time series techniques are therefore essential in

Economics:

- properties of macroeconomic time series
- persistence of macro shocks
- testing economic theories
- transmission of monetary policy

Finance:

- predictability of returns
- testing and estimating asset price models
- properties of price formation processes

Stochastic processes:

Economic time series are viewed as realizations of stochastic processes, that is, of a sequence of random variables over time (that are typically not independent).

Idea of randomness:

draws from distributions, no certain numbers - not deterministic but stochastic!
However, we observe only one (possible) realization of the stochastic process!

$\Rightarrow$We call $\{X_t\}$ a stochastic process or sequence of random variables

and

$\{x_t\}$ the realization of the stochastic process or sequence of real numbers (that we do observe). Hence, we have observed the specific sample $(x_1, x_2, \ldots, x_t)$.

Because of the dependencies between the random variables $\{\ldots X_{t-2}, X_{t-1} \ldots\}$ we have a "more complex" structure than in the cross- sectional case with independent random variables $\{X_1, X_2 \ldots\}$

As we have only one realization of the stochastic process, we need to reduce complexity.
$\rightarrow$ Two "required" concepts in time series analysis:

1. **stationarity:** the distribution doesn't change over time/what matters is the relative position in the sequence but the moments remain the same across time.

2. **ergodicity:** there might be dependencies of the random variables over time, but these dependencies get smaller and smaller for larger time lags.

II. Basic Concepts

—

II.1 Mathematical Techniques of Time Series Analysis [Hamilton (1994), Appendix A]

Required techniques:

Complex numbers, unit circle, employing difference- and lag operators, solving stochastic difference equations

Unit circle

Basics:

The algebraic equation

$$x^2 - 2ax + (a^2 + b^2) = 0$$

has the following formal solution:

$$x = a \pm b\sqrt{-1}$$

But these solutions are defined in the numerical range of real numbers just for $b = 0$.

Solution

Definition of a set $\mathbb{C}$ which contains complex numbers $\mathbb{R} \subset \mathbb{C}$

Requirements for the set $\mathbb{C}$:

1. The sum (product) of real numbers as elements of $\mathbb{C}$ is identical with the sum (product) that is defined for real numbers.
2. The set $\mathbb{C}$ contains an element with the property $i^2 = -1$.
3. For each element z of $\mathbb{C}$ there are two real numbers a, b, such that the complex number z can be expressed as $z = a + ib$, where a is the real part of z and b the imaginary part of z.

We will specify this definition in more detail by defining a 2x2 matrix:

$$a := \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \quad a \in \mathbb{R}$$

$$i := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

We define the complex number $a + bi$ as

$$a + bi := \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + \begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \qquad a, b \in \mathbb{R}$$

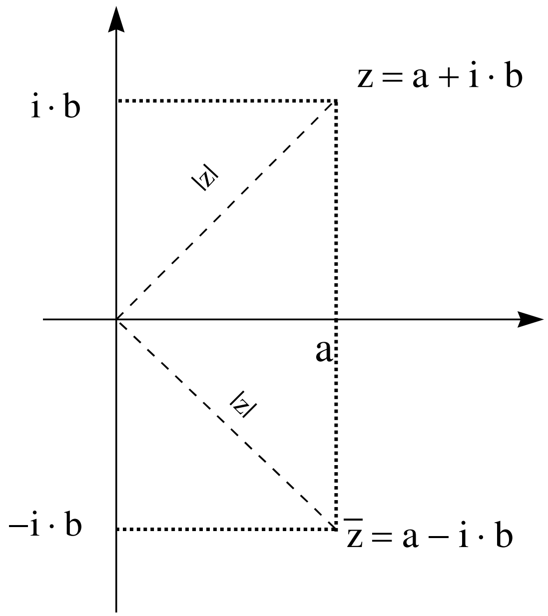
The set of (2x2) matrices illustrates, by addition and multiplication of matrices, a model for complex numbers. The complex number $z = a + ib$ is called purely imaginary, whenever $a = 0$ and $b \neq 0$. It is called purely real, whenever $b = 0$.

The complex number $\overline{z} = a - ib$ is the complex conjugate of $z = a + ib$.

Example:

The equation $x^2 + c = 0$, where $c > 0$ can be solved with the purely imaginary number $z_1 = i\sqrt{c}$ and $z_2 = -i\sqrt{c}$, as $z_1^2 = z_2^2 = -c$. The numbers z_1 and z_2 are said to be complex conjugate.

Visualization of the complex numbers in an Argand diagram:



The points on the horizontal axis correspond to the real numbers. The points on the vertical axis correspond to the purely imaginary numbers. Each point in the plane matches exactly one complex number.

The real number $|z| = \sqrt{a^2 + b^2}$ is called the **absolute value** of $z = a + ib$.

$|z|$ is the distance to the origin.

As it is obvious from the formula this absolute value is identical to the absolute value of real numbers.

Important rules from calculus:

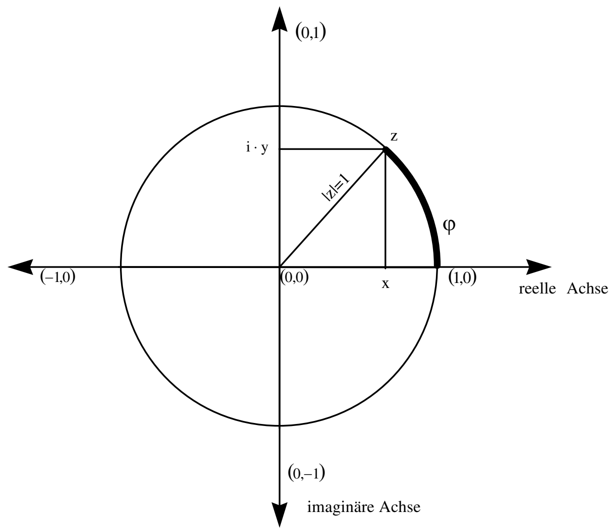
$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

$$(a + ib) - (c + id) = (a - c) + i(b - d)$$

$$(a + ib) \cdot (c + id) = ac - bd + i(ad + bc)$$

Trigonometric representation of complex numbers

A complex number $z = a + ib$ of the absolute value 1 satisfies $x^2 + y^2 = 1$. It is referred to as z being an element of the unit circle in the Argand diagram.



The circumference of the unit circle is 2π. The length of the arc from $(1, 0)$ to $(0, 1), (-1, 0), (0, -1)$ equals $\frac{\pi}{2}, \pi, \frac{3\pi}{2}$.

ϱ is the length of the circular arc from $(1, 0)$ to z

$$\cos(\varphi) := x$$

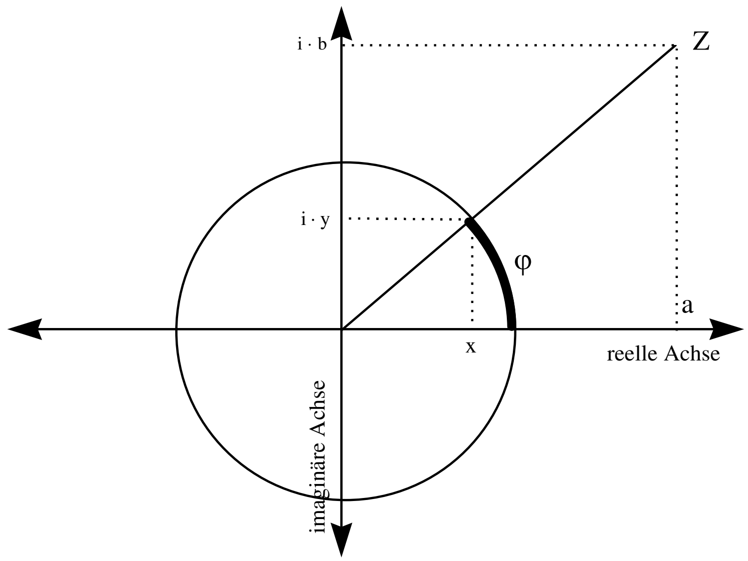
$$\sin(\varphi) := y \quad \text{if } y \neq 0$$

$$\tan(\varphi) := \tfrac{y}{x} \quad \text{if } x \neq 0$$

Hence, the complex number z on the unit circle can be expressed as:

$$z = \cos(\varphi) + i \cdot \sin(\varphi)$$

An arbitrary complex number $z = a + ib$ has the absolute value $R = \sqrt{a^2 + b^2}$. It can be expressed as $z = R(x + iy)$, where $x = \frac{a}{R}$, $y = \frac{b}{R}$ and (x, y) are elements of the unit circle.



Hence, z has the trigonometric form: $z = R \cdot (\cos(\varphi) + i \sin(\varphi))$

$\Rightarrow$ Polar coordinate representation of z

Moivre's theorem: For each complex number $z \neq 0$ and each rational number q it has to hold that $z^q = R^q \left[\cos(q\varphi) + i \sin(q\varphi)\right]$

Exponential representation of complex numbers

$$e^x = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^5}{5!} + \ldots \quad \text{(Power series expansion)}$$

where $x = i\varphi$ holds due to $i^2 = -1, i^3 = -i, i^4 = 1, i^5 = i$

$$\begin{aligned} e^{i\varphi} &= 1 + i\varphi - \frac{\varphi^2}{2!} - i\frac{\varphi^3}{3!} + \frac{\varphi^4}{4!} + i\frac{\varphi^5}{5!} - \frac{\varphi^6}{6!} - i\frac{\varphi^7}{7!} \ldots \\ &= \left[1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} - \frac{\varphi^6}{6!} + \ldots\right] + i\left[\varphi - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} - \frac{\varphi^7}{7!} + \ldots\right] \\ &= \cos(\varphi) + i\sin(\varphi) \end{aligned}$$

The representation of a complex number $z = a + ib$ by means of $z = Re^{i\varphi}$ using $R = |z|$, $\tan(\varphi) = \frac{b}{a}$ is called the exponential form.

II.2 (Stochastic) Difference Equations

[Hamilton (1994), Chapter 1]

First order difference equation

Dynamic properties of

$$y_t = \phi y_{t-1} + w_t \tag{1}$$

w_t can be a random variable. Then: First order stochastic difference equation

Example:

Equation describing the demand for money [Goldfeld (1973)] for the USA m_j (log real demand for money) as a function of log aggregate income (real) I_t, the logarithmic interest rate on deposits r_{Gt} and the interest rate on bonds r_{Ct}:

$$m_t = 0.27 + 0.72m_{t-1} + 0.19I_t - 0.045r_{Gt} - 0.019r_{Ct} \tag{2}$$

Hence, this is just a special case of equation (1) with

$$w_t = 0.27 + 0.19 I_t - 0.045 r_{Gt} - 0.019 r_{Ct}$$

$$y_t = m_t$$

$$\phi = 0.72$$

Aim:

Understanding the dynamic behavior of y if w changes.

Point in time	Equation
0	$y_0 = \phi y_{-1} + w_0$
1	$y_1 = \phi y_0 + w_1$
2	$y_2 = \phi y_1 + w_2$
$\vdots$	$\vdots$
t	$y_t = \phi y_{t-1} + w_t$

$\Rightarrow$ If the starting value y_{-1} for $t = -1$ and w_t for $0, 1, ..., t$ is known, recursive substitution can be used to evaluate the sequence y_t:

$$y_t = \phi^{t+1} y_{-1} + \phi^t w_0 + \phi^{t-1} w_1 + \phi^{t-2} w_2 + \ldots + \phi w_{t-1} + w_t \tag{3}$$

Dynamic behavior

If w_0 changes and $w_1 \ldots w_t$ are not affected of the change, the effect on y_t is:

$$y_t = \frac{\partial y_t}{\partial w_0} = \phi^t$$

Dynamic multiplier = (impulse-response function)

The intensity of the effect of the dynamic multiplier depends on the time span $0 - t$ and the parameter ϕ.

Let the dynamic simulation start in t:

$$y_{t+j} = \phi^{j+1} y_{t-1} + \phi^j w_t + \phi^{j-1} w_{1+1} + \ldots + w_{t+j}$$

Size and sign of ϕ determine the sequence of dynamic multipliers.

The effect of w_t on y_{t+j} is:

$$\frac{\partial y_{t+j}}{\partial w_t} = \phi^j$$

Thus, the dynamic multiplier depends just on j, the time span between w_t and y_{t+j}.

Therefore we have exponential growth/augmention for $\phi > 1$, a geometric decreasing development for $0 < \phi < 1$, oscillating decline for $-1 < \phi < 0$, explosive oscillating behavior for $\phi < -1$.

Higher order difference equations

Generalization of a p-th order difference equation

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + w_t \tag{4}$$

Aim: Explaining the dynamic behavior of equation (4).

To do so: Writing the p-th order difference equation as vector difference equation of order one. We need the following notation:

$$\boldsymbol{\xi}_t \equiv \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix} \quad (p \times 1) - \text{vector}$$

$$\mathbf{F} \equiv \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \quad (p \times p) - \text{matrix}$$

$$\mathbf{v}_t \equiv \begin{pmatrix} w_t \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (p \times 1) - \text{vector}$$

For $p = 1$ (first order difference equation) we have $\mathbf{F} = \phi$ (scalar).

Define a first order vector-difference equation:

$$\boldsymbol{\xi}_t = \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$$

Recursion analogous to the case of a first order difference equation:

For time $t = 0$: $\boldsymbol{\xi}_0 = \mathbf{F}\boldsymbol{\xi}_{-1} + \mathbf{v}_0$
For time $t = 1$: $\boldsymbol{\xi}_1 = \mathbf{F}\boldsymbol{\xi}_0 + \mathbf{v}_1 = \mathbf{F}(\mathbf{F}\boldsymbol{\xi}_{-1} + \mathbf{v}_0) + \mathbf{v}_1 = \mathbf{F}^2\boldsymbol{\xi}_{-1} + \mathbf{F}\mathbf{v}_0 + \mathbf{v}_1$
For time $t = t$:

$$\boldsymbol{\xi}_t = \mathbf{F}^{t+1}\boldsymbol{\xi}_{-1} + \mathbf{F}^t\mathbf{v}_0 + \mathbf{F}^{t-1}v_1 + \ldots + \mathbf{F}\mathbf{v}_{t-1} + \mathbf{v}_t \tag{5}$$

Of special significance for the dynamics: First row of system (5) for time t.

Definition: $f_{11}^{(t)}$ is the $(1,1)$ element of $\mathbf{F}_t$, $f_{12}^{(t)}$ is the $(1,2)$ element of $\mathbf{F}_t$.

For the first row of $\boldsymbol{\xi}_t = \ldots$ we get:

$$\begin{aligned} y_t = \quad & f_{11}^{(t+1)} \quad y_{-1} + f_{12}^{(t+1)} y_{-2} + \ldots + f_{1p}^{(t+1)} y_{-p} + f_{11}^{(t)} w_0 + \\ & f_{11}^{(t-1)} \quad w_1 + \ldots + f_{11}^{(1)} w_{t-1} + w_t \end{aligned}$$

$\Rightarrow \quad y_t$ is a function of p initial values of y and the entire history of w.

Starting the dynamic simulation in t:

$$\boldsymbol{\xi}_{t+j} = \mathbf{F}^{j+1}\boldsymbol{\xi}_{t-1} + \mathbf{F}^{j}\mathbf{v}_t + \mathbf{F}^{j-1}\mathbf{v}_{t+1} + \ldots + \mathbf{F}\mathbf{v}_{t+j-1} + \mathbf{v}_{t+j}$$

for a p-th order difference equation the impulse-response function is

$$\frac{\partial y_{t+j}}{\partial w_t} = f_{11}^{(j)} \tag{6}$$

For $j = 1$ this is given by the $(1, 1)$ element of $\mathbf{F}$, or the parameter ϕ_1!

For each p-th order system the effect of an increase in w_t on y_{t+1} is as follows:

$$\frac{\partial y_{t+1}}{\partial w_t} = \phi_1$$

Expansion of F^2 yields:

$$\frac{\partial y_{t+2}}{\partial w_t} = \phi_1^2 + \phi_2$$

This is the $(1, 1)$ element of $\mathbf{F}^2$.

In order to describe the dynamic behavior of higher order difference equations analytically (e.g. when is the system explosive?) the eigenvalues of the matrix $\mathbf{F}$ are analyzed.

$\Rightarrow$ Matrix algebra [see for example Hamilton Appendix A]

Eigenvalues/characteristic roots of a matrix $\mathbf{F}$ are the solutions for the following equation:

$$|\mathbf{F} - \lambda \mathbf{I}_p| = \mathbf{0}$$

Where $\mathbf{I}_p$ is a p-th order identity matrix. For a system of difference equations of second order this means

$$\left| \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| = \left| \begin{pmatrix} (\phi_1 - \lambda) & \phi_2 \\ 1 & -\lambda \end{pmatrix} \right| = \lambda^2 - \phi_1 \lambda - \phi_2 = 0$$

$\Rightarrow$ characteristic equation

Hence, the two eigenvalues are:

$$\lambda_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}, \; \lambda_2 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

$\Rightarrow$ Eigenvalues can be complex numbers

For difference equations of order p it holds generally that the eigenvalues of $\mathbf{F}$ can be computed as solutions to the characteristic equation:

$$\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \ldots - \phi_{p-1}\lambda - \phi_p = 0$$

Proposition from matrix algebra [see for example Hamilton (1994), Appendix A]

If the eigenvalues of a $(p \times p)$ matrix $\mathbf{F}$ differ, then there is a non-singular matrix $\mathbf{T}$, such that

$$\mathbf{F} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^{-1}$$

where $\mathbf{\Lambda}$ is a $(p \times p)$ matrix containing the eigenvalues of $\mathbf{F}$, which are arranged in the following fashion

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & \lambda_p \end{pmatrix}$$

Hence, we can write:

$$\mathbf{F}^2 = \mathbf{T\Lambda T}^{-1} \cdot \mathbf{T\Lambda T}^{-1} = \mathbf{T\Lambda}^2\mathbf{T}^{-1}$$

Due to the diagonal structure of $\mathbf{\Lambda}$ it holds that

$$\mathbf{\Lambda}^2 = \begin{pmatrix} \lambda_1^2 & 0 & \dots & 0 \\ 0 & \lambda_2^2 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & \lambda_p^2 \end{pmatrix}$$

Generally it must hold that

$$\mathbf{F}^j = \mathbf{T}\mathbf{\Lambda}^j\mathbf{T}^{-1} \tag{7}$$

The diagonal structure of $\mathbf{\Lambda}^j$ is still kept:

$$\mathbf{\Lambda}^j = \begin{pmatrix} \lambda_1^j & 0 & \dots & 0 \\ 0 & \lambda_2^j & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & \lambda_p^j \end{pmatrix}$$

Defining t_{ij} as the element of the i-th row and j-th column of $\mathbf{T}$ and defining t^{ij} as the element of the i-th row and j-th column of $\mathbf{T}^{-1}$, then by multiplying the matrices one can write the $(1,1)$-th element of $\mathbf{F}^j$ as:

$$f_{11}^{(j)} = [t_{11}t^{11}]\lambda_1^j + [t_{12}t^{21}]\lambda_2^j + \ldots + [t_{1p}t^{p1}]\lambda_p^j = c_1\lambda_1^j + c_2\lambda_2^j + \ldots + c_p\lambda_p^j$$

where $c_i = [t_{1i}t^{i1}]$ (To show this, write equation (7) extensively!)

$(c_1 + c_2 + \ldots + c_p)$ is the $(1,1)$ element of $\mathbf{T}\mathbf{T}^{-1} = \mathbf{I}_p$, such that $c_1 + c_2 + \ldots + c_p = 1$

Substitution into equation (6) yields

$$\frac{\partial y_{t+j}}{\partial w_t} = c_1 \lambda_1^j + c_2 \lambda_2^j + \ldots + c_p \lambda_p^j$$

The impulse-response function of order j is a weighted average of the p eigenvalues raised to the j-th power.

For $p = 1$ the characteristic equation states:

$$\lambda_1 - \phi_1 = 0 \quad \Rightarrow \quad \lambda_1 = \phi_1$$

The dynamic multiplier is then given by

$$\frac{\partial y_{t+j}}{\partial w_t} = c_1 \lambda_1^j = \phi_1^j \quad \text{as } c_1 = 1 \text{ (see above)}$$

If there is at least one eigenvalue of $\mathbf{F}$ with an absolute value > 1 the system is explosive, because:

the eigenvalue with the largest absolute value dominates the dynamic multiplier in an exponential function. For real eigenvalues with an absolute value < 1 the dynamic multiplier converges either geometrically or oscillating against zero.

(Compute the dynamic multiplier of equation $y_t = 0.6y_{t-1} + 0.2y_{t-2} + w_t$)

Complex eigenvalues for $p = 2$:

Eigenvalues of $\mathbf{F}$ are complex, if $\phi_1^2 + 4\phi_2 < 0$. Writing the solutions of the characteristic polynomial as complex numbers

$$\lambda_1 = a + ib, \quad \lambda_2 = a - ib, \text{ where } a = \frac{\phi_1}{2}, \quad b = 0.5\sqrt{-\phi_1^2 - 4\phi_2}.$$

To show the dynamic of the system of difference equations, we use the polar coordinate representation:

$$\lambda_1 = R\,[\cos(\rho) + i\sin(\rho)]$$

where $R = \sqrt{a^2 + b^2}, \quad \cos(\rho) = \frac{a}{R}, \quad \sin(\rho) = \frac{G}{R}$

In exponential representation:

$$\lambda_1 = R[e^{i\rho}]$$

$$\lambda_1^j = R^j[e^{i\rho j}] = R^j\,[\cos(\rho j) + i\sin(\rho j)]$$

The complex conjugate λ_1 can be derived as follows:

$$\lambda_2^j = R^j[e^{-i\rho j}] = R^j\left[\cos(\rho j) - i\sin(\rho j)\right]$$

Substitution yields

$$\begin{aligned}
\frac{\partial y_{t+j}}{\partial w_t} &= c_1\lambda_1^j + c_2\lambda_2^j \\
&= c_1 R^j\left[\cos(\rho j) + i\sin(\rho j)\right] + c_2 R^j\left[\cos(\rho j) - i\sin(\rho j)\right] \\
&= \left[c_1 + c_2\right] R^j\cos(\rho j) + i\left[c_1 - c_2\right] R^j\sin(\rho j)
\end{aligned}$$

It can be shown, that these are also complex conjugates [proof: see Hamilton (1994) p. 15]:

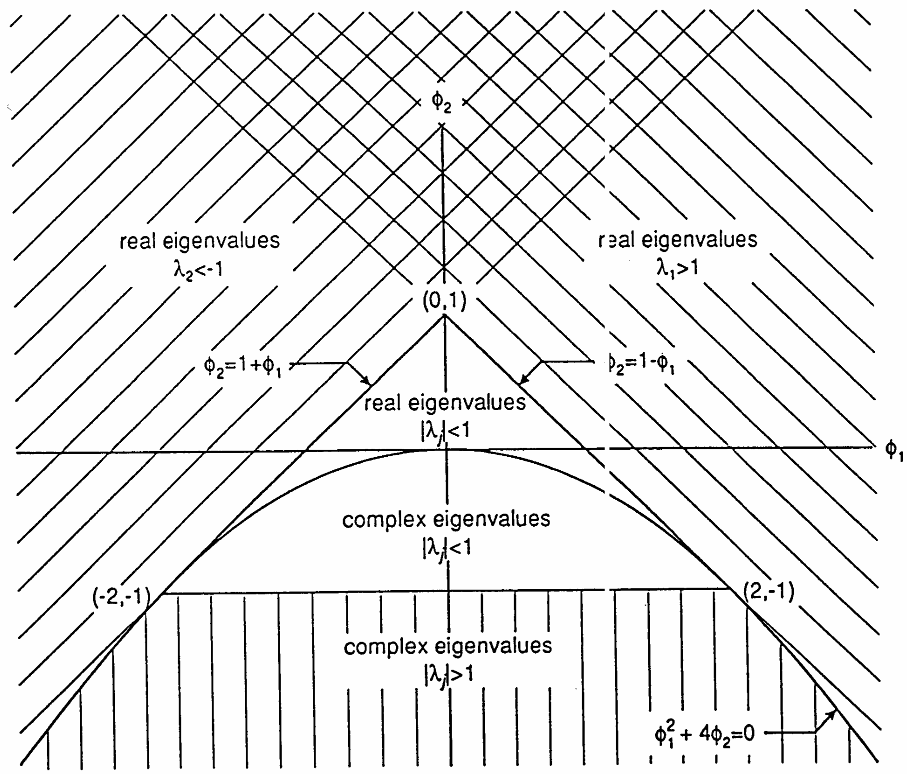
$$c_1 = \alpha + \beta i, \quad c_2 = \alpha - \beta i$$

Substitution yields the real multipliers:

$$c_1\lambda_1^j + c_2\lambda_2^j = 2\alpha R^j \cos(\rho j) - 2\beta R^j \sin(\rho j)$$

$\Rightarrow$ If the eigenvalues are greater than 1 in absolute terms the system explodes at a rate R^j.

For $R = 1$ (the eigenvalues are on the unit circle) the multipliers are periodic sine-cosine-combinations. Only if $R < 1$ (the eigenvalues are inside the unit circle) the amplitude of the multipliers decreases at a rate R^j.



Due to the enormous significance of second order difference equations Sargent's so-called stationarity triangle (1981). A simple derivation [Hamilton (1994) p. 17f.]

II.3 Using Lag Operators

[Hamilton (1994), Chapter 2]

Comment on the notation

The notation of a time series y_t is an abbreviated representation.

The fact, that y_t does not just denote one observation, but a complete time series can be accounted for by using the extensive expression $\{y_t\}_{t=-\infty}^{\infty}$.

Thus: An arithmetic operation $x_t = by_t$ generates not only a new value, but $\{x_t\}_{t=-\infty}^{\infty}$, i.e. a new time series! This holds as well for all the other possible arithmetic operators.

A very important operator, that creates a new time series, is the lag operator.

It is defined as:

$$Lx_t \equiv x_{t-1},$$

where $y = Lx_t$ creates a new time series from $\{x_t\}_{t=-\infty}^{\infty}$. This new time series is denoted by $\{y_t\}_{t=-\infty}^{\infty}$.

It is written:

$$L^2 x_t = L(Lx_t) = L(x_{t-1}) = x_{t-2}$$

For each integer value k:

$$L^k x_t = x_{t-k}$$

Arithmetic operators and lag operators are commutative

$$L(\beta x_t) = \beta L x_t$$

and distributive:

$$L(x_t + w_t) = Lx_t + Lw_t$$

Using the lag operator manipulation of time series is possible. It works analogous to the manipulations done by the common arithmetic operators. Therefore, it can be stated, that x_t is „multiplied“ by L to express that the lag operator operates on x_t.

Example:

$$y_t = (a + bL)Lx_t = (aL + bL^2)x_t = ax_{t-1} + bx_{t-2}$$

An important, later implemented example:

$$(1 - \lambda_1 L)(1 - \lambda_2 L)x_t = \left(1 - (\lambda_1 + \lambda_2)L + \lambda_1\lambda_2 L^2\right) x_t \quad (8)$$

$$= x_t - (\lambda_1 + \lambda_2)x_{t-1} + \lambda_1\lambda_2 x_{t-2} \quad (9)$$

$\Rightarrow$ Lag polynomials can be compared to simple polynomials such as $a \cdot z + b \cdot z^2$ (where z is a real number).

Main difference:

The term $a \cdot z + b \cdot z^2$ adds up to a real number, while $a \cdot L + b \cdot L^2$ operating on a time series $\{_t\}_{t=-\infty}^{\infty}$ produces a new series $\{y_t\}_{t=-\infty}^{\infty}$.

If $x_t = c$ for all t then: $Lx_t = c$.

Practical implementation of the lag operators: Analysis of the dynamics of difference equations

First order difference equation:

$$y_t = \phi y_{t-1} + w_t \quad \Rightarrow \quad y_t = \phi L y_t + w_t \quad \Rightarrow \quad y_t - L y_t = w_t \quad \Rightarrow \quad (1 - \phi L) y_t = w_t \tag{10}$$

In textbooks mainly the inverse representation $y_t = (1 - \phi L)^{-1} w_t$ is printed.

We will explain the relevance of the expression $(1 - \phi L)^{-1}$.

To do so:

"Multiplication" of equation (10) with the lag polynomial $\left(1 + \phi L + \phi^2 L^2 + \phi^3 L^3 \ldots \phi^t L^t\right)$:

$$\left(1 + \phi L + \phi^2 L^2 + \phi^3 L^3 \ldots \phi^t L^t\right) (1 - \phi L)\, y_t = \left(1 + \phi L + \phi^2 L^2 + \phi^3 L^3 \ldots \phi^t L$$

"Expanding" the left hand side (exercise!) yields:

$$(1 - \phi^{t+1} L^{t+1}) y_t = \left(1 + \phi L + \phi^2 L^2 + \phi^3 L^3 \ldots \phi^t L^t\right) w_t$$

Written extensively:

$$y_t = \phi^{t+1} y_{-1} + w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \ldots + \phi^t w_0$$

This is the same result as we got above by recursive substitution!

Property of the operator $\left(1 + \phi L + \phi^2 L^2 + \phi^3 L^3 \ldots \phi^t L^t\right)$, if

- ◇ t gets large,
- ◇ $|\phi < 1|$ is bounded for all t and
- ◇ $|y_t| < y^u$ is bounded for all t,

$$(1 - \phi^{t+1} L^{t+1}) y_t \cong y_t$$

$$\left(1 + \phi L + \phi^2 L^2 + \phi^3 L^3 \ldots \phi^t L^t\right) (1 - \phi L)\, y_t \cong y_t$$

This yields the following result:

$$(1 - \phi L)^{-1} = \lim_{j \to \infty} \left(1 + \phi L + \phi^2 L^2 + \phi^3 L^3 \ldots \phi^j L^j\right)$$

Dynamics of difference equations can also be analyzed by means of the lag operator.

First, dynamics for a second order difference equation

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t$$

$$\left(1 - \phi_1 L - \phi_2 L^2\right) y_t = w_t$$

$\Rightarrow$ Second order lag polynomial factorization of the lag polynomial results in:

$$\left(1 - \phi_1 L - \phi_2 L^2\right) = (1 - \lambda_1 L)(1 - \lambda_2 L) = 1 - [\lambda_1 + \lambda_2]L + [\lambda_1 \lambda_2]L^2 \quad (11)$$

Example:

If $\phi_1 = 0.6$ and $\phi_2 = 0.08 \quad \Rightarrow \quad \lambda_1 = 0.4$ and $\lambda_2 = 0.2$

We will show, that λ_1, λ_2 from equation (11) are identical to the eigenvalues of the matrix $\mathbf{F}$ (see above).

(Remember: Stability ("stationarity") is determined by the eigenvalues of the (2×2) matrix $\mathbf{F}$)

Furthermore, we search for: values λ_1, λ_2 for which equation (11) is fulfilled!

To do so:

Auxiliary construction: We use a number z, that can be substituted for the lag operator L in equation (11):

$$(1 - \phi_1 z - \phi_2 z^2) = (1 - \lambda_1 z)(1 - \lambda_2 z) \qquad (12)$$

The right hand side of equation (12) is 0, if $z = \lambda_1^{-1}$ or $z = \lambda_2^{-1}$.

Thus, it is made clear why we substituted L out with z: $L = \lambda_1^{-1}$ would not have a reasonable interpretation!

z is just to be used as intermediate replacement character for solving for λ_1, λ_2!

$z = \lambda_1^{-1}$ or $z = \lambda_2^{-1}$ have to set the left hand side of equation (12) equal to zero.

$(1 - \phi_1 z - \phi_2 z^2) = 0$ holds for

$$z_1 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}, \; z_2 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$$

z_1, z_2 set the left hand side of equation (12) to 0. We can compute $\lambda_1 = z_1^{-1}, \lambda_2 = z_2^{-1}$.

There is also a more direct way to compute λ_1, λ_2:

To do so:

Division of equation (12) by z^2:

$$(z^{-2} - \phi_1 z^{-1} - \phi_2) = (z^{-1} - \lambda_1)(z^{-1} - \lambda_2)$$

Defining $\lambda = z^{-1}$ yields

$$(\lambda^2 - \phi_1 \lambda - \phi_2) = (\lambda - \lambda_1)(\lambda - \lambda_2) \tag{13}$$

The values of λ, which equalize the right hand side to zero are $\lambda = \lambda_1, \lambda = \lambda_2$. These values have to equalize the left hand side of equation (13) to zero as well:

$$\lambda^2 - \phi_1 \lambda - \phi_2 = 0$$

$$\lambda_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}, \ \lambda_2 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

Hence, it follows: λ_1 and λ_2 are identical to the eigenvalues of the matrix $\mathbf{F}$, which determine the dynamics of the system of difference equations.

These eigenvalues can be calculated by factorizing the lag polynomial $(1 - \phi_1 L - \phi_2 L^2)$ and computing the nulls of the corresponding polynomial $(\lambda^2 - \phi_1 \lambda - \phi_2)$ or $1 - \phi_1 z - \phi_2 z^2$.

Calculate λ_1, λ_2 for a second order difference equation with $\phi_1 = 0.6$ and $\phi_2 = 0.08$.

Be careful: In many textbooks the representations are not clear: Therefore, when is a system of second order difference equations stable?

We have seen:

- ◇ if the eigenvalue λ_1, λ_2 of the (2×2) matrix $\mathbf{F}$ are < 1 in absolute terms (lie inside the unit circle)
- ◇ if the solutions to λ_1 and λ_2 of $(\lambda^2 - \phi_1 \lambda - \phi_2) = 0$ lie inside the unit circle
- ◇ if the solutions to z_1, z_2 where $\lambda_1 = z_1^{-1}, \lambda_2 = z_2^{-1}$ of $(1 - \phi_1 z - \phi_2 z^2) = 0$ lie outside the unit circle.

All three statements are equivalent.

Generalization of the p-th order difference equation:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + w_t$$

$$\left(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p\right) y_t = w_t$$

Factorization of the lag polynomial results in:

$$\left(1-\phi_1 L-\phi_2 L^2-\ldots-\phi_p L^p\right)=(1-\lambda_1 L)(1-\lambda_2 L)\ldots(1-\lambda_p L) \quad (14)$$

As seen above: Substitution of the lag operator by the number z:

$$\left(1-\phi_1 z-\phi_2 z^2-\ldots-\phi_p z^p\right)=(1-\lambda_1 z)(1-\lambda_2 z)\ldots(1-\lambda_p z) \quad (15)$$

The right hand side of equation (15) is zero, whenever $z=\lambda_1^{-1}, z=\lambda_2^{-1}, \ldots, z=\lambda_p^{-1}$. These values also have to equalize the left hand side to zero.

Equalizing the left hand side to zero and multiplying it with z^{-p} and $\lambda \equiv z^{-1}$ yields

$$\left(\lambda^p-\phi_1 \lambda^{p-1} z-\phi_2 \lambda^{p-2}-\ldots-\phi_{p-1}\lambda-\phi_p\right)=0 \quad (16)$$

Equation (16) is identical to the formula we found for the eigenvalues of $\mathbf{F}$ in the case of a p-th order difference equation.

It follows: The nulls of equation (16) are identical to the eigenvalues of the matrix $\mathbf{F}$, which determines the dynamics of the system of difference equations.

These eigenvalues can be computed by first factorizing the lag polynomial $\left(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p\right)$, second derivation of the nulls $\lambda_1, \ldots, \lambda_p$ of the corresponding polynomial $\left(\lambda^p - \phi_1 \lambda^{p-1} z - \phi_2 \lambda^{p-2} - \ldots - \phi_{p-1}\lambda - \phi_p\right)$. Equivalently the nulls $z_1, \ldots, z_p$ of the polynomial $\left(1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p\right)$ (where $z = \lambda_1^{-1}, z = \lambda_2^{-1}, \ldots, z = \lambda_p^{-1}$) can be derived in order to get the eigenvalues.

Three equivalent statements about stability („stationarity") of difference equations of p-th order can be made (and are often confused in textbooks).

A p-th order difference equation is stable, if:

- ◇ the eigenvalues of the $(p \times p)$ matrix $\mathbf{F}$ are within the unit circle.
- ◇ the solutions to $\lambda_1, \ldots, \lambda_p$ of the polynomial $\left(\lambda^p - \phi_1 \lambda^{p-1} z - \phi_2 \lambda^{p-2} - \ldots - \phi_{p-1}\lambda - \phi_p\right)$ are within the unit circle.
- ◇ the solutions $z_1, z_2, \ldots, z_p$ to the polynomial $\left(1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p\right)$ are outside the unit circle.

II.4 Stationarity and Ergodicity

[Hayashi 2.2]

1. Weak/Covariance stationarity

A stochastic process X_t is weakly/covariance stationary if

$$\begin{aligned} \mathbb{E}(X_t) &= \mu \quad \forall\, t \\ \mathrm{Var}(X_t) &= \sigma^2 \quad \forall\, t \\ \mathrm{Cov}(X_t, X_{t-j}) &= \gamma_j \quad \forall\, t \end{aligned}$$

$\Rightarrow$ The mean, variance and autocovariances do not depend on t.
The autocovariances only depend on the distance j,

for example: $\mathrm{Cov}(x_3, x_5) = \mathrm{Cov}(x_{98}, x_{100})$.

2. Strict stationarity

A stochastic process X_t is strictly stationary if its distribution does not depend on t:

$$F_{X_{t_1},\ldots,X_{t_n}}(x_1,\ldots,x_n) = F_{X_{t_1+j},\ldots,X_{t_n+j}}(x_1,\ldots,x_n).$$

So, the joint distribution of two or more random variables in the sequence does not depend on t,

for example: $F_{X_{100},X_{200}}(a,b) = F_{X_{900},X_{1000}}(a,b)$.

- If a sequence is strictly stationary and the variance and covariances are finite, then the sequence is also weakly stationary.
- In the remainder of the course "stationary" means covariance stationary, and therefore we always check for covariance stationarity of a given stochastic process.
- Special case: **Gaussian process**
  As the first two moments are sufficient to identify the normal distribution, for the Gaussian process weak stationarity also implies strict stationarity.

3. Trend stationarity and difference stationarity

- A stochastic process X_t is trend stationary if the process is stationary after subtracting a (usually linear) function of time t, which is called time trend.

- A stochastic process X_t is difference stationary if the process is not stationary, but its first difference, $X_t - X_{t-1}$, is stationary.
  X_t is also called integrated of order 1, $I(1)$-process or a stochastic process with a unit root.

4. Ergodicity and the Ergodic Theorem

- A stochastic process X_t is ergodic if the dependencies between X_t and X_{t-j} get weaker and weaker over time.

- We consider two different definitions:
  a) Hayashi and b) Hamilton.

a) Ergodicity following Hayashi:

A <u>stationary</u> process is ergodic if for any two bounded functions $f : \mathbb{R}^k \rightarrow \mathbb{R}$ and $g : \mathbb{R}^l \rightarrow \mathbb{R}$

$$\lim_{n\to\infty} \quad \mathsf{E}\left[f(z_i,\ldots,z_{i+k}) \cdot g(z_{i+n},\ldots,z_{i+n+l})\right]$$
$$= \mathsf{E}\left[f(z_i,\ldots,z_{i+k})\right] \cdot \mathsf{E}\left[g(z_{i+n},\ldots,z_{i+n+l})\right]$$

$\Rightarrow$ A stationary process is ergodic if it is asymptotically independent, that is, if any two random variables positioned far apart in the sequence, are almost independently distributed.

$\Rightarrow$ Problem: this definition of ergodicity is difficult to check!

b) Ergodicity following Hamilton:
A stationary **Gaussian** process X_t is ergodic if

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty \qquad \text{"absolute summability"}$$

with

$$\begin{aligned} \gamma_0 &= \text{Var}(X_t) \qquad \text{and} \\ \gamma_j &= \text{Cov}(X_t, X_{t-j}); \quad j = 1, 2, \ldots \end{aligned}$$

$\Rightarrow$ In order to check for ergodicity:

1. Is the process stationary Gaussian? Yes $\rightarrow$ 2.
2. Find the autocovariances γ_j and sum them up.
3. Is the sum finite? Yes: the process is stationary and ergodic!

The Ergodic Theorem:

If X_t is a stationary and ergodic process, then any moment of this process is consistently estimated by the sample moment.

5. The autocorrelation function (ACF)

The jth-order autocorrelation function is defined as:

$$\rho_j := \frac{\gamma_j}{\gamma_0} = \frac{\text{Cov}(X_t, X_{t-j})}{\text{Var}(X_t)}; \quad j = 0, 1, 2 \dots$$

with $-1 \leq \rho_j \leq 1$.

The plot of ρ_j against $j = 0, 1, 2 \dots$ is called the correlogram.

III. ARMA Models and Stationarity Tests

—

III.1 Modeling Univariate Time Series: ARMA Models

[Hamilton: 43-61, 64-71]

[Hayashi: 365 - 386]

A general **ARMA**(p, q) model is defined as the stochastic process $\{Y_t\}$ that evolves as:

$$\begin{aligned} Y_t = \quad c + & \quad \underbrace{\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p}}_{\text{AR (autoregressive)-part}} \\ & + \underbrace{\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}}_{\text{MA (moving average)-part}} + \varepsilon_t \end{aligned}$$

where $\{\varepsilon_t\}$ is **Gaussian White Noise**, that is:

$$\begin{aligned} \mathbb{E}(\varepsilon_t) &= 0 \\ \text{Var}(\varepsilon_t) &= \mathbb{E}(\varepsilon_t^2) = \sigma^2 \qquad \forall\, t \\ \text{Cov}(\varepsilon_t, \varepsilon_{t-j}) &= \mathbb{E}(\varepsilon_t \cdot \varepsilon_{t-j}) = 0 \qquad \forall\, j \neq 0 \\ \text{and} \quad \varepsilon_t &\sim N(0; \sigma^2) \end{aligned}$$

Firstly, we are interested in:

i) Is a given ARMA(p, q) process stationary and ergodic?

ii) How does its joint distribution look like?

iii) How can the parameters $c, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q$ be estimated?

iv) How can we forecast the time series?

Reference:
Hamilton: p.43-61 and 64-71
Hayashi: p.365-386

A. Moving Average Processes

1. **MA(1)-process**
2. MA(q)-process
3. MA(∞)-process

1. MA(1)-process

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

with $\{\varepsilon_t\}$: Gaussian White Noise

Checking for stationarity:

$$\mathbb{E}(Y_t) = \mu + \theta_1 \mathbb{E}(\varepsilon_{t-1}) + \mathbb{E}(\varepsilon_t) = \mu \qquad \forall\, t$$

$$\begin{aligned}
\gamma_0 \quad = \mathsf{Var}(Y_t) \quad &= \mathbb{E}[(Y_t - \mu)^2] = \mathbb{E}[(\theta_1 \varepsilon_{t-1} + \varepsilon_t)^2] \\
&= \mathbb{E}[\theta_1^2 \varepsilon_{t-1}^2 + 2\theta_1 \varepsilon_{t-1} \varepsilon_t + \varepsilon_t^2] \\
&= \theta_1^2 \sigma^2 + 0 + \sigma^2 \\
&= (1 + \theta_1^2)\sigma^2 \qquad \forall\, t
\end{aligned}$$

$$\begin{aligned}
\gamma_1 &= \mathsf{Cov}(Y_t, Y_{t-1}) = \mathbb{E}[(Y_t - \mu)(Y_{t-1} - \mu)] \\
&= \mathbb{E}[(\theta_1 \varepsilon_{t-1} + \varepsilon_t)(\theta_1 \varepsilon_{t-2} + \varepsilon_{t-1})] \\
&= \mathbb{E}[\theta_1^2 \varepsilon_{t-1} \varepsilon_{t-2} + \theta_1 \varepsilon_{t-1}^2 + \theta_1 \varepsilon_t \varepsilon_{t-2} + \varepsilon_t \varepsilon_{t-1}] \\
&= 0 + \theta_1 \sigma^2 + 0 + 0 \\
&= \theta_1 \sigma^2 \qquad \forall\, t
\end{aligned}$$

Higher order covariances are all zero:

$$\gamma_j = \mathsf{Cov}(Y_t, Y_{t-j}) = 0 \qquad \forall\ j > 1$$

$\Rightarrow$ $\{Y_t\}$ is (covariance) stationary! Is it also ergodic?

$$\sum_{j=0}^{\infty} \gamma_j = (1 + \theta_1^2)\sigma^2 + |\theta_1|\sigma^2 < \infty$$

$\Rightarrow$ **The MA(1)-process is stationary and ergodic!**

The autocorrelations for the MA(1)-process are given by:

$$\rho_j = \frac{\gamma_j}{\gamma_0} \quad \text{for} \quad j = 0, 1, 2, \ldots$$

Therefore, $\rho_0 = 1$ (always) and for the MA(1)-process we get:

$$\rho_1 \quad = \quad \frac{\theta_1}{(1 + \theta_1^2)} \quad \text{with}$$

$$\rho_1 \quad > \quad 0 \quad \text{for} \quad \theta_1 > 0 \quad \text{and}$$

$$\rho_1 \quad < \quad 0 \quad \text{for} \quad \theta_1 < 0.$$

As **for** $\mathbf{j} > \mathbf{1}$: $\gamma_j = 0 \quad \Rightarrow \quad \rho_j = 0!$
Hence, the autocorrelations are useful to identify the process!

2. MA(q)-process

$$Y_t = \mu + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

normally with $\theta_0 = 1$.

Checking for stationarity and ergodicity: $\rightarrow$ See Hamilton p. 50

Result:

$$\mathbb{E}(Y_t) = \mu$$

$$
\begin{aligned}
\gamma_0 &= \text{Var}(Y_t) = (\theta_0^2 + \theta_1^2 + \ldots + \theta_q^2)\sigma^2 \quad \forall\, t \\
\gamma_j &= \text{Cov}(Y_t, Y_{t-j}) \\
&= (\theta_j\theta_0 + \theta_{j+1}\theta_1 + \ldots + \theta_q\theta_{q-j})\sigma^2 \quad \text{for } j = 1, \ldots, q \\
\gamma_j &= 0 \quad \text{for } j > q\ !
\end{aligned}
$$

Checking for ergodicity:

$$
\sum_{j=0}^{\infty} |\gamma_j| < \infty \qquad \text{for } q < \infty.
$$

$\Rightarrow$ **The MA(q)-process is stationary and ergodic** (for finite q)!

3. MA(∞)-process

If $q \to \infty$: the complete history of the ε's matters! (often in econometrics)

$$
\begin{aligned}
Y_t &= \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \\
&= \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}
\end{aligned}
$$

Is the MA(∞)-process also stationary and ergodic?

If $\sum_{j=0}^{\infty} |\psi_j| < \infty$ (the coefficients are absolutely summable), then the MA(∞)-process is stationary and ergodic!

Why do we need the condition $\sum_{j=0}^{\infty} |\psi_j| < \infty$? Because then:

$$\begin{aligned}
\mathbb{E}(Y_t) &= \mu + \psi_0 \mathbb{E}(\varepsilon_t) + \psi_1 \mathbb{E}(\varepsilon_{t-1}) + \ldots \\
&= \mu + \underbrace{(\psi_0 + \psi_1 + \ldots)}_{\text{finite}} \underbrace{\mathbb{E}(\varepsilon_t)}_{0} \\
&= \mu
\end{aligned}$$

and $\gamma_0 = \text{Var}(Y_t) = \ldots = (\psi_0^2 + \psi_1^2 + \ldots)\sigma^2.$

As $\sum_{j=0}^{\infty} |\psi_j| < \infty$ implies that $\sum_{j=0}^{\infty} {\psi_j}^2 < \infty$ (square summability),
[proof see Hamilton p.69-70],

γ_0 converges to a finite number if $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Similarly,

$$\gamma_j = \text{Cov}(Y_t, Y_{t-j}) = \ldots = (\psi_j \psi_0 + \psi_{j+1}\psi_1 + \ldots)\sigma^2$$

converges also to a finite number if $\sum_{j=0}^{\infty} |\psi_j| < \infty$.
[proof see Hamilton p.70]
Hence, the MA(∞)-process is stationary if $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

And as $\sum_{j=0}^{\infty} |\psi_j| < \infty$ also implies that $\sum_{j=0}^{\infty} |\gamma_j| < \infty$,

the MA(∞)-process is also ergodic if $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

B. Autoregressive Processes

1. **AR(1)-process**

2. AR(2)-process

3. AR(p)-process

4. Invertibility of AR processes

1. AR(1)-process

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t \tag{17}$$

with $\{\varepsilon_t\}$: Gaussian White Noise.

Remember: A first-order linear difference equation is given by

$$Y_t = Y_{t-1} + w_t.$$

For the AR(1)-process: $w_t = c + \varepsilon_t$.

As ε_t is a stochastic process, the AR(1)-process is a first-order stochastic linear difference equation.

As we already showed, Y_t can be written as:

$$Y_t = \phi^{t+1} Y_{-1} + \phi^t w_o + \ldots + \phi^2 w_{t-2} + \phi w_{t-1} + w_t$$

with the dynamic multiplier ϕ^j.

Hence, the effects of the past innovations ε only die out for $|\phi| < 1$, and under this condition the difference equation is stable!

$\Rightarrow$ **The AR(1)-process is only stationary and ergodic if $|\phi| < 1$!**

The AR(1)-process can be written as:

$$Y_t = (c + \varepsilon_t) + \phi(c + \varepsilon_{t-1}) + \phi^2(c + \varepsilon_{t-2}) + \phi^3(c + \varepsilon_{t-3}) + \ldots$$

$$= c\underbrace{(1 + \phi + \phi^2 + \phi^3 + \ldots)}_{\frac{1}{1-\phi} \quad \text{if} \quad |\phi| < 1} + \underbrace{\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \ldots}_{\text{MA}(\infty) - \text{process}}$$

$$Y_t = \mu + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \ldots$$

with $\mathbb{E}(Y_t) = \mu = \dfrac{c}{1-\phi}$.

Checking stationarity and ergodicity for this MA(∞)-process:

$$\sum_{j=0}^{\infty}|\psi_j| = \sum_{j=0}^{\infty}|\phi^j| = \frac{1}{1-|\phi|} < \infty \quad \text{if} \quad |\phi| < 1$$

$\longrightarrow$ stationary and ergodic!

The variance is given by:

$$\begin{aligned}\gamma_0 &= \mathbb{E}[(Y_t-\mu)^2] = \mathbb{E}[(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \ldots)^2] \\ &= (1+\phi^2+\phi^4+\phi^6+\ldots)\sigma^2 \\ &= \frac{1}{1-\phi^2}\,\sigma^2 \qquad (\text{if } |\phi| < 1)\end{aligned}$$

Similarly, we get the autocovariances for $|\phi| < 1$:

$$\begin{aligned}\gamma_j &= \mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)] \\ &= \mathbb{E}[(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \ldots) \\ &\quad (\varepsilon_{t-j} + \phi\varepsilon_{t-j-1} + \phi^2\varepsilon_{t-j-2} + \ldots)]\end{aligned}$$

$$\begin{aligned}\Rightarrow \qquad \gamma_1 &= (\phi + \phi^3 + \phi^5 + \ldots)\,\sigma^2 \\ &= \phi(1 + \phi^2 + \phi^4 + \ldots)\,\sigma^2 \\ &= \frac{\phi}{1-\phi^2}\,\sigma^2\end{aligned}$$

$$
\begin{aligned}
\Rightarrow \qquad \gamma_2 &= (\phi^2 + \phi^4 + \ldots)\,\sigma^2 \\
&= \phi^2(1 + \phi^2 + \phi^4 + \ldots)\,\sigma^2 \\
&= \frac{\phi^2}{1-\phi^2}\,\sigma^2 \\
&\vdots \\
\Rightarrow \qquad \gamma_j &= \phi^j(1 + \phi^2 + \phi^4 + \ldots)\,\sigma^2 = \frac{\phi^j}{1-\phi^2}\,\sigma^2
\end{aligned}
$$

and the autocorrelations:

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi^j$$

⇒ If $|\phi| < 1$, ρ_j decays for $j = 1, 2, \ldots$, but there is no abrupt stop as for a MA(q)-process!

Alternatively, the moments of the AR(1)-process can be calculated by "brute force", that is under the assumption that the AR(1)-process is covariance-stationary:

$$\begin{aligned} Y_t &= c + \phi Y_{t-1} + \varepsilon_t \\ \mathbb{E}(Y_t) &= c + \phi\, \mathbb{E}(Y_{t-1}) + \mathbb{E}(\varepsilon_t). \end{aligned}$$

As $\mathbb{E}(Y_t) = \mathbb{E}(Y_{t-1}) = \mu$ for a covariance-stationary AR(1)-process:

$$\begin{aligned} \mu &= c + \phi\mu + 0 \\ \Rightarrow \quad \mu &= \frac{c}{1-\phi} \end{aligned}$$

Substituting $c = \mu(1-\phi)$ into (17), we get:

$$\begin{aligned} Y_t &= \mu(1-\phi) + \phi Y_{t-1} + \varepsilon_t \\ Y_t - \mu &= \phi(Y_{t-1} - \mu) + \varepsilon_t \end{aligned}$$

Therefore, the variance is:

$$\begin{aligned} \gamma_0 &= \mathbb{E}[(Y_t - \mu)^2] \\ &= \mathbb{E}[(\phi(Y_{t-1} - \mu) + \varepsilon_t)^2] \\ &= \phi^2 \mathbb{E}[(Y_{t-1} - \mu)^2] + 2\phi \mathbb{E}[(Y_{t-1} - \mu)\varepsilon_t] + \mathbb{E}[\varepsilon_t^2] \\ &= \phi^2 \cdot \gamma_0 + 0 + \sigma^2 \\ \Rightarrow \quad \gamma_0 &= \frac{1}{1-\phi^2}\,\sigma^2 \end{aligned}$$

γ_j : $\longrightarrow$ See Hamilton p.53

Note: Using the Lag operator L, the AR(1)-process can be written as:

$$\begin{aligned} Y_t &= \phi L Y_t + \varepsilon_t \qquad \text{(with } c = 0\text{)} \\ (1-\phi L)Y_t &= \varepsilon_t \\ Y_t &= (1-\phi L)^{-1}\varepsilon_t \\ &= (1+\phi L+\phi^2 L^2+\ldots)\varepsilon_t \\ &= \varepsilon_t+\phi\varepsilon_{t-1}+\phi^2\varepsilon_{t-2}+\ldots \end{aligned}$$

which is a MA(∞)-process and therefore called the MA representation of the AR(1) process.

2. AR(2)-process

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{18}$$

which is a second-order stochastic linear difference equation with $w_t = c + \varepsilon_t$ and $\{\varepsilon_t\}$ Gaussian White Noise. This stochastic process can also be written using the lag operator L as:

$$(1 - \phi_1 L - \phi_2 L^2) Y_t = c + \varepsilon_t$$

or in the factorized form:

$$(1 - \lambda_1 L)(1 - \lambda_2 L) Y_t = c + \varepsilon_t.$$

As we saw in II.1, this difference equation is only stable if the eigenvalues λ_1 and λ_2 of the matrix $F = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$, which are the solutions λ_1 and λ_2 of the characteristic polynomial

$$\lambda^2 - \phi_1\lambda - \phi_2 = 0,$$

lie <u>inside</u> the unit circle (are less than 1 in modulus for complex numbers). As we also showed, you can alternatively check, if the solutions z_1 and z_2 of the lag polynomial

$$1 - \phi_1 z - \phi_2 z^2 = 0$$

lie <u>outside</u> the unit circle (are greater than 1 in modulus).
As the AR(2)-process is a second-order stochastic linear difference equation, those same conditions must be fulfilled for the AR(2)-process to be **stationary**!

Then, there also exists an expression for $(1 - \phi_1 L - \phi_2 L^2)^{-1}$ so that the AR(2)-process can also be written as a MA(∞)-process:

$$Y_t = (1 - \phi_1 L - \phi_2 L^2)^{-1} c + (1 - \phi_1 L - \phi_2 L^2)^{-1} \varepsilon_t$$

where

$$\begin{aligned}
(1 - \phi_1 L - \phi_2 L^2)^{-1} &= (1 - \lambda_2 L)^{-1}(1 - \lambda_1 L)^{-1} \\
&= (1 + \lambda_2 L + \lambda_2^2 L^2 + \ldots)(1 + \lambda_1 L + \lambda_1^2 L^2 + \ldots) \\
&= 1 + \psi_1 L + \psi_2 L^2 + \ldots \\
&= \psi(L)
\end{aligned}$$

$$\begin{aligned}
\text{with} \quad \psi_1 &= \lambda_1 + \lambda_2 \\
\psi_2 &= \lambda_1^2 + \lambda_2^2 + \lambda_1 \cdot \lambda_2 \\
&\vdots
\end{aligned}$$

Hence, the MA(∞)-representation of the AR(2)-process is given by:

$$Y_t = \frac{c}{1-\phi_1-\phi_2} + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \ldots$$

with

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

and

$$\psi_j = c_1\lambda_1^j + c_2\lambda_2^j$$

where $c_1 + c_2 = 1$ (for a proof see Hamilton p.12).

Therefore, the MA representation of the AR(2)-process can be written shortly as:

$$Y_t = \mu + \psi(L)\varepsilon_t.$$

Substituting $c = \mu(1 - \phi_1 - \phi_2)$ in (18), we get:

$$\begin{aligned} Y_t &= \mu(1 - \phi_1 - \phi_2) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \\ (Y_t - \mu) &= \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{t-2} - \mu) + \varepsilon_t \end{aligned}$$

Multiplying by $(Y_{t-j} - \mu)$ and taking expectations results in:

$$\begin{aligned} \mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)] &= \phi_1 \mathbb{E}[(Y_{t-1} - \mu)(Y_{t-j} - \mu)] \\ &+ \phi_2 \mathbb{E}[(Y_{t-2} - \mu)(Y_{t-j} - \mu)] \\ &+ \mathbb{E}[\varepsilon_t (Y_{t-j} - \mu)] \end{aligned}$$

$$\Rightarrow \quad \gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} \quad \text{for} \quad j = 1, 2, ... \qquad (19)$$

Thus, the autocovariances follow the same second-order difference equation as the process for Y_t.

By dividing (19) through γ_0 we get the autocorrelations as:

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} \quad \text{for} \quad j = 1, 2, ...$$

As $\rho_0 = 1$ and $\rho_{-1} = \rho_1$ the autocorrelation for $j = 1$ is given by:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 \\ \Rightarrow \quad \rho_1 &= \frac{\phi_1}{1 - \phi_2}. \end{aligned}$$

For $j = 2$:

$$\begin{aligned} \rho_2 &= \phi_1 \rho_1 + \phi_2 \\ &= \frac{\phi_1^2}{1-\phi_2} + \phi_2 \end{aligned}$$

and so on.

Similarly (→ See Hamilton p.57-58), it can be shown that:

$$\gamma_0 = \frac{(1-\phi_2)\sigma^2}{(1-\phi_2)[(1-\phi_2)^2 - \phi_1^2]}.$$

3. AR(p)-process

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \tag{20}$$

which is a pth-order stochastic linear difference equation with $w_t = c + \varepsilon_t$ and $\{\varepsilon_t\}$ Gaussian White Noise.

This stochastic process can also be written using the lag operator L as:

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) Y_t = c + \varepsilon_t.$$

As we have already shown, this difference equation is only stable if the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ of the matrix F,

$$F = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

which are the solutions $\lambda_1, \lambda_2, \ldots, \lambda_p$ of the characteristic polynomial

$$\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \ldots - \phi_{p-1}\lambda - \phi_p = 0,$$

lie <u>inside</u> the unit circle (are less than 1 in modulus for complex numbers).

As we have also shown, you can alternatively check, if the solutions $z_1, z_2, \ldots, z_p$ of the lag polynomial

$$1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0$$

lie <u>outside</u> the unit circle (are greater than 1 in modulus).

As the AR(p)-process is a pth-order stochastic linear difference equation, those same conditions must be fulfilled for the AR(p)-process to be **stationary**!

Then, there exists an expression for $(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1}$ so that the AR(p)-process can also be expressed as a MA(∞)-process:

$$Y_t = (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1} c + (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1} \varepsilon_t$$

where

$$\begin{aligned}
& (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1} \\
= \quad & (1 - \lambda_p L)^{-1} \cdots (1 - \lambda_1 L)^{-1} \\
= \quad & (1 + \lambda_p L + \lambda_p^2 L^2 + \ldots) \cdots (1 + \lambda_1 L + \lambda_1^2 L^2 + \ldots) \\
= \quad & 1 + \psi_1 L + \psi_2 L^2 + \ldots \\
= \quad & \psi(L).
\end{aligned}$$

It can also be shown that

$$\psi_j = c_1\lambda_1^j + c_2\lambda_2^j + \ldots + c_p\lambda_p^j$$

with $\sum_{i=1}^{p} c_i = 1$ (for a proof see Hamilton p.12).

Hence, the MA(∞)-representation of the AR(p)-process is given by:

$$Y_t = \frac{c}{1 - \phi_1 - \ldots - \phi_p} + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \ldots$$

with

$$\mathbb{E}(Y_t) = \mu = \frac{c}{1 - \phi_1 - \ldots - \phi_p}.$$

Therefore, the MA representation of the AR(p)-process can also be written shortly as:

$$Y_t = \mu + \psi(L)\varepsilon_t.$$

As for a stationary AR(p)-process

$$\sum_{j=0}^{\infty} |\psi_j| < \infty,$$

the process is also **ergodic**.

Substituting $c = \mu(1 - \phi_1 - \ldots - \phi_p)$ in (20), we get:

$$
\begin{aligned}
Y_t &= \mu(1 - \phi_1 - \ldots - \phi_p) + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \\
(Y_t - \mu) &= \phi_1(Y_{t-1} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + \varepsilon_t
\end{aligned}
$$

Multiplying by $(Y_{t-j} - \mu)$ and taking expectations results in:

$$
\begin{aligned}
\mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)] &= \phi_1 \mathbb{E}[(Y_{t-1} - \mu)(Y_{t-j} - \mu)] + \ldots \\
&+ \phi_p \mathbb{E}[(Y_{t-p} - \mu)(Y_{t-j} - \mu)] \\
&+ \mathbb{E}[\varepsilon_t (Y_{t-j} - \mu)]
\end{aligned}
$$

$$
\Rightarrow \quad \gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \ldots + \phi_p \gamma_{j-p} \quad \text{for} \quad j = 1, 2, \ldots \tag{21}
$$

Again, the autocovariances follow the same pth-order difference equation as the process for Y_t.

By dividing (21) through γ_0 we get the autocorrelations as:

$$\rho_j = \phi_1 \rho_{j-1} + \ldots + \phi_p \rho_{j-p} \quad \text{for} \quad j = 1, 2, ...$$

Those equations are called the **Yule-Walker equations** and can be solved recursively as we did in the case of the AR(2)-process.

4. Invertibility of AR processes

As all stationary AR(p)-processes have a MA(∞) representation, it can also be shown that a MA(q) process has an AR(∞) representation if the so-called **invertibility conditions** are fulfilled. However, those invertibility conditions resemble the **stationarity conditions** of the AR-process!

C. ARMA Processes

Combining an MA(q) and an AR(p) part, we obtain the general ARMA(p, q) model:

$$\begin{aligned} Y_t &= c + \underbrace{\phi_1 Y_{t-1} + ... + \phi_p Y_{t-p}}_{\text{AR-part}} \\ &\quad + \underbrace{\theta_1 \varepsilon_{t-1} + ... + \theta_p \varepsilon_{t-p}}_{\text{MA-part}} + \varepsilon_t \end{aligned}$$

where $\{\varepsilon_t\}$ is Gaussian White Noise.

As the MA(q) part is always a stationary process, the AR(p) part, that is to say the parameters $\phi_1, ..., \phi_p$, determine if the ARMA(p, q) process is stationary.

Using the lag operator L, the ARMA(p,q)-process can be written as:
$(1-\phi_1 L - ... - \phi_p L^p)Y_t = c_t + (1+\theta_1 L + ... + \theta_q L^q)\varepsilon_t$

If the AR part is stationary, there exists an expression for $(1-\phi_1 L - ... - \phi_p L^p)^{-1}$, so that the ARMA($p,q$)-process has the following MA(∞) representation:

$$
\begin{aligned}
Y_t &= \mu + (1-\phi_1 L - ... - \phi_p L^p)^{-1}(1+\theta_1 L + ... + \theta_q L^q)\varepsilon_t \\
&= \mu + \frac{(1+\theta_1 L + ... + \theta_q L^q)}{(1-\phi_1 L - ... - \phi_p L^p)}\varepsilon_t \\
&= \mu + (1+\psi_1 L + \psi_2 L^2 + ...)\varepsilon_t \\
&= \mu + \psi(L)\varepsilon_t
\end{aligned}
$$

with $\mu = \frac{c}{1-\phi_1-...-\phi_p}$ as for the AR(p)-process.

The stationarity of the AR(p) part also guarantees that:

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

$\Rightarrow$ the process is ergodic!

As we did for the AR(p)-process, we can write the ARMA(p, q)-process in terms of deviations from the mean μ in order to derive the autocovariances:

$$\begin{aligned}(Y_t - \mu) &= \phi_1(Y_{t-1} - \mu) + ... + \phi_p(Y_{t-p} - \mu) \\ &\quad + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q} + \varepsilon_t \\ &\vdots \\ \gamma_j &= \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + ... + \phi_p\gamma_{j-p} \quad \text{for} \quad j > q!\end{aligned}$$

For $j \leq q$, the MA part also effects the autocovariances. Hence, the autocovariances as well as the autocorrelations of the ARMA(p, q)-process have more complicated characteristics than those of an AR(p)- or MA(q)-process!

III.2 Parameter Estimation of ARMA Processes

[Hamilton (1994), Chapter 3, 5]

Aim:

Estimation of the model parameters $\boldsymbol{\theta} = \left(c, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, \sigma^2\right)'$ of an ARMA(p, q) process

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \ldots + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

$\{\varepsilon_t\}_{t \in T}$ White Noise with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t^2) = \sigma^2$ from a time series that contains T observations $(y_1, y_2, \ldots, y_T)$

Maximum likelihood (ML) estimation

$\Rightarrow$ Distributional assumption for ε_t. Typically: $\varepsilon_t \sim i.i.d.N\left(0, \sigma^2\right)$

Computation of the likelihood function, i.e. the "likelihood" to observe a time series $(y_1, y_2, \ldots, y_T)$ given the assumption of a specific parametric stochastic process.

Parameter vector $\boldsymbol{\theta} = \left(c, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, \sigma^2\right)'$ which maximizes the likelihood function: Maximum likelihood estimator.

Maximum Likelihood Estimation of a stationary AR(1) , i.e. ARMA$(1,0)$-Process

$$Y_t = c + \phi y_{t-1} + \varepsilon_t$$

$\{\varepsilon_t\}_{t\in T}$ White Noise with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t^2) = \sigma^2$. Additional Assumption: $\varepsilon_t \sim i.i.d.N\left(0,\sigma^2\right)$.

We search for estimators of the unknown parameters $\boldsymbol{\theta} = (c, \phi, \sigma^2)'$.

$$\mathbb{E}(Y_t) = \mathbb{E}(Y_1) = \frac{c}{1-\phi}$$

$$\mathbb{E}(Y_t) = \mathbb{E}(Y_1 - \mu)^2 = \frac{\sigma^2}{1-\phi^2}$$

where ε_t is normally distributed $\quad\Rightarrow\quad y_1 \sim N\left(\frac{c}{1-\phi}, \frac{\sigma^2}{1-\phi^2}\right)$

Likelihood contribution y_1: $f_{Y_1}(y_1;\boldsymbol{\theta}) = f_{Y_1}(y_1; c, \phi, \sigma^2) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2/(1-\phi^2)}} \exp\left[\frac{-\{y_1 - [c/(1-\phi)]\}}{2\sigma^2/(1-\phi^2)}\right.$

Consider y_1: Density of $(y_2|Y_1 = y_1) \quad N\left((c + \phi y_1), \sigma^2\right)$ i.e.

$$f_{Y_2|Y_1}(y_2|y_1;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\{y_2 - c - \phi y_1)\}^2}{2\sigma^2}\right]$$

Joint density function of the first and second observation

$$f_{Y_2,Y_1}(y_2, y_1;\boldsymbol{\theta}) = f_{Y_2|Y_1}(y_2|y_1;\boldsymbol{\theta}) \cdot f_{Y_1}(y_1;\boldsymbol{\theta})$$

Analogous:

$$
\begin{aligned}
f_{Y_3|Y_2,Y_1}(y_3|y_2,y_1;\boldsymbol{\theta}) &= f_{Y_3|Y_2}(y_3|y_2;\boldsymbol{\theta}) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\{y_3 - c - \phi y_2\}^2}{2\sigma^2}\right]
\end{aligned}
$$

$$
\begin{aligned}
f_{Y_3,Y_2,Y_1}(y_3,y_2,y_1;\boldsymbol{\theta}) &= f_{Y_3|Y_2,Y_1}(y_3|y_2,y_1;\boldsymbol{\theta}) \cdot f_{Y_2,Y_1}(y_2,y_1;\boldsymbol{\theta}) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\{y_3 - c - \phi y_2\}^2}{2\sigma^2}\right] \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\{y_2 - c - \phi y_1\}^2}{2\sigma^2}\right] \cdot \\
&\quad \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2/(1-\phi^2)}} \exp\left[\frac{-\{y_1 - [c/(1-\phi)]\}^2}{2\sigma^2/(1-\phi^2)}\right]
\end{aligned}
$$

Generally:

$$
\begin{aligned}
f_{Y_t|Y_{t-1},Y_{t-2},\ldots,Y_1}(y_t|y_{t-1},y_{t-2},\ldots,y_1;\boldsymbol{\theta}) &= f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\boldsymbol{\theta}) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\{y_t - c - \phi y_{t-1}\}^2}{2\sigma^2}\right]
\end{aligned}
$$

Joint Density of the sample, i.e. likelihood function:

$$f_{Y_T,Y_{T-1},\ldots,Y_1}\left(y_T, y_{T-1}, y_{T-2}, \ldots, y_1; \boldsymbol{\theta}\right) = f_{Y_T|Y_{T-1}}\left(y_t|y_{t-1};\boldsymbol{\theta}\right) \cdot f_{Y_{T-1},Y_{T-2},\ldots,Y_1}\left(y_{T-1}, y_{T-2}, \ldots\right.$$

$$= f\left(y_1;\theta\right) \cdot \prod_{t=2}^{T} f_{Y_t|Y_{t-1}}\left(y_t|y_{t-1};\boldsymbol{\theta}\right)$$

Log likelihood function:

$$\log L = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi^2}\right) - \left[\frac{-\{y_1 - [c/(1-\phi)]\}^2}{2\sigma^2/(1-\phi^2)}\right] - \left[\frac{T-1}{2}\right]\log(2\pi) -$$

$$- \left[\frac{T-1}{2}\right]\log(\sigma^2) - \sum_{t=2}^{T}\left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}\right]$$

The system is maximized by solving for nulls of the first derivatives subject to $\boldsymbol{\theta} = (c, \phi, \sigma^2)'$.

System of equations is non-linear in the parameters $\boldsymbol{\theta} = (c, \phi, \sigma^2)'$ $\Rightarrow$ numerical optimization

Summary: Hamilton (1994), p. 133-142.

Maximum Likelihood Estimation of a stationary AR(p) , i.e. ARMA$(p, 0)$-process

Aim:

Estimating $\boldsymbol{\theta} = (c, \phi_1, \phi_2, \ldots, \phi_p \sigma^2)'$ of an ARMA$(p, 0)$-process is defined with

$$\left[1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \ldots - \phi_p L^p\right] Y_t = c + \varepsilon_t$$

$\{\varepsilon_t\}_{t \in T}$ Gaussian White Noise with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t^2) = \sigma^2$. Additional Assumption: $\varepsilon_t \sim i.i.d. N\left(0, \sigma^2\right)$.

$y^{(p)} = (y_1, y_2, \ldots, y_p)'$: $(p \times 1)$ vector of the first p observations of the time series

$\mu^{(p)}$: $(p \times 1)$ vector of expectations of the first p observations $E\left(y^{(p)}\right)$.

vector consists of p elements: $\mu = \frac{c}{1 - \phi_1 - \phi_2 - \ldots - \phi_p}$

$\sigma^2 \mathbf{V}^{(p)}$: $(p \times p)$ variance covariance matrix of $y^{(p)}$:

$$\begin{pmatrix} \mathbb{E}(Y_1-\mu)^2 & \mathbb{E}(Y_1-\mu)(Y_2-\mu) & \dots & \dots & \mathbb{E}(Y_1-\mu)(Y_p-\mu) \\ \mathbb{E}(Y_1-\mu)(Y_2-\mu) & \mathbb{E}(Y_2-\mu)^2 & & \dots & \vdots \\ \vdots & \vdots & \mathbb{E}(Y_3-\mu)^3 & \vdots & \vdots \\ \vdots & & & \ddots & \\ \mathbb{E}(Y_1-\mu)(Y_p-\mu) & \dots & & \dots & \mathbb{E}(Y_p-\mu)^2 \end{pmatrix}$$

$y^{(p)} \sim N\left(\mu^{(p)}, \sigma^2 \mathbf{V}^{(p)}\right)$. Joint density function of the first p observations (i. e. likelihood contribution):

$$\begin{aligned} f_{Y_1,Y_2,\ldots,Y_p}(y_1, y_2, \ldots, y_p; \boldsymbol{\theta}) = \quad & (2\pi)^{-p/2} \quad \left|\sigma^{-2}(\mathbf{V}^{(p)})^{-1}\right|^{1/2} \cdot \\ & \exp \quad \left[-\frac{1}{2\sigma^2}(y^{(p)} - \mu^{(p)})'(\mathbf{V}^{(p)})^{-1}(y^{(p)} - \mu^{(p)})\right] = \\ & (2\pi)^{-p/2} \quad (\sigma^{-2})^{-p/2}\left|(\mathbf{V}^{(p)})^{-1}\right|^{1/2} \cdot \\ & \exp \quad \left[-\frac{1}{2\sigma^2}(y^{(p)} - \mu^{(p)})'(\mathbf{V}^{(p)})^{-1}(y^{(p)} - \mu^{(p)})\right] \end{aligned}$$

Using $|\alpha \mathbf{A}| = \alpha^n |\mathbf{A}|$.

Consider p preceding observations, then the tth observation is normally distributed with expectation $c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \ldots + \phi_p y_{t-p}$ and variance σ^2.

When we condition only the last p observations are of interest for t. Therefore for $t > p$

$$f_{Y_t|Y_{t-1},Y_{t-2},\ldots,Y_1}\left(y_t|y_{t-1},y_{t-2},\ldots,y_1;\boldsymbol{\theta}\right) = f_{Y_t|Y_{t-1},Y_{t-2},\ldots,Y_{t-p}}\left(y_t|y_{t-1},y_{t-2},\ldots,y_{t-p};\boldsymbol{\theta}\right) =$$

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{\{y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3} - \ldots - \phi_p y_{t-p}\}^2}{2\sigma^2}\right]$$

The joint density (= likelihood) function is:

$$f_{Y_t,Y_{t-1},\ldots,Y_1}(y_t,y_{t-1},y_{t-2},\ldots,y_1;\boldsymbol{\theta}) = f_{Y_p|Y_{p-1},Y_{p-2}\ldots,Y_1}(y_{p-1},y_{p-2},\ldots,y_1;\boldsymbol{\theta}) \cdot \prod_{t=p+1}^{T} f_{Y_t|Y_{t-1},Y_{t-2},\ldots,Y_{t-p}}(y_t|y_{t-1},y_{t-2},\ldots,y_{t-p};\boldsymbol{\theta})$$

Log likelihood:

$$
\begin{aligned}
\log L = \quad & - \quad \frac{p}{2}\log(2\pi) - \frac{p}{2}\log\left(\sigma^2\right) + \frac{1}{2}\left(\mathbf{V}^{(p)}\right)^{-1} - \frac{1}{2\sigma^2}\left(y^{(p)} - \mu^{(p)}\right)'\left(\mathbf{V}^{(p)}\right)^{-1}\left(y^{(p)} - \mu^{(p)}\right) \\
& - \quad \frac{T-p}{2}\cdot\log(2\pi) - \frac{T-p}{2}\cdot\log(\sigma^2) - \sum_{t=p+1}^{T}\left[\frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3} - \ldots - \phi_p y_{1-p})^2}{2\sigma^2}\right] \\
& - \quad \frac{T}{2}\cdot\log(2\pi) - \frac{T}{2}\cdot\log(\sigma^2) + \log(\sigma^2)\frac{1}{2}\log(\mathbf{V}^{(p)})^{-1} - \frac{1}{2\sigma^2}\left(y^{(p)} - \mu^{(p)}\right)'\left(\mathbf{V}^{(p)}\right)^{-1}\left(y^{(p)} - \mu^{(p)}\right) \\
& - \quad \sum_{t=p+1}^{T}\left[\frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3} - \ldots - \phi_p y_{1-p})^2}{2\sigma^2}\right]
\end{aligned}
$$

Setting the first derivatives equal to zero: Resulting system of equations is non-linear in the parameters.

$\Rightarrow$ Numerical optimization.

Avoiding numerical optimization techniques: Conditional likelihood function

$$f_{Y_t,Y_{t-1},\ldots,Y_{p+1}|Y_p,\ldots,Y_1}(y_t, y_{t-1}, y_{t-2}, \ldots, y_{p+1}|y_p, \ldots, y_1; \boldsymbol{\theta}) =$$
$$-\frac{T-p}{2}\cdot\log(2\pi) - \frac{T-p}{2}\cdot\log(\sigma^2) - \sum_{t=p+1}^{T}\left[\frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3} - \ldots - \phi_p y_{1-p})^2}{2\sigma^2}\right]$$

Identical asymptotic distributions for large samples. Conditional log likelihood:

Maximization yields the same result as minimization

$\sum_{t=p+1}^{T}\left[(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3} - \ldots - \phi_p y_{1-p})^2\right]$

$\Rightarrow$ Conditional ML-estimation of an AR(p)-process: Result is identical to Least Squares Estimation. Asymptotic properties of (exact) ML-estimation and OLS-estimation are equivalent.

Conditional Maximum Likelihood estimation of an MA(1), i.e. ARMA$(0,1)$-process

Aim: Estimation of the parameters of an MA(q) process $\boldsymbol{\theta} = \left(\mu, \theta_1, \theta_2, \ldots, \theta_q, \sigma^2\right)'$

Conditional Maximum Likelihood estimation does not result in a simplified estimating equation for the parameters MA(1):

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

$\{\varepsilon_t\}_{t \in T}$ Gaussian White Noise with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t^2) = \sigma^2$.

Additional assumption: $\varepsilon_t \sim i.i.d.N\left(0, \sigma^2\right)$.

Conditioning is more difficult compared to the AR: ε is not directly observable

If ε_{t-1} were known:

$$Y_t|\varepsilon_{t-1} \sim N\left(\mu + \theta\varepsilon_{t-1}, \sigma^2\right)$$

$$f_{Y_t|\varepsilon_{t-1}}(y_t|\varepsilon_{t-1};\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\{y_t-\mu-\theta\varepsilon_{t-1}\}^2}{2\sigma^2}\right]$$

If additionally $\varepsilon_0 = 0$ were known $\Rightarrow$

$$Y_1|\varepsilon_0 \sim N\left(\mu, \sigma^2\right)$$

and: $\varepsilon_1 = y_1 - \mu$

$$\Rightarrow \quad f_{Y_2|Y_1,\varepsilon_0=0}(y_t|y_1,\varepsilon_0 = 0;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\{y_2-\mu-\theta\varepsilon_1\}^2}{2\sigma^2}\right]$$

$\varepsilon_2 = y_2 - \mu - \theta\varepsilon_1$ is also to be derived. $\Rightarrow$ If $\varepsilon_0 = 0$, then the sequence $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T\}$ can be iteratively computed from $\varepsilon_t = y_t - \mu - \theta\varepsilon_{t-1}$ for a given $\boldsymbol{\theta} = (\mu, \theta, \sigma^2)'$.

Conditional density of the tth observation, conditioned on the past observations and $\varepsilon_0 = 0$

$$f_{Y_t|Y_{t-1},Y_{t-2},\ldots,Y_1,\varepsilon_0=0}\left(y_t|y_{t-1}, y_{t-2}, \ldots, y_{t-q}, \varepsilon_0 = 0; \boldsymbol{\theta}\right) =$$

$$= f_{Y_t|\varepsilon_{t-1}}(y_t|\varepsilon_{t-1}; \boldsymbol{\theta}) =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\{\varepsilon_t\}^2}{2\sigma^2}\right]$$

$$\log L = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=1}^{T}\left[\frac{\varepsilon_t^2}{2\sigma^2}\right]$$

For each choice of the parameter vector $\boldsymbol{\theta} = (\mu, \theta, \sigma^2)'$

$\Rightarrow$ Recursion from $\varepsilon_t = y_t - \mu - \theta\varepsilon_{t-1}$

$\Rightarrow$ Sequence of $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T\}$.

Analytical solution for nulls of the conditional log likelihood of an MA(1) process is not available.

Alternative method of computation instead of recursion $\varepsilon_t = y_t - \mu - \theta\varepsilon_{t-1}$:

$$\varepsilon_t = (y_t - \mu) - \theta(y_{t-1} - \mu) + \theta^2(y_{t-2} - \mu) - \ldots + (-1)^{t-1}\theta^{t-1}(y_1 - \mu) + (-1)^t\theta^t\varepsilon_0$$

$|\theta|$ smaller than 1: effects of conditioning get weaker over time.

$\Rightarrow$ Conditional log likelihood is a good approximation of the exact likelihood function

$|\theta| > 1$: cumulation of the effects of conditioning.

Parameter estimator $|\theta| > 1$: results can not be used

$\Rightarrow$ Exact likelihood has to be known.

MA(q) estimation is analogous to conditional ML estimation.

Conditional Maximum Likelihood estimation of an MA(q), i.e. ARMA($0, q$)-process

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

$\{\varepsilon_t\}_{t \in T}$ White Noise with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t^2) = \sigma^2$. Additional assumption: $\varepsilon_t \sim i.i.d.N\left(0, \sigma^2\right)$.

Conditioning on the first q values of the innovation $\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{-q+1} = 0$.

Analogous to MA(1): Recursive construction $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T\}$:

$$\varepsilon_t = y_t - \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

for $t = 1, 2, \ldots, T$

$$\log L = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=1}^{T}\left[\frac{\varepsilon_t^2}{2\sigma^2}\right]$$

Effects of conditioning: Stability of the difference equation $\varepsilon_t = y_t - \mu - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}$?

Is the solution to $\left(1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q\right) = 0$ within the unit circle? (asked differently: are the eigenvalues of F outside?)

$\Rightarrow$ Identification of the exact likelihood function.

Exact Maximum Likelihood estimation of an MA(1), i.e. ARMA$(0,1)$-process

1. Kalman-Filter-Approach [see Hamilton (1994), p. 372 ff.]

2. Triangular factorization of the variance covariance matrix of the MA(1) process

   $(T \times 1)$ vector of realizations of the stochastic process:

   $\mathbf{y} \equiv (y_1, y_2, \ldots, y_t)'$

   $(T \times 1)$ vector of expectations $\boldsymbol{\mu} \equiv (\mu, \mu, \ldots, \mu)'$ and

   $(T \times T)$ variance covariance-matrix of an MA(1): $Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$

$$\boldsymbol{\Omega} = \sigma^2 \cdot \begin{pmatrix} (1+\theta^2) & \theta & 0 & \ldots & 0 \\ \theta & (1+\theta^2) & \theta & 0 & \vdots \\ 0 & \theta & \ldots & & 0 \\ \vdots & 0 & & \ddots & \theta \\ 0 & 0 & \ldots & \theta & (1+\theta^2) \end{pmatrix}$$

Implementing Gaussian White Noise innovations ⇒ joint density (=likelihood): T-variate normal distribution

$$(2\pi)^{-T/2}|\boldsymbol{\Omega}|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'(\boldsymbol{\Omega})^{-1}(\mathbf{y}-\boldsymbol{\mu})\right] \tag{22}$$

Maximization of equation (22)? If there are many observations in the time series: Numerical instabilities when inverting $\boldsymbol{\Omega}$.

Solution: Triangular factorization

$$\mathbf{\Omega} = \mathbf{ADA}'$$

$\mathbf{A}$: $(T \times T)$ matrix, only on and below the main diagonal there are elements unequal to zero. There are only ones on the main diagonal.

$\mathbf{D}$: Diagonal matrix, i.e. only the elements on the main diagonal of the $(T \times T)$ matrix are unequal to zero.

$$\mathbf{\Omega} = \mathbf{ADA}'$$

Writing the matrices $\mathbf{\Omega}$, $\mathbf{A}$, and $\mathbf{D}$ out

$$\mathbf{\Omega} = \sigma^2 \cdot \begin{pmatrix} (1+\theta^2) & \theta & 0 & \dots & 0 \\ \theta & (1+\theta^2) & \theta & 0 & \vdots \\ 0 & \theta & \dots & & 0 \\ \vdots & 0 & & \ddots & \theta \\ 0 & 0 & \dots & \theta & (1+\theta^2) \end{pmatrix}$$

$$\mathbf{A} = \cdot \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{\theta}{(1+\theta^2)} & 1 & 0 & 0 & \vdots \\ 0 & \frac{\theta(1+\theta^2)}{1+\theta^2+\theta^4} & \dots & & 0 \\ \vdots & 0 & & \ddots & 0 \\ 0 & 0 & \dots & \frac{\theta(1+\theta^2+\theta^4+\ldots+\theta^{2(T-2)})}{1+\theta^2+\theta^4+\ldots+\theta^{2(T-1)}} & 1 \end{pmatrix}$$

$$\mathbf{D} = \sigma^2 \cdot \begin{pmatrix} 1+\theta^2 & 0 & 0 & \dots & 0 \\ 0 & \frac{1+\theta^2+\theta^4}{1+\theta^2} & 0 & 0 & \vdots \\ 0 & 0 & \frac{1+\theta^2+\theta^4+\theta^6}{1+\theta^2+\theta^4} & \dots & 0 \\ \vdots & 0 & & \ddots & 0 \\ 0 & 0 & \dots & 0 & \frac{+\theta^2+\theta^4+\dots+\theta^{2T}}{1+\theta^2+\theta^4+\dots+\theta^{2(T-1)}} \end{pmatrix}$$

Derivation: see Hamilton (1994)

Alternative notation of the MA(1) likelihood:

Construction of an auxiliary time series: $\tilde{\mathbf{y}} = \mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu})$

$\mathbf{A}$ has on its main diagonal only ones $\Rightarrow \quad |\mathbf{A}| = 1$

$\Rightarrow \quad |\boldsymbol{\Omega}| = |\mathbf{A}|\,|\mathbf{D}|\,|\mathbf{A}'| = |\mathbf{D}|$

$\Rightarrow$ Likelihood function of the MA(1):

$$
\begin{aligned}
& (2\pi)^{-T/2}|\boldsymbol{\Omega}|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'(\boldsymbol{\Omega})^{-1}(\mathbf{y}-\boldsymbol{\mu})\right] \\
= \quad & (2\pi)^{-T/2}|\mathbf{D}|^{-1/2}\exp\left[-\frac{1}{2}\tilde{\mathbf{y}}'\mathbf{D}^{-1}\tilde{\mathbf{y}}\right]
\end{aligned}
$$

where $\boldsymbol{\Omega}^{-1} = \mathbf{A}^{-1}\mathbf{D}^{-1}\mathbf{A}^{-1}$

Numerical instability when computing the auxiliary time series (due to inversion of the $(T \times T)$ matrix)?

$$\tilde{\mathbf{y}} = \mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \quad \Rightarrow \quad \mathbf{A}\tilde{\mathbf{y}} = (\mathbf{y} - \boldsymbol{\mu})$$

System of equations with T equations.

First line: $\tilde{y}_1 = y_1 - \mu$

tth line: $\tilde{y}_t = y_t - \mu - \frac{1+\theta^2+\theta^4+\ldots+\theta^{2(t-2)}}{1+\theta^2+\theta^4+\ldots+\theta^{2(t-1)}}\tilde{y}_{t-1}$

$\Rightarrow$ Iterative computation of $\tilde{y}_t$, starting with $\tilde{y}_1 = y_1 - \mu$

Numerical instability when computing the inverse of D ($(T \times T)$ matrix)?

$\mathbf{D}$ is a diagonal matrix $\Rightarrow$ $|\mathbf{D}|$ is the product of the terms on the main diagonal $|\mathbf{D}| = \prod_{t=1}^{T} d_{tt}$

Inverse of $\mathbf{D}$: Diagonal matrix with reciprocal values on the main diagonal of $\mathbf{D}$

$$\Rightarrow \quad \tilde{\mathbf{y}}'\mathbf{D}^{-1}\tilde{\mathbf{y}} = \sum_{t=1}^{T} \frac{\tilde{y}_t^2}{d_{tt}}$$

Log likelihood function of an MA(1) process $\log \Big((2\pi)^{-T/2} |\boldsymbol{\Omega}^{-1/2}| \exp \big[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})'(\boldsymbol{\Omega})^{-1}(\mathbf{y} - \boldsymbol{\mu}$

$$\log L = \frac{T}{2}\log(2\pi) - \frac{1}{2}\left(\sum_{t=1}^{T} \log d_{tt}\right) - \frac{1}{2}\left(\sum_{t=1}^{T} \frac{\tilde{y}_t^2}{d_{tt}}\right)$$

Simply evaluate it recursively!

Conditional Maximum Likelihood Estimation of an ARMA(p, q)-process

We search for: Estimator for the parameter vector of an ARMA(p, q) process

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

$\{\varepsilon_t\}_{t \in T}$ White Noise with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t^2) = \sigma^2$. Additional assumption: $\varepsilon_t \sim i.i.d.N\left(0, \sigma^2\right)$.

Likelihood is conditioned on p initial values $y^{(0)} = \{y_0, y_{-1}, \ldots, y_{-p+1}\}$ and q initial innovations $\varepsilon^{(0)} = \{\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{-q+1}\}$.

For given $y^{(0)}, \varepsilon^{(0)}, \boldsymbol{\theta}\left(c, \phi_1, \phi_2, \phi_3, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, \sigma^2\right)'$ $\Rightarrow$ recursive computation of $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T\}$ from $\{y_1, y_2, \ldots, y_T\}$

$$\varepsilon_t = y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \ldots - \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

$$\log L = \log f_{Y_T,Y_{T-1},Y_{T-2},\ldots,Y_1|\varepsilon^{(0)},y^{(0)}}\left(y_T, y_{T-1}, y_{T-2}, \ldots, y_{T-p}|\varepsilon^{(0)}, y^{(0)};\boldsymbol{\theta}\right)$$

$$= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=1}^{T}\left[\frac{\varepsilon_t^2}{2\sigma^2}\right]$$

compare MA(1)

Initial values of the vectors $y^{(0)}$ and $\varepsilon^{(0)}$ e.g. on expectations:

$\varepsilon_s = 0$ for $s = 0, -1, ..., -q+1$ and $y_s = \frac{c}{(1-\phi_1-\phi_2-...-\phi_p)}$ for $s = 0, -1, .., -p+1.$

or observed values $y_1, y_2, \ldots, y_p$ as starting values.

Examination for MA(q) part: Stability of the difference equation

$$\varepsilon_t = y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \ldots - \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

Solutions of

$$\left(1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q\right) = 0$$

outside the unit circle? (possibly eigenvalues of F inside the unit circle?)

If we do not have an exact likelihood function, e.g. Kalman-Filter approach (Hamilton (1994, p.372 ff.)

Wold's decomposition theorem (WDT)

Consider: stationary AR(p) [and ARMA(p,q)] process have MA(∞) representation:

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

$\{\varepsilon_t\}_{t \in T}$ White Noise process and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$

Wold's decomposition theorem: All covariance stationary processes with expectation 0 can be written in the form:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \kappa_t$$

where $\psi_0 = 1$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ and κ_t uncorrelated with ε_{t-j}. κ_t can be expressed by a linear function of preceding values of Y_t: linear deterministic component of Y_t

$\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$: linear stochastic component of Y_t. $\kappa_t = 0 \Rightarrow \quad Y_t$ is a purely stochastic process.

Implications of Wold's decomposition theorem for modeling

Additional assumptions regarding the MA parameter $(\psi_1, \psi_2, \ldots)$ are necessary to make use of the WDT.

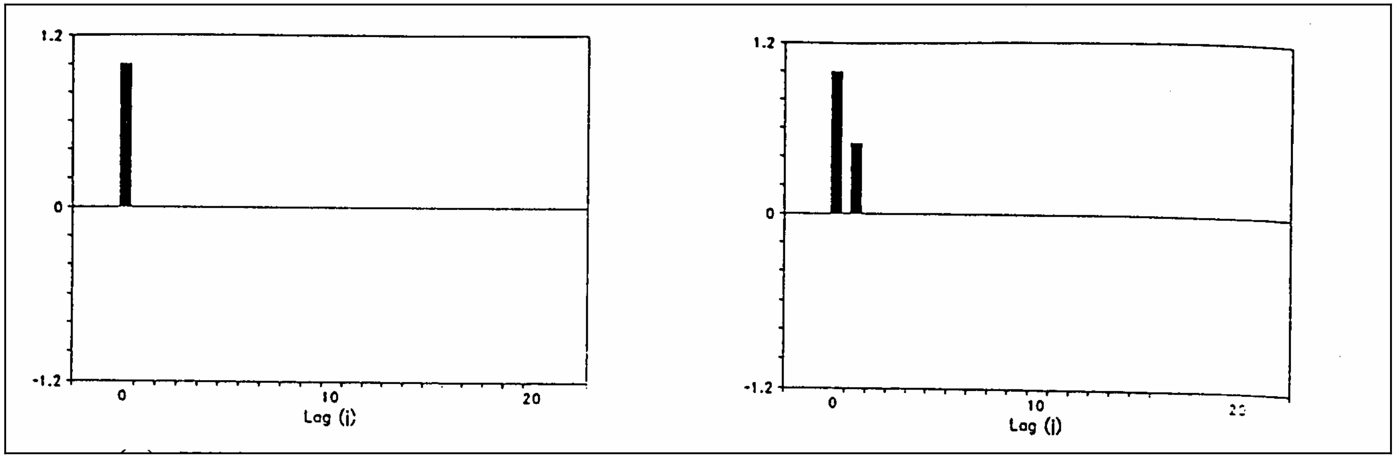
If not there were infinitely many possible parameters.

ARMA(p, q) pose a structure on $\psi(L)$: Infinite lag polynomial as function of the ARMA parameter $\boldsymbol{\theta} = \left(c, \phi_1, \phi_2, , \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, \sigma^2\right)'$

$$\sum_{j=0}^{\infty} \psi_j L^j = \psi(L) = \frac{\theta(L)}{\phi(L)} = \frac{1+\theta_1 L+\theta_2 L^2+\ldots+\theta_q L^q}{1-\phi_1 L-\phi_2 L^2-\ldots-\phi_p L^p}$$

Estimation of $\theta(L)$ and $\phi(L)$ from the sample.

Box-Jenkins modeling philosophy



1. Transform the data, until the assumption of covariance stationarity is met (building differences, logs)
2. First try to model the transformed time series with small values p and q (stage of identification). Compare the empirical ACF with the theoretical ACF of the ARMA(p, q) process.

3. Estimate the parameter $\theta(L)$ and $\phi(L)$ (stage of estimation)
4. Specification tests (possibly iteration for identification)

   Testing for uncorrelated estimated residuals. (Ljung-Box statistic)

Under the null hypothesis, $y_t \sim N\left(\mu, \sigma^2\right)$ the test statistic $Q(k) = \frac{T}{T+2}\sum_{i=1}^{k}(T - i)^{-1} r_i^2$ is asymptotically $\chi^2(k)$.

T: number of observations,

r_i^2: squared autocorrelation of order i,

k: number of accounted autocorrelations

Akaike/Schwartz information criterion

$$AIC^A(p,q) = \ln(\hat{\sigma}^2) + 2(p+q)T^{-1}$$

$$AIC^B(p,q) = -2\ln(L) + 2(p+q)$$

$$SBC^A(p,q) = \ln(\hat{\sigma}^2) + (p+q)T^{-1}\ln T$$

$$SBC^B(p,q) = -2\ln(L) + (p+q)\ln T$$

III.3 Stationarity Tests (Dickey Fuller Test)

[Hamilton (1994), p.502;
Hayashi (2000), Chapter 9.3/9.4]

The work horse to test for non-stationarity: Dickey-Fuller tests
Basics: Unit Root Processes vs. Trend Stationary Processes:

Two types of non-stationarity

$$y_t = \mu + y_{t-1} + u_t \tag{23}$$

$$y_t = \alpha + \beta \cdot t + u_t$$

Equation (23) is a special case of:

$$y_t = \mu + \phi y_{t-1} + u_t$$

There are three cases possible:

$$|\phi| < 1$$
$$|\phi| > 1$$
$$|\phi| = 1$$

$$y_t = \phi y_{t-1} + u_t = \phi u_{t-1} + \phi^2 u_{t-2} + \phi^3 u_{t-3} + \ldots + \phi^t u_0 + \phi^{t+1} y_{-1} + u_t$$

The work horse to test for non-stationarity: Dickey-Fuller tests
Basics: Unit Root Processes vs. Trend Stationary Processes:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \ldots + \phi_p y_{t-p} + u_t$$

Explosive? Stationary? Permanent Effects (Unit root)?

$$y_t = f^1 u_{t-1} + f^2 u_{t-2} + f^3 u_{t-3} + \ldots + f^t u_0 + y_{-1}^{t+1} + u_t$$

Compute p eigenvalues of $\mathbf{F}$, where $\mathbf{F}$:

$$\mathbf{F} \equiv \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \qquad \left| \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| = \lambda^2 - \phi_1 \lambda - \phi_2 = 0$$

absolute value largest root $= 1$: unit root process for $p = 2$

The work horse to test for non-stationarity: Dickey-Fuller tests
Basics: Unit Root Processes vs. Trend stationary processes:

Two types of non-stationarity

$$y_t = y_{t-1} + u_t \quad \text{or} \quad y_t = \mu + y_{t-1} + u_t \tag{24}$$

$$y_t = \alpha + \beta \cdot t + u_t$$

Equation (24) is a special case of:

$$y_t = \mu + \phi y_{t-1} + u_t$$

There are three cases possible with $\mu = 0$:

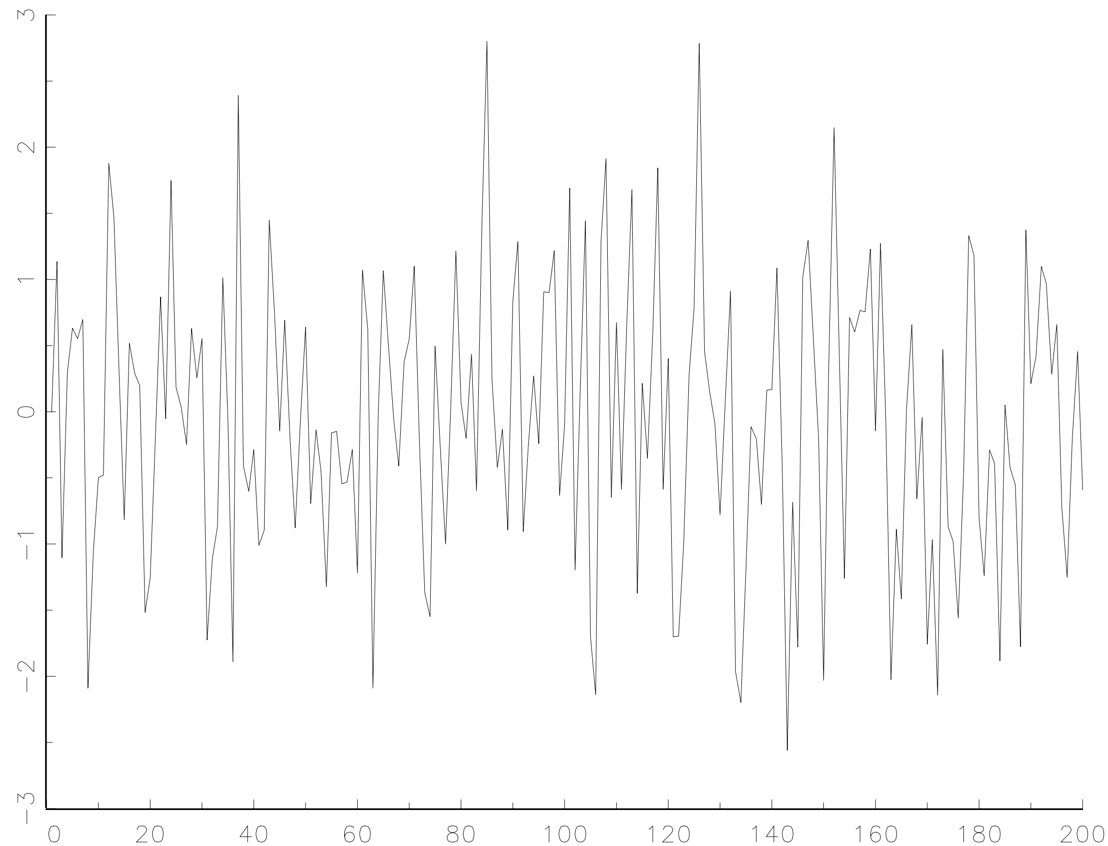
$|\phi| < 1$

$|\phi| > 1$

$|\phi| = 1$

$$y_t = \phi y_{t-1} + u_t = \phi u_{t-1} + \phi^2 u_{t-2} + \phi^3 u_{t-3} + \ldots + \phi^t u_0 + \phi^{t+1} y_{-1} + u_t$$

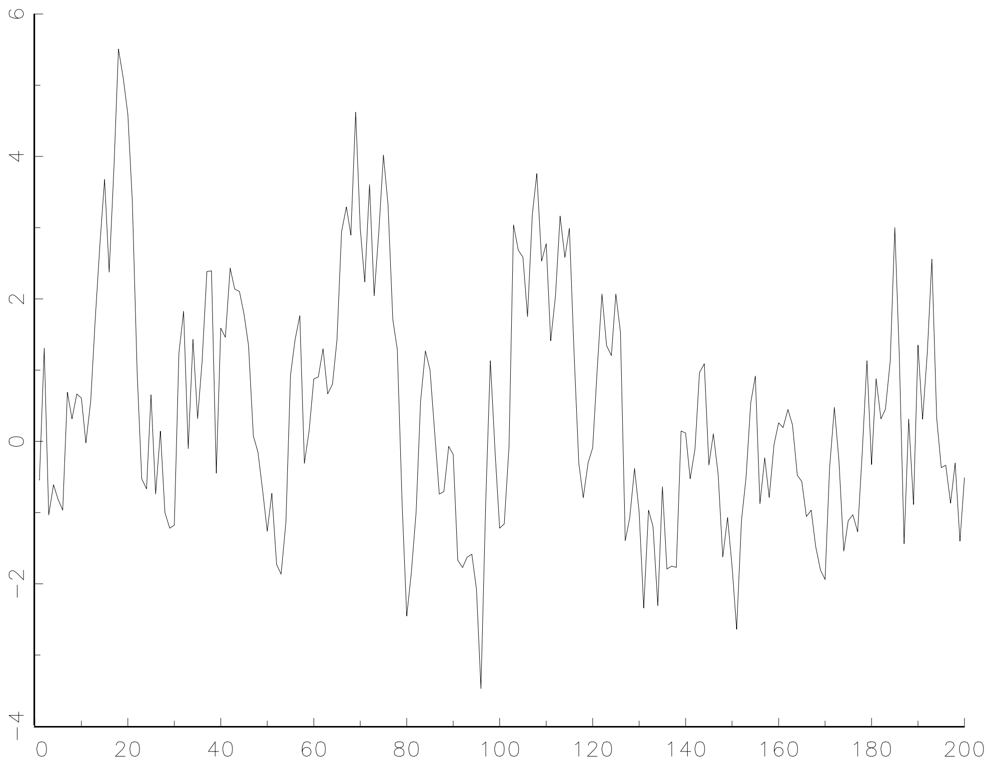
Realization of a White Noise process $y_t = u_t$



Realization of a stationary process (autoregressive process of order one)

$$y_t = 0.8y_{t-1} + u_t$$

$$y_0 = 0$$



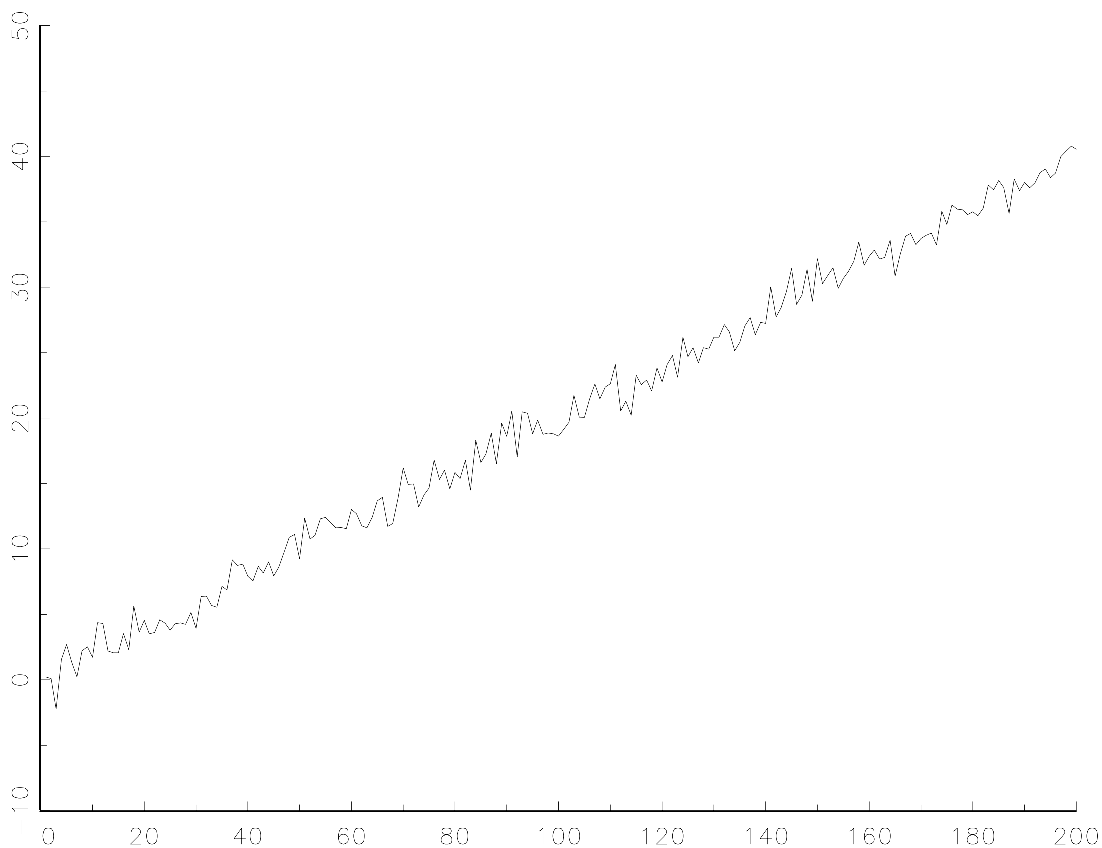
Realization of a random walk without drift

$$y_t = y_{t-1} + u_t$$

$$y_0 = 0$$

Realization of a trend-stationary process

$$y_t = 0.2 \cdot t + u_t$$



Realization of a random walk with drift

$y_t = 0.2 + y_{t-1} + u_t$

$y_0 = 0$

The work horse to test for non-stationarity: Dickey Fuller tests

Basic idea: Test whether $a_1 = 1$ in $y_t = a_1 y_{t-1} + u_t$

Run a regression, back out $\hat{a}_1$, $s.e.(\hat{a}_1)$

Calculate t-statistic: $\tau = \frac{\hat{a}_1 - 1}{s.e.(\hat{a}_1)}$

Distribution of τ under the null: non-standard. Obtained by simulations. Refer to tables (e.g. in Hamilton)

Equivalent (and usually done):

$$y_t - y_{t-1} = \Delta y_t = (a_1 - 1) y_{t-1} + u_t = \gamma \cdot y_{t-1} + u_t$$

$$\Rightarrow \quad \tau = \frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$$

The work horse to test for non-stationarity: Dickey-Fuller test statistics

Related tests. Look at your data! Estimated models:

$$y_t = a_0 + a_1 y_{t-1} + u_t \qquad y_t = a_0 + a_1 y_{t-1} + a_2 t + u_t$$

$$\Delta y_t = a_0 + \gamma \cdot y_{t-1} + u_t \qquad \Delta y_t = a_0 + \gamma \cdot y_{t-1} + a_2 t + u_t$$

Test whether $a_1 = 1$, $\gamma = 0$ respectively.

Run regression, back out $s.e.(\hat{\gamma})$

Calculate t-statistic: $\tau_\mu = \frac{\hat{\gamma}}{s.e.(\hat{\gamma})} \qquad \tau = \frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$

both have under the null hypothesis $\gamma = 0$ non-standard distributions: look up the correct quantile table!!

Critical values (quantiles) for Dickey-Fuller test statistics

STATISTICAL TABLES

Table A **Empirical Cumulative Distribution of τ**

Probability of a Smaller Value Sample Size	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
No Constant or Time ($a_0 = a_2 = 0$)				τ				
25	−2.66	−2.26	−1.95	−1.60	0.92	1.33	1.70	2.16
50	−2.62	−2.25	−1.95	−1.61	0.91	1.31	1.66	2.08
100	−2.60	−2.24	−1.95	−1.61	0.90	1.29	1.64	2.03
250	−2.58	−2.23	−1.95	−1.62	0.89	1.29	1.63	2.01
300	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	2.00
∞	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	2.00
Constant ($a_2 = 0$)				τ_μ				
25	−3.75	−3.33	−3.00	−2.62	−0.37	0.00	0.34	0.72
50	−3.58	−3.22	−2.93	−2.60	−0.40	−0.03	0.29	0.66
100	−3.51	−3.17	−2.89	−2.58	−0.42	−0.05	0.26	0.63
250	−3.46	−3.14	−2.88	−2.57	−0.42	−0.06	0.24	0.62
500	−3.44	−3.13	−2.87	−2.57	−0.43	−0.07	−0.24	0.61
∞	−3.43	−3.12	−2.86	−2.57	−0.44	−0.07	0.23	0.60
Constant + time				τ_τ				
25	−4.38	−3.95	−3.60	−3.24	−1.14	−0.80	−0.50	−0.15
50	−4.15	−3.80	−3.50	−3.18	−1.19	−0.87	−0.58	−0.24
100	−4.04	−3.73	−3.45	−3.15	−1.22	−0.90	−0.62	−0.28
250	−3.99	−3.69	−3.43	−3.13	−1.23	−0.92	−0.64	−0.31
500	−3.98	−3.68	−3.42	−3.13	−1.24	−0.93	−0.65	−0.32
∞	−3.96	−3.66	−3.41	−3.12	−1.25	−0.94	−0.66	−0.33

Source: This table was constructed by David A. Dickey using Monte Carlo methods. Standard errors of the estimates vary, but most are less than 0.20. The table is reproduced from Wayne Fuller. *Introduction to Statistical Time Series.* (New York: John Wiley). 1976.