Advanced Time Series Analysis

Lecture Notes

Prof. Dr. Joachim Grammig

Department of Econometrics, Statistics and Empirical Economics

Winter Term 2007/08

Table of contents

Table of contents (continued)

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!

Stochastic processes (continued):

$\Rightarrow$ 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 $\{X_1, X_2 \ldots\}$

Stochastic processes (continued):

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 the above 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

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

Detailed specification (continued)

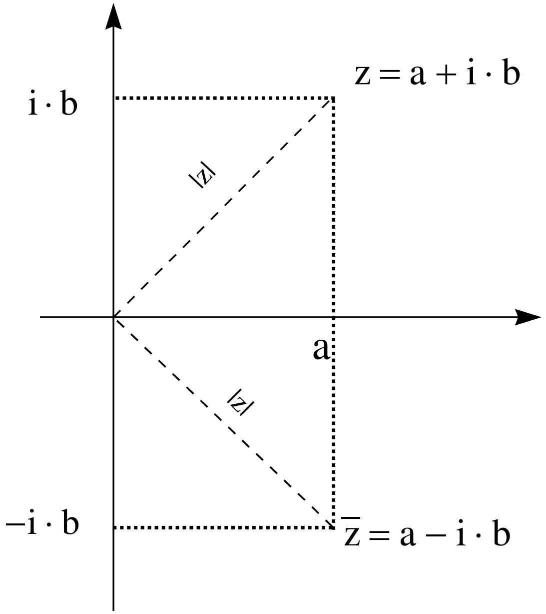
The set of (2x2) matrices illustrates, by addition and multiplication of matrices, a model for complex numbers. The complex number $z = a + ib$ is called purely imaginary, whenever $a = 0$ and $b \neq 0$. It is called purely real, whenever $b = 0$.

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:

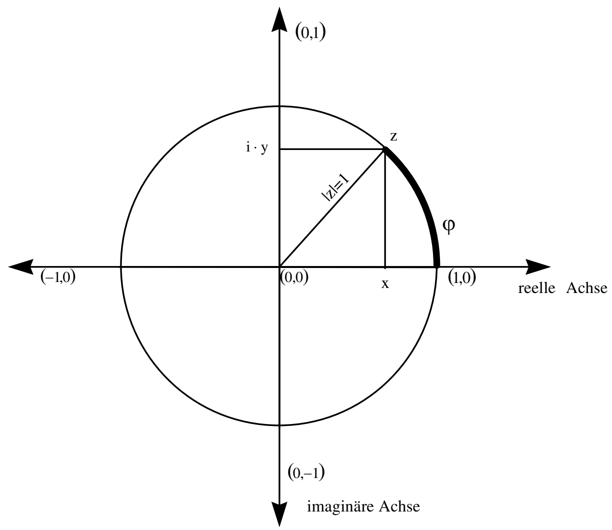
$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

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

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

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 $\frac{\pi}{2}, \pi, \frac{3\pi}{2}$.

Trigonometric representation of complex numbers (continued)

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

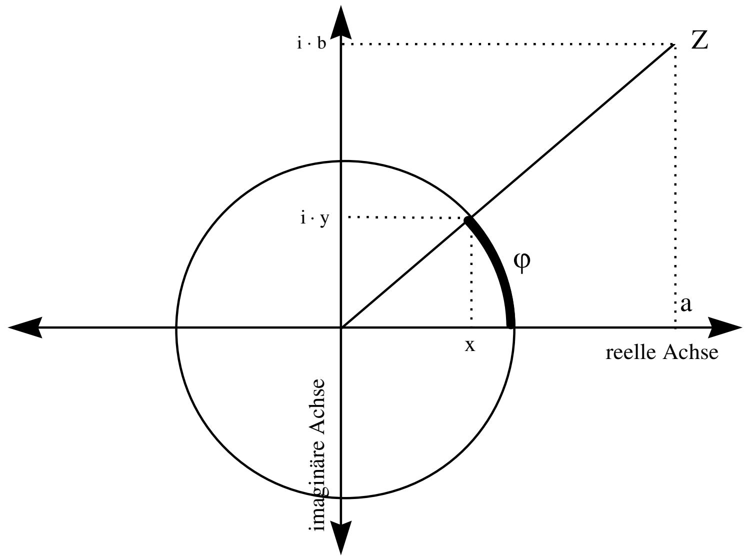
$$\cos(\varphi) := x$$
$$\sin(\varphi) := y \quad \text{if } y \neq 0$$
$$\tan(\varphi) := \frac{y}{x} \quad \text{if } x \neq 0$$

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.

Trigonometric representation of complex numbers (continued)



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 $z^q = R^q [\cos(q\varphi) + i \sin(q\varphi)]$

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, $m_t = 0.27 + 0.72m_{t-1} + 0.19I_t - 0.045r_{Gt} - 0.019r_{Ct}$ **is just a special case of equation** (1) **with**

$$w_t = 0.27 + 0.19I_t - 0.045r_{Gt} - 0.019r_{Ct}, \quad y_t = m_t, \quad \phi = 0.72$$

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

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

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

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.

Dynamic Simulation

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).

Explaining the dynamic behavior of equation

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

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

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

Explaining the dynamic behavior of equation (continued)

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

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

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

$\Rightarrow$ 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:

$$\frac{\partial y_{t+2}}{\partial w_t} = \phi_1^2 + \phi_2 \quad \text{This is the } (1,1) \text{ element of } \mathbf{F}^2.$$

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

$\mathbf{I}_p$ 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 e.g. 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

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

where $\boldsymbol{\Lambda}$ is a $(p \times p)$ matrix containing the eigenvalues of $\mathbf{F}$

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

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

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

Substitution into equation (9) 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:

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

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

Substitution yields the real multipliers

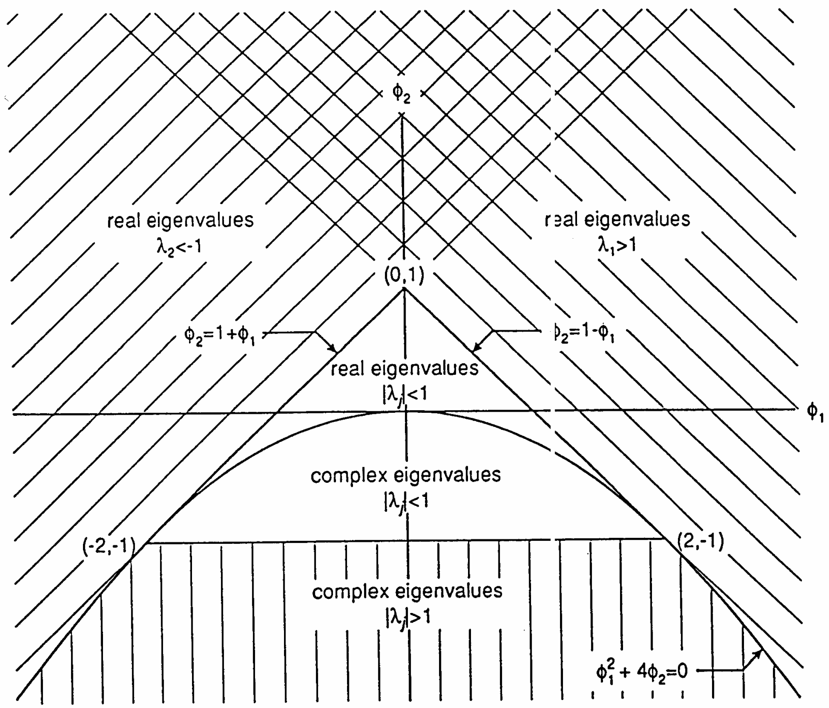
$$c_1\lambda_1^j + c_2\lambda_2^j = 2\alpha R^j \cos(\rho j) - 2\beta R^j \sin(\rho j)$$

$\Rightarrow$ 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:

$$\begin{aligned}(1-\lambda_1 L)(1-\lambda_2 L)x_t &= \left(1-(\lambda_1+\lambda_2)L+\lambda_1\lambda_2 L^2\right)x_t \\ &= x_t-(\lambda_1+\lambda_2)x_{t-1}+\lambda_1\lambda_2 x_{t-2}\end{aligned}$$

$\Rightarrow$ 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:

$$y_t = \phi y_{t-1} + w_t \qquad \Rightarrow \qquad y_t = \phi L y_t + w_t \tag{7}$$

$$\Rightarrow \quad y_t - L y_t = w_t \qquad \Rightarrow \qquad (1 - \phi L) y_t = w_t \tag{8}$$

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 (7) with the lag polynomial $\left(1 + \phi L + \phi^2 L^2 + \phi^3 L^3 \ldots \phi^t L^t\right)$:

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

"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 $\left(1+\phi L+\phi^2L^2+\phi^3L^3\ldots\phi^tL^t\right)$, if

◇ t gets large,

◇ $|\phi<1|$ is bounded for all t and

◇ $|y_t|<y^u$ is bounded for all t,

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

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

This yields the following result:

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

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:

$$\left(1 - \phi_1 L - \phi_2 L^2\right) = (1 - \lambda_1 L)(1 - \lambda_2 L) = 1 - [\lambda_1 + \lambda_2] L + [\lambda_1 \lambda_2] L^2 \quad (9)$$

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 (9) 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 (9) is fulfilled!

Searching for values λ_1, λ_2 for which equation (9) is fulfilled

Auxiliary construction: We use a number z, that can be substituted for the lag operator L in equation (9):

$$(1 - \phi_1 z - \phi_2 z^2) = (1 - \lambda_1 z)(1 - \lambda_2 z) \tag{10}$$

The right hand side of equation (10) 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 (10) equal to zero.

Searching for values λ_1, λ_2 for which equation (9) is fulfilled (cont.)

$(1 - \phi_1 z - \phi_2 z^2) = 0$ 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 (10) 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:

Division of equation (10) by z^2:

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

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

$$(\lambda^2 - \phi_1 \lambda - \phi_2) = (\lambda - \lambda_1)(\lambda - \lambda_2) \tag{11}$$

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 (11) 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 $\phi_2 = 0.08$.

Concluding remarks

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:

$$\left(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p\right) = (1 - \lambda_1 L)(1 - \lambda_2 L) \ldots (1 - \lambda_p L) \tag{12}$$

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

$$\left(1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p\right) = (1 - \lambda_1 z)(1 - \lambda_2 z) \ldots (1 - \lambda_p z) \tag{13}$$

The right hand side of equation (13) 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.

Generalization of the p-th order difference equation (cont.)

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

$$\left(\lambda^p - \phi_1 \lambda^{p-1} z - \phi_2 \lambda^{p-2} - \ldots - \phi_{p-1}\lambda - \phi_p\right) = 0 \qquad (14)$$

Equation (14) 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 (14) 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

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

$\Rightarrow$ The mean, variance and autocovariances do not depend on t.

The autocovariances only depend on the distance j.

Example: $\mathsf{Cov}(x_3, x_5) = \mathsf{Cov}(x_{98}, x_{100})$.

2. Strict stationarity

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

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

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

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

Implications from stationarity

- 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 stationary process is ergodic if for any two bounded functions $f : \mathbb{R}^k \to \mathbb{R}$ and $g : \mathbb{R}^l \to \mathbb{R}$

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

$\Rightarrow$ 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!

5. 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.

6. The autocorrelation function (ACF)

The jth-order autocorrelation function is defined as:

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

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

The plot of ρ_j against $j = 0, 1, 2 \ldots$ 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

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

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

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

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:

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

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

The autocovariance of an MA(1)

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

Higher order covariances are all zero

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

$\Rightarrow$ $\{Y_t\}$ 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** $\mathbf{j > 1}$: $\quad \gamma_j = 0 \quad \Rightarrow \quad \rho_j = 0!$

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:

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

Further results of the MA(q)-process

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

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:

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

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

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

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t \tag{15}$$

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:

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

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

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

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

$\Rightarrow$ 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 $|\phi| < 1$:

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

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

Autocovariances for $|\phi| < 1$ (continued):

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

and the autocorrelations:

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

$\Rightarrow$ 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 (15), 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 : $\rightarrow$ 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

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{16}$$

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 inside 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 outside 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:

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

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 (16), 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:

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

$$
\Rightarrow \quad \gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} \quad \text{for} \quad j = 1, 2, ... \tag{17}
$$

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

By dividing (17) 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 ($\Rightarrow$ 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

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

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, the 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 (18), 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:

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

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

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

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

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

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.

Note: 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: Parameter Estimation of ARMA Processes

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

Parameter Estimation of ARMA Processes

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 $\boldsymbol{\theta} = (c, \phi, \sigma^2)'$.

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

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

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

Likelihood contribution y_1:

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

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:

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

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

$$
\begin{aligned}
f_{Y_T,Y_{T-1},\ldots,Y_1}(y_T, y_{T-1}, y_{T-2}, \ldots, y_1; \boldsymbol{\theta}) = \quad & f_{Y_T|Y_{T-1}}(y_t|y_{t-1}; \boldsymbol{\theta}) \\
\cdot \quad & f_{Y_{T-1},Y_{T-2},\ldots,Y_1}(y_{T-1}, y_{T-2}, \ldots, y_1; \boldsymbol{\theta}) \\
= \quad & f(y_1; \theta) \cdot \prod_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \boldsymbol{\theta})
\end{aligned}
$$

Log likelihood function:

$$
\begin{aligned}
\log L = \quad & - \quad \frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi^2}\right) - \left[\frac{-\{y_1 - [c/(1-\phi)]\}^2}{2\sigma^2/(1-\phi^2)}\right] - \\
& - \quad \left[\frac{T-1}{2}\right]\log(2\pi) - \left[\frac{T-1}{2}\right]\log(\sigma^2) - \sum_{t=2}^{T}\left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}\right]
\end{aligned}
$$

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

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

$y^{(p)} = (y_1, y_2, \ldots, y_p)'$: $(p \times 1)$ 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: $\mu = \frac{c}{1-\phi_1-\phi_2-\ldots-\phi_p}$

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

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

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

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

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

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

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

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 $t > p$

$$
\begin{aligned}
& f_{Y_t|Y_{t-1},Y_{t-2},\ldots,Y_1}(y_t|y_{t-1},y_{t-2},\ldots,y_1;\boldsymbol{\theta}) = \\
= \quad & f_{Y_t|Y_{t-1},Y_{t-2},\ldots,Y_{t-p}}(y_t|y_{t-1},y_{t-2},\ldots,y_{t-p};\boldsymbol{\theta}) = \\
= \quad & \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{\{y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3} - \ldots - \phi_p y_{t-p}\}^2}{2\sigma^2}\right]
\end{aligned}
$$

The joint density (= likelihood) function is:

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

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

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

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 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 $\varepsilon_0 = 0$

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

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

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

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

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

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

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

$|\theta|$ 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. If $|\theta| > 1$: 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 $t = 1, 2, \ldots, T$

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

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

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

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

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

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

$$(2\pi)^{-T/2}|\boldsymbol{\Omega}|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'(\boldsymbol{\Omega})^{-1}(\mathbf{y}-\boldsymbol{\mu})\right] \qquad (20)$$

Maximization of equation (20)? If there are many observations in the time series: Numerical instabilities when inverting $\boldsymbol{\Omega}$.

Numerical instabilities when inverting $\mathbf{\Omega}$

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

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

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

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

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

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

Log likelihood function of an MA(1) process

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

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

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

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

$\{\varepsilon_t\}_{t \in T}$ 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 $\varepsilon^{(0)} = \{\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{-q+1}\}$.

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

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 $\{y_1, y_2, \ldots, y_T\}$

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

$$\begin{aligned} \log L &= \log f_{Y_T, Y_{T-1}, Y_{T-2}, \ldots, Y_1 | \varepsilon^{(0)}, y^{(0)}} \left(y_T, y_{T-1}, y_{T-2}, \ldots, y_{T-p} | \varepsilon^{(0)}, y^{(0)}; \boldsymbol{\theta} \right) \\ &= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \sum_{t=1}^{T} \left[\frac{\varepsilon_t^2}{2\sigma^2} \right] \end{aligned}$$

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

$$\begin{aligned}\log L &= \log f_{Y_T,Y_{T-1},Y_{T-2},\ldots,Y_1|\varepsilon^{(0)},y^{(0)}}\left(y_T,y_{T-1},y_{T-2},\ldots,y_{T-p}|\varepsilon^{(0)},y^{(0)};\boldsymbol{\theta}\right)\\ &= -\frac{T}{2}\log(2\pi)-\frac{T}{2}\log(\sigma^2)-\sum_{t=1}^{T}\left[\frac{\varepsilon_t^2}{2\sigma^2}\right]\end{aligned}$$

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 $y_s = \frac{c}{\left(1-\phi_1-\phi_2-...-\phi_p\right)}$ 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

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

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: $Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ with $\{\varepsilon_t\}_{t \in T}$ 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. $\kappa_t = 0 \Rightarrow Y_t$ 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.

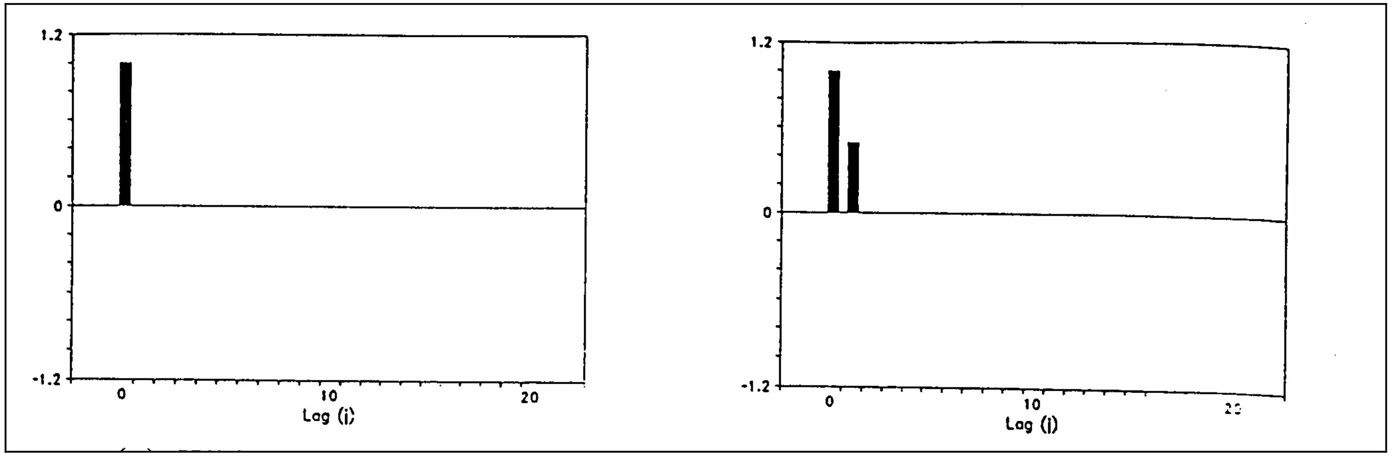
If not there were infinitely many possible parameters.

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.

Box-Jenkins modeling philosophy (continued)

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

$$y_t = \mu + y_{t-1} + u_t \tag{21}$$

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

Equation (21) 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 $\mathbf{F}$:

$$\mathbf{F} \equiv \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

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

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

$$y_t = y_{t-1} + u_t \quad \text{or} \quad y_t = \mu + y_{t-1} + u_t \tag{22}$$

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

Equation (22) is a special case of:

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

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

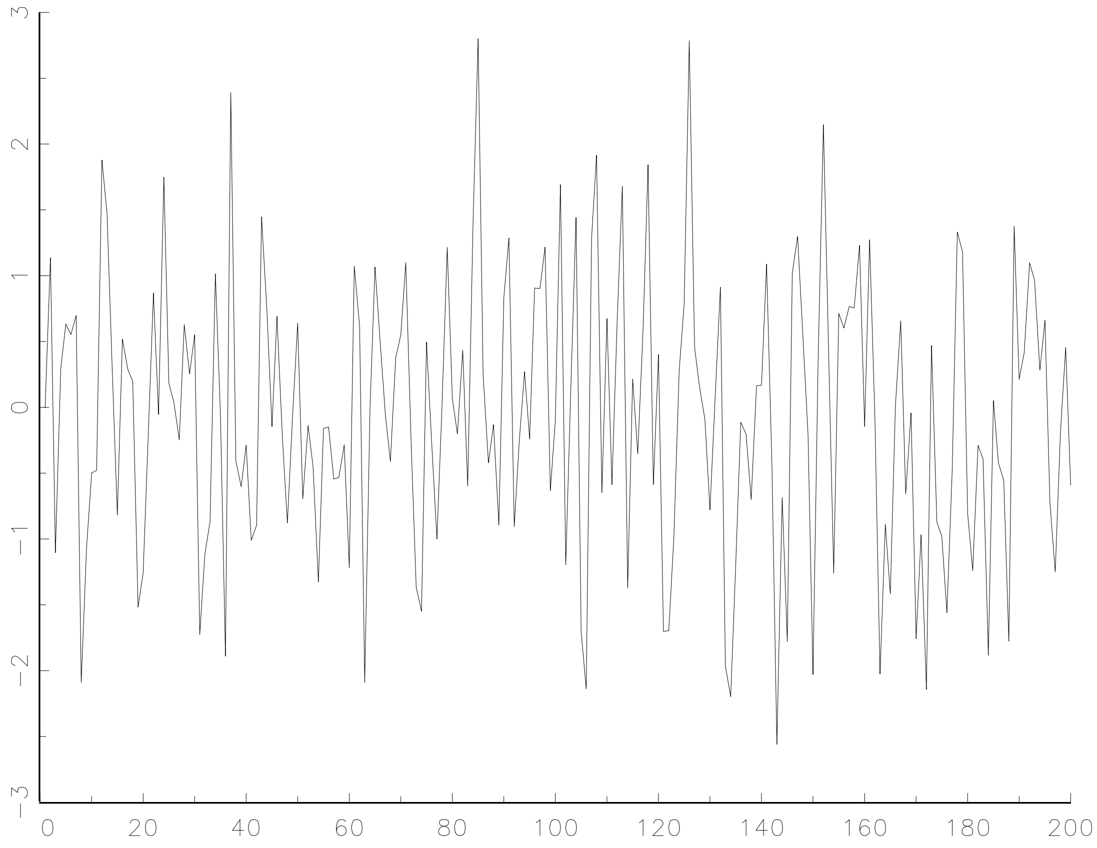
$|\phi| < 1$

$|\phi| > 1$

$|\phi| = 1$

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

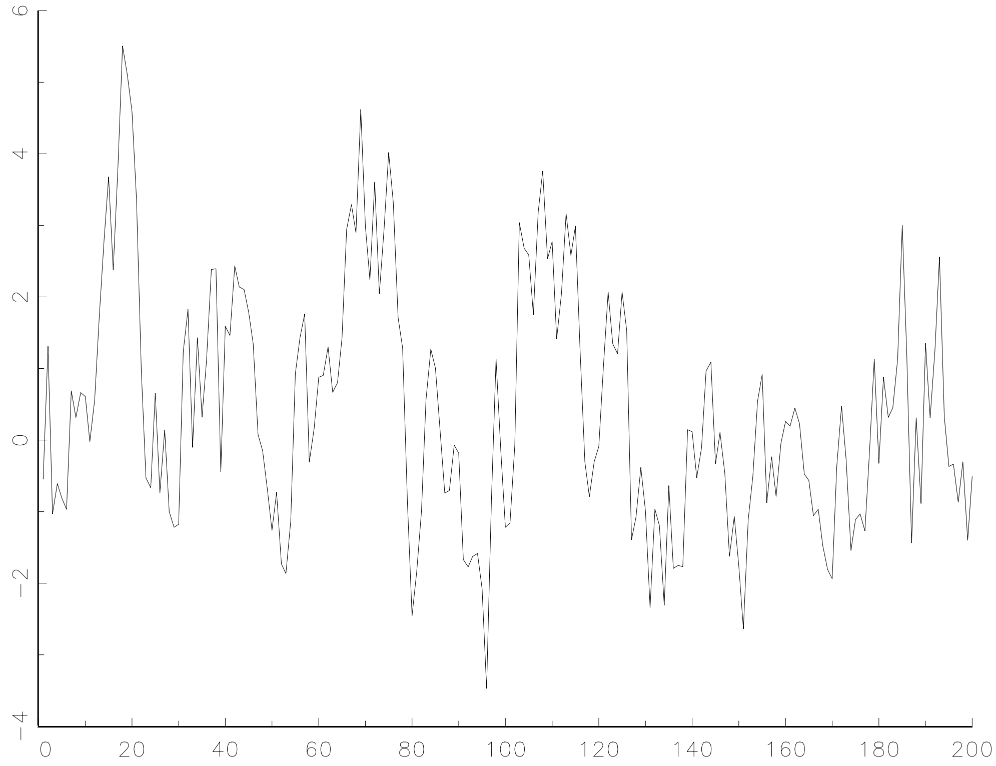
Realization of a White Noise process $y_t = u_t$



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

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

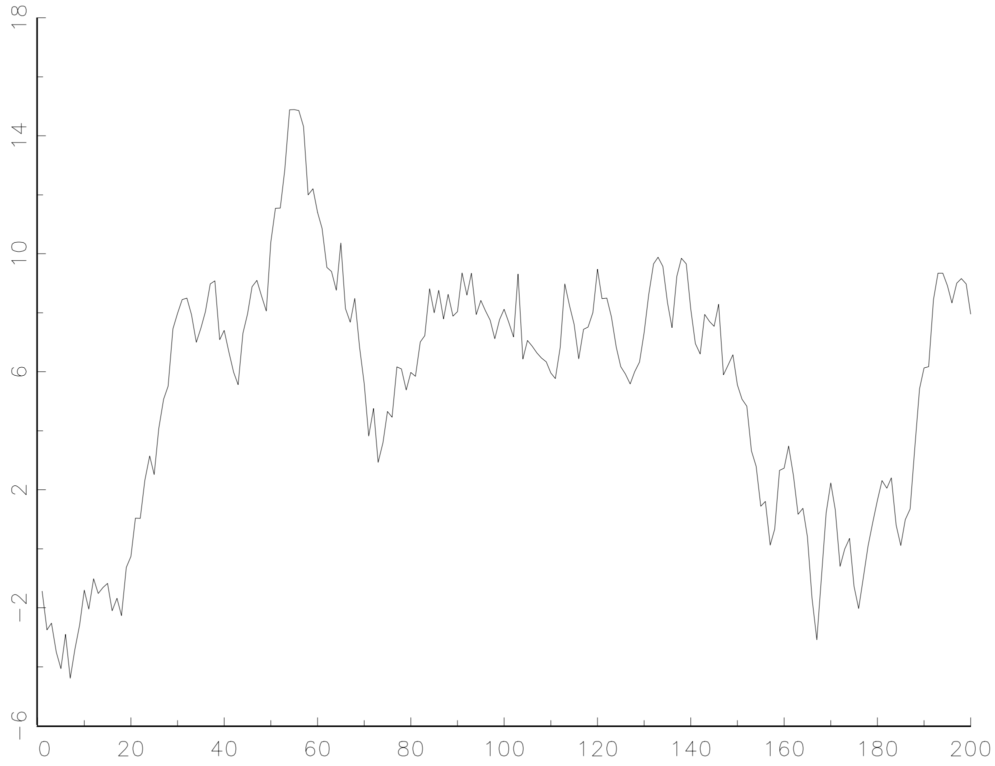
$$y_0 = 0$$



Realization of a random walk without drift

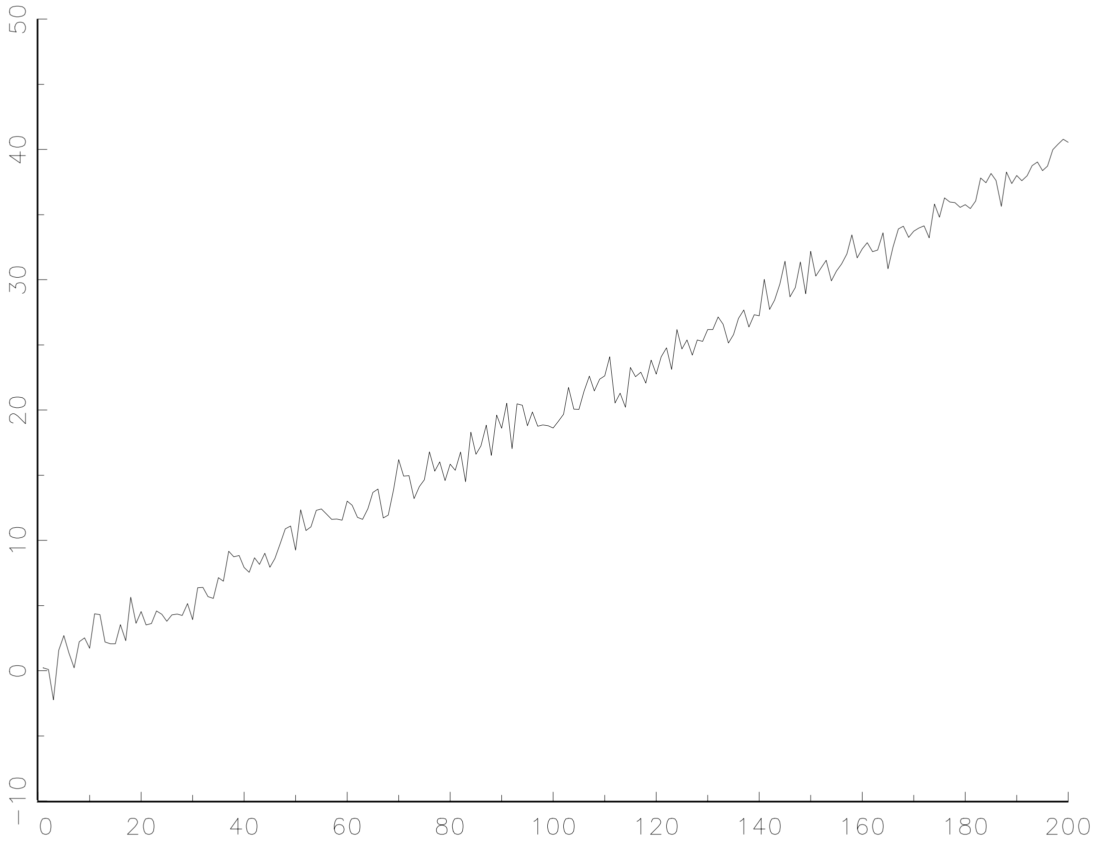
$$y_t = y_{t-1} + u_t$$

$$y_0 = 0$$



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.

IV. Univariate GARCH Models

[Hamilton (1994), Chapter 21]

Conditional vs. unconditional distributions and moments

joint distribution $f_{X_t X_{t+1}}(x_t, x_{t+1})$

marginal distribution $f_{X_t}(x_t)$

conditional distribution of $X_t | X_{t-1} \sim ?$

unconditional moments $\mathbb{E}(X_t), \text{Var}(X_t), \ldots$

conditional moments $\mathbb{E}(X_{t+1} | X_t, X_{t-1}, \ldots), \ldots$

Unconditional moments $\neq$ conditional moments $\Rightarrow$ predictability?

To study the time series properties of asset prices and returns we review some fundamentals of time series analysis

Weak stationarity

$$\left.\begin{aligned} \mathbb{E}(X_t) &= \mu \\ \text{Var}(X_t) &= \sigma^2 \\ \text{Cov}(X_t, X_{t-j}) &= \gamma_j \end{aligned}\right\} \quad \text{unconditional moments are not time dependent}$$

serial dependence $\text{Cov}(X_t, X_{t-j}) = \gamma_j \neq 0$ for $j \neq 0$

$\Rightarrow$ predictability?

Martingale Processes

$\mathbb{E}\left(X_{t+1}|\, I_t\right) = X_t \quad I_t$: information available time t

$\{X_t, X_{t-1}\ldots\} \subset I_t$

$\{X_t\}$ A martingale w.r.t I_t

$\mathbb{E}\left(X_{t+1}|I_t\right)$ **best** forecast of X_{t+1} in terms of
$MSE = \mathbb{E}\left[(X^*_{t+1|t} - X_{t+1})^2\right]$

Using $\mathbb{E}\left(X_{t+1}|I_t\right)$ for $X^*_{t+1|t}$ yields smallest MSE
[Proof: Hamilton (1994): Time Series Analysis, page 72f.]

For a martingale: "best" forecast of tomorrow: observed value of process today

Are asset prices martingales?

$\mathbb{E}\left(Y_{t+1}|I_t\right) = 0$ a martingale difference process

$Y_t = X_t - X_{t-1} \quad \mathbb{E}\left(X_{t+1} - X_t|I_t\right) = 0 \quad (X_t \in I_t)$

Future **changes** are not forecastable using past information (do not improve MSE)

Hypothesis:
Do asset prices follow a martingale process $\Rightarrow$ i.e. price changes unforecastable?

Theory?

Practice?

Marginal utility weighted prices follow martingales (in the absence of dividends)

$$\mathbb{E}_t\left(m_{t+1}\,x_{t+1}\right) = p_t \quad x_{t+1} = p_{t+1} + d_{t+1}$$

$p_t = \mathbb{E}\left(\beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} | I_t\right)$ compare to $\mathbb{E}\left(X_{t+1}|I_t\right) = X_t$

Assume $\beta \approx 1$ and no dividends $d_{t+1} = 0$

$$\mathbb{E}\left(u'(c_{t+1})p_{t+1}|I_t\right) = u'\left(c_t\right)p_t \quad u'\left(c_t\right)p_t \equiv \widetilde{p}_t$$

$$\mathbb{E}\left(\widetilde{p}_{t+1}|I_t\right) = \widetilde{p}_t$$

Marginal utility weighted prices follow a martingale process

In a risk neutral world with no dividends and no time preferences prices follow a martingale. Predictability in the short run?

In a **risk neutral world** $u'(c_t)$ constant:

$$\mathbb{E}\left(p_{t+1}|I_t\right) = p_t$$

Short run, high frequency (e.g. daily): $\beta = 1, c_t$ almost constant, $\Rightarrow$ $u'(c_t) = u'(c_{t+1})$

$$\Rightarrow \mathsf{E}\left(p_{t+1}|I_t\right) = p_t \text{ in short run!}$$

no better forecast of p_{t+1} than p_t in terms of MSE
$\mathbb{E}_t\left(\frac{p_{t+1}}{p_t}\right) = \mathbb{E}_t\left(R_{t+1}\right) = 1$ (coin flips) $\underbrace{\mathbb{E}\left(R_{t+1} - 1\right) = 0}_{\text{net return}}$

In practice...
technical analysis, trend lines, resistance lines, double shoulders...

Predictability in the short run? (1)

An ARMA model for asset returns?

$$R_{t+1} = c + \phi_1 R_t + + \ldots \phi_p R_{t-p+1} + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} + \theta_p \varepsilon_{t-p+1} + \varepsilon_{t+1}$$

where

$$\mathbb{E}(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma^2, Cov(\varepsilon_t \varepsilon_{t-j}) = 0 \quad j > 0$$

$$\mathbb{E}\left(\varepsilon_t | I_{t-1}\right) = 0$$

A useful model?

$$\mathbb{E}\left(R_{t+1} | I_t\right) = c + \phi_1 R_t + \phi_2 R_{t-1} \cdot\cdot + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1}$$

$$\{R_t, R_{t-1}, \varepsilon_t, \varepsilon_{t-1} = \cdot\cdot\} \subset I_t$$

If theory correct $\phi_1 = \phi_2 \ldots = \theta_1 = \theta_2 = 0$

Predictability in the short run? (2)

Some specific martingale processes

Random Walk	$p_t = p_{t-1} + \varepsilon_t$	$\mathbb{E}_t(\varepsilon_t) = 0$
RW type 1	$\{\varepsilon_t\}$ i.i.d	independent, identically distributed
RW type 2	$\{\varepsilon_t\}$ independent, but not necessarily identically distributed	
RW type 3	$\{\varepsilon_t\}$ uncorrelated (less restrictive than independence)	

Tests for random walk hypothesis of asset prices (Chapter 3 in Campbell/Lo/McKinlay) Only weak evidence for short run predictability of asset returns.
(Microstructure effects: bid/ask bounce)

Predictability in the long run?

From

$$\mathbb{E}_t\left(R^i_{t+1}\right) - \quad R^f_t = -\frac{Cov_t\left(m_{t+1}, R^i_{t+1}\right)}{\mathsf{E}_t(m_{t+1})}$$

using $m_t = \beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma_t}$ and log-normal consumption growth $\frac{c_{t+1}}{c_t}$

$$\mathbb{E}_t\left(R^i_{t+1}\right) - R^f_t \approx \gamma_t \sigma_t(\Delta \ln c_{t+1})\sigma_t(R^i_{t+1})\rho_t(m_{t+1}, R^i_{t+1})$$

When setting up a time series model in finance we usually use log returns (1)

It is useful to take 'log returns' (continuously compounded returns)

Use $r_{t+1} = \ln\left(\frac{p_{t+1}}{p_t}\right)$ instead of $R_{t+1} = \frac{p_{t+1}}{p_t}$ (gross return)

$$r_{t+1} \approx \frac{p_{t+1} - p_t}{p_t} = \frac{p_{t+1}}{p_t} - 1 = R_{t+1} - 1 \quad \text{(net return)}$$

e.g.
$p_{t+1} = 105 \qquad p_t = 100$
$R_{t+1} = 1.05$ net return $= 0.05$
$r_t = 0.049$

Continuous compounding between t and $t+1$

When setting up a time series model in finance we usually use log returns (2)

Distributional assumption for $R_{t+1} = \frac{p_{t+1}}{p_t}; [0, \infty)$

Normal distribution? $R_{t+1} - 1; [-1, \infty)$

Assume: $r_t = \ln\left(\frac{p_{t+1}}{p_t}\right) \sim N(\mu, \sigma^2)$

$$\Rightarrow \quad \frac{p_{t+1}}{p_t} = \exp(r_t) \sim \text{lognormal defined on} (0, \infty)$$

When setting up a time series model in finance we usually use log returns (3)

Multi-period returns: $k > 1$

gross returns: $\frac{p_{t+k}}{p_t} = \underbrace{\frac{p_{t+1}}{p_t} \cdot \frac{p_{t+2}}{p_{t+1}} \cdot \frac{p_{t+3}}{p_{t+2}} \cdot \ldots \cdot \frac{p_{t+k}}{p_{t+k-1}}}_{\text{multiplicative}}$

log returns: $\ln\left(\frac{p_{t+k}}{p_t}\right) = \underbrace{\ln\left(\frac{p_{t+1}}{p_t}\right) + \ln\left(\frac{p_{t+2}}{p_{t+1}}\right) \cdots}_{\text{additive}}$

$$= \ln\left(\frac{p_{t+1}}{p_t} \cdot \frac{p_{t+2}}{p_{t+1}} \cdot \frac{p_{t+3}}{p_{t+2}} \cdot \ldots \cdot \frac{p_{t+k}}{p_{t+k-1}}\right)$$

When setting up a time series model in finance we usually use log returns (4)

(Asymptotic) distribution of sum of (normal) random variables known.

Distribution of product of random variables more difficult, especially asymptotic distribution

LLNs and CLTs exist for sums of random variables

Stylized facts of financial return data

- low serial correlation in (log) returns (in line with theory, if prices are martingales)
- significant correlation in squared returns

A simple model to account for these stylized facts

$$r_t = c + u_t$$

$\mathbb{E}(u_t) = 0$, $\mathsf{Var}(u_t) = \mathbb{E}(u_t^2) = \sigma^2$, $\mathbb{E}(u_t u_{t-j}) = 0$ for $j \neq 0$

$\Rightarrow r_t$ and u_t White Noise, $\underbrace{\mathbb{E}(r_t) = c \text{ and } \mathsf{Var}(r_t) = \sigma^2}_{\text{unconditional}}$

$\mathsf{Cov}(r_t, r_{t-j}) = 0 \quad \forall j \neq 0$

The success of Engle‘s ARCH is due to the fact that the model can take into account the fundamental time series properties of asset prices (1)

For the AutoRegressive Conditional Heteroscedasticity (ARCH) model Engle assumes

$$u_t = \sqrt{h_t} \cdot \varepsilon_t$$

1. $\varepsilon_t \sim N(0,1)$ i.i.d.

2. $\mathbb{E}(\varepsilon_t | I_{t-1}) = 0$ exogenous identical shocks (unpredictable)

3. $h_t = f(u_{t-1}^2)$ or $h_t = f(|u_{t-1}|)$ or longer lags of absolute or squared returns. ARCH(1): $h_t = d + a_1 u_{t-1}^2$

The success of Engle's ARCH is due to the fact that the model can take into account the fundamental time series properties of asset prices (2)

For the ARCH(1) specification

$$h_t = d + a_1 u_{t-1}^2$$

$$\begin{aligned}
\mathsf{E}_{t-1}(r_t) &= c + \sqrt{h_t} \cdot \mathsf{E}_t(\varepsilon_t) \\
&= c + \sqrt{d + a_1 u_{t-1}^2} \cdot 0 = c \\
Var_{t-1}(r_t) &= \mathrm{Var}(c) + (d + a_1 u_{t-1}^2) Var_t(\varepsilon_t) \\
&= d + a_1 u_{t-1}^2 \\
&= h_t \quad \text{(conditional variance, } \sqrt{h_t} \text{ conditional "volatility")}
\end{aligned}$$

Remark:

Volatility sometimes defined as **annualized** (log) return standard deviation. With $\sigma_t = \sqrt{h_t}$ the standard deviation of daily log returns we annualize $\sigma_{ann.} = T \cdot \sigma_t$ (T number of trading days)

(ML) estimated coefficient a_1 significantly different from zero (positive)!

$\Rightarrow$ variance of return in $t+1$ predictable given time t information!

A plethora of conditional volatility models have been proposed

- ♢ asymmetric responses of return variance to positive or negative return shock?
- ♢ persistence of shocks (ARCH) only one lag period or longer?

  $\Rightarrow$ GARCH: $h_t = d + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$
- ♢ Long memory, fractionally integrated GARCH
- ♢ shocks ε normally distributed?
  fat tails, skewness (large negative shocks more likely)
- ♢ How to ensure non-negativity of conditional variance h_t?
- ♢ ARCH-in-Mean $r_t = d + \delta h_t + \sqrt{h_t}\varepsilon_t$
- ♢ multivariate extensions $\Rightarrow$ multivariate ARCH: Conditional covariances of asset returns (correlations) predictable.

A successful model: Nelson‘s E-ARCH model (1)

Nelson’s Exponential ARCH
Standard assumptions:

$$
\begin{aligned}
r_t &= \mu + u_t \\
u_t &= \sqrt{h_t} v_t
\end{aligned}
$$

where $E_{t-1}(v_t) = 0$ and $E_{t-1}(v_t^2) = 1$

$\{v_t\}$ iid, i.e. $Var_{t-1}(r_t) = h_t$ r_t a White Noise process as above.

However, log of conditional variance evolves as:

$$\ln h_t = \zeta + \pi\{|v_{t-1}| - \mathbb{E}(|v_{t-1}|) + \aleph v_{t-1}\}$$

A successful model: Nelson‘s E-ARCH model (2)

$$\ln h_t = \zeta + \pi\{|v_{t-1}| - \mathbb{E}(|v_{t-1}|) + \aleph v_{t-1}\}$$

Non-negativity of $h_t = Var_{t-1}(r_t)$ guaranteed.

Asymmetric effects positive and negative return shocks possible:

$\pi > 0 \rightarrow$ deviation of absolute iid shock $|v_{t-1}|$ from expectation $\mathbb{E}(|v_{t-1}|)$ increases volatility forecast (c.p.)

$-1 < \aleph < 0$ positive return shock $v_{t-1} > 0$ increases volatility forecast h_{t+1} less than negative return shock $v_{t-1} < 0$

$\aleph < -1$ positive return shock $v_{t-1} > 0$ c.p. decreases volatility forecast h_{t+1}.

A successful model: Nelson‘s E-ARCH model (3)

$$\ln h_t = \zeta + \pi\{|v_{t-1}| - \mathbb{E}(|v_{t-1}|) + \aleph v_{t-1}\}$$

Economic explanation (heuristic) for $-1 < \aleph < 0$:

Leverage effect

stock prices $\downarrow \Rightarrow$ value of ratio $\frac{\text{value equity}}{\text{corporate dept}} \downarrow \Rightarrow$

risk of holding stocks increases.

Note: Extendable to E-GARCH model (lagged $\ln h_{t-j}$ on right hand side)

Uses of ARCH type models

Forecast return variances and covariances for VAR models

volatility forecast to feed in Black/Scholes formula (practitioners approach)

Estimate and forecast time varying beta

$$\beta_{it} = \frac{Cov_{t-1}(R_t^m, R_t^i)}{Var_{t-1}(R_t^m)} \Rightarrow \text{asset pricing}$$

Modeling evolution of conditional covariance in same fashion: Bivariate ARCH models

Portfolio selection: forecast variance-covariance matrix of assets in portfolio (multivariate ARCH models)

V. Vector Autoregressions (Basics)

[Hamilton (1994), Chapter 11;
Hayashi (2000), Chapter 6.3/6.4;
Enders (1995), p. 291-331 (-343 for those who want more)]

An SVAR model to analyze short term effects of Swiss monetary policy

$(\Delta \log p_t, \Delta \log y_t, \Delta \log m_t, \Delta r_t)$ with $\Delta \log p_t = \log p_t - \log p_{t-1}$

$p =$ Consumer price index
$y =$ GDP in 1990 Swiss francs
$m =$ money stock M1
$r =$ quarterly average of three month Swiss franc LIBOR rate of interest

Denote: $\tilde{p}_t = \Delta log p_t \quad \tilde{y}_t = \Delta log y_t \quad \tilde{m}_t = \Delta log m_t \quad \tilde{r}_t = \Delta r_t$

$$\tilde{p}_t = b_{10} \qquad -b_{12}\tilde{y}_t - b_{14}\tilde{r}_t \qquad +\gamma_{11}\tilde{p}_{t-1}\gamma_{12}\tilde{y}_{t-1}+\gamma_{13}\tilde{m}_{t-1}+\gamma_{14}\tilde{r}_{t-1}+\ldots+\varepsilon_{1t}$$

$$\tilde{y}_t = b_{20} - b_{21}\tilde{p}_t \qquad -b_{24}\tilde{r}_t \qquad +\gamma_{21}\tilde{p}_{t-1}\gamma_{22}\tilde{y}_{t-1}+\gamma_{23}\tilde{m}_{t-1}+\gamma_{24}\tilde{r}_{t-1}+\ldots+\varepsilon_{2t}$$

$$\tilde{m}_t = b_{30} - b_{31}\tilde{p}_t - b_{32}\tilde{y}_t - b_{34}\tilde{r}_t \qquad +\gamma_{31}\tilde{p}_{t-1}\gamma_{32}\tilde{y}_{t-1}+\gamma_{33}\tilde{m}_{t-1}+\gamma_{34}\tilde{r}_{t-1}+\ldots+\varepsilon_{3t}$$

$$\tilde{r}_t = b_{40} - b_{41}\tilde{p}_t - b_{42}\tilde{y}_t - \qquad b_{43}\tilde{m}_t+\gamma_{41}\tilde{p}_{t-1}\gamma_{42}\tilde{y}_{t-1}+\gamma_{43}\tilde{m}_{t-1}+\gamma_{44}\tilde{r}_{t-1}+\ldots+\varepsilon_{4t}$$

Structural Vector Autoregression (SVAR) in primitive form

No distinction between exogenous and endogenous variables.

$\{\varepsilon_{it}\}$ White Noise

$$\mathbb{E}(\varepsilon_{it}) = 0,$$

$$\text{Cov}(\varepsilon_{it}\varepsilon_{it-k}) = 0 \text{ for all } k \neq 0$$

$$\text{Var}(\varepsilon_{it}) = \sigma_i^2, \quad \text{Cov}(\varepsilon_{it}\varepsilon_{jt}) \quad = 0 \text{ for all } j \neq i$$

The key tool to trace short run effects of monetary policy with an SVAR is the impulse-response function

Q: How are monetary shocks absorbed by the system?

Tool: Impulse-Response function $\Rightarrow$ Traces the dynamics of the system

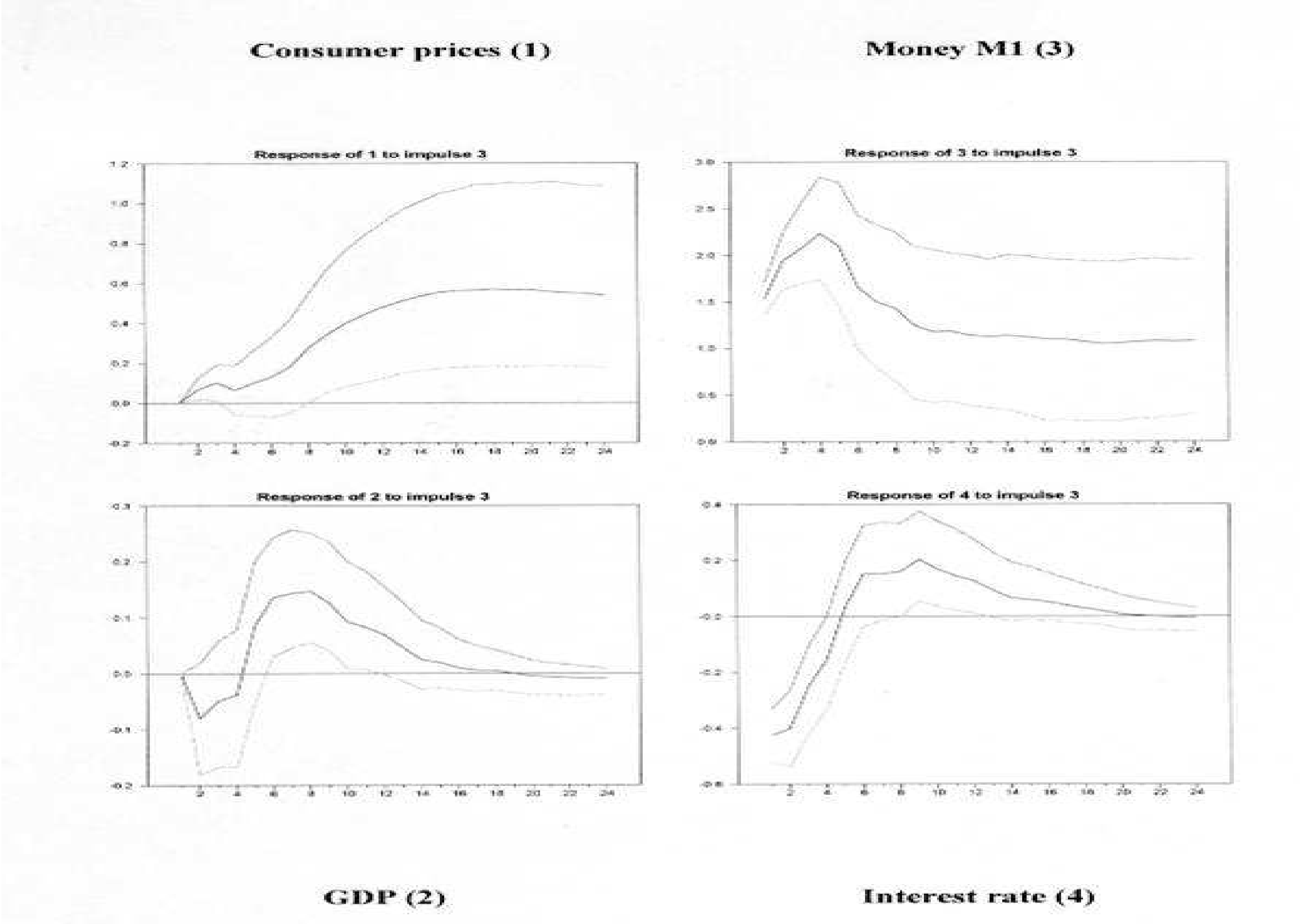
Based on the reduced form of the system. VAR in standard form:

$$
\begin{aligned}
\tilde{p}_t &= a_{11}\tilde{p}_{t-1} + a_{12}\tilde{y}_{t-1} + a_{13}\tilde{m}_{t-1} + a_{14}\tilde{r}_{t-1}... + e_{1t} \\
\tilde{y}_t &= a_{21}\tilde{p}_{t-1} + a_{22}\tilde{y}_{t-1} + a_{23}\tilde{m}_{t-1} + a_{24}\tilde{r}_{t-1}... + e_{2t} \\
\tilde{m}_t &= a_{31}\tilde{p}_{t-1} + a_{32}\tilde{y}_{t-1} + a_{33}\tilde{m}_{t-1} + a_{34}\tilde{r}_{t-1}... + e_{3t} \\
\tilde{r}_t &= a_{41}\tilde{p}_{t-1} + a_{42}\tilde{y}_{t-1} + a_{43}\tilde{m}_{t-1} + a_{44}\tilde{r}_{t-1}... + e_{4t}
\end{aligned}
$$

Estimate parameters by OLS: $\{\tilde{p}_t\}$, $\{\tilde{y}_t\}$, $\{\tilde{m}_t\}$, $\{\tilde{p}_{t-1}\}$

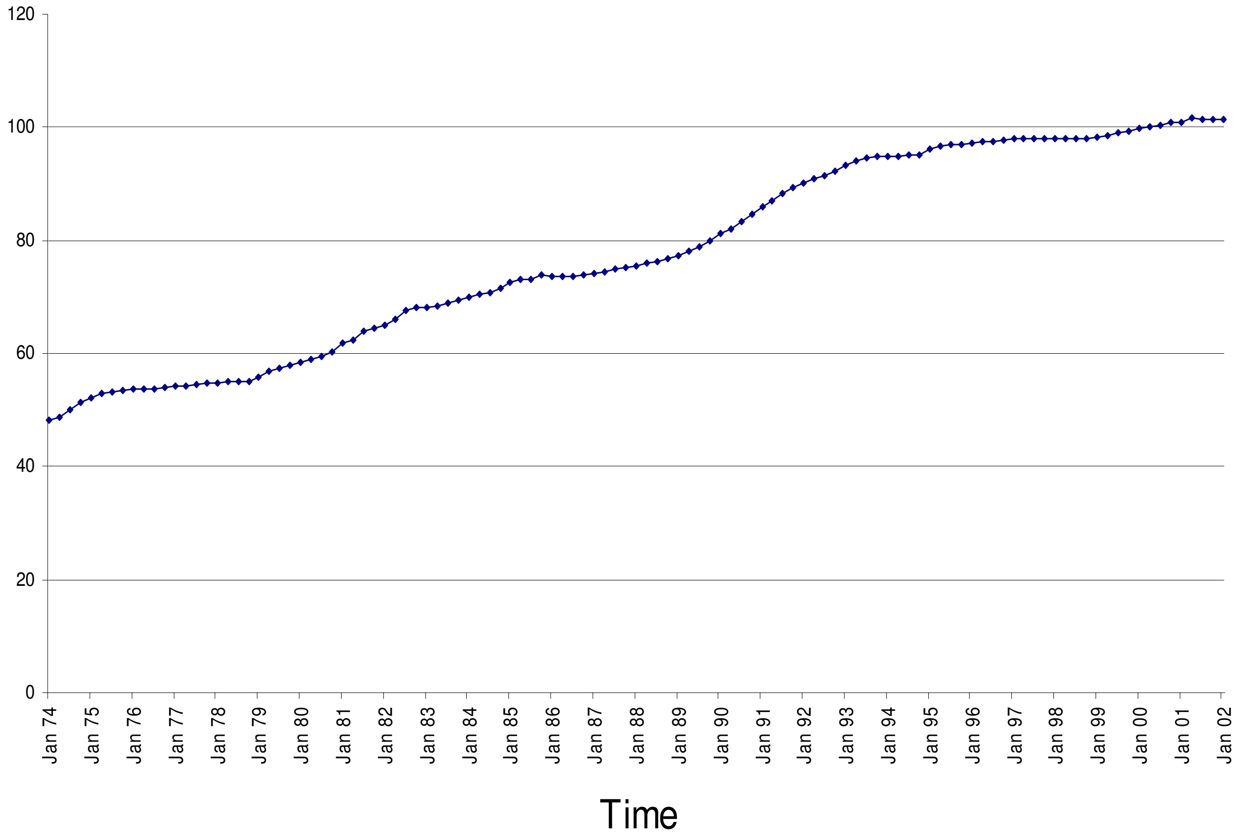
Ceteris paribus shock of the system, trace forward responses of the system.

Key tool to trace short run effects of monetary policy with SVAR is the impulse-response function

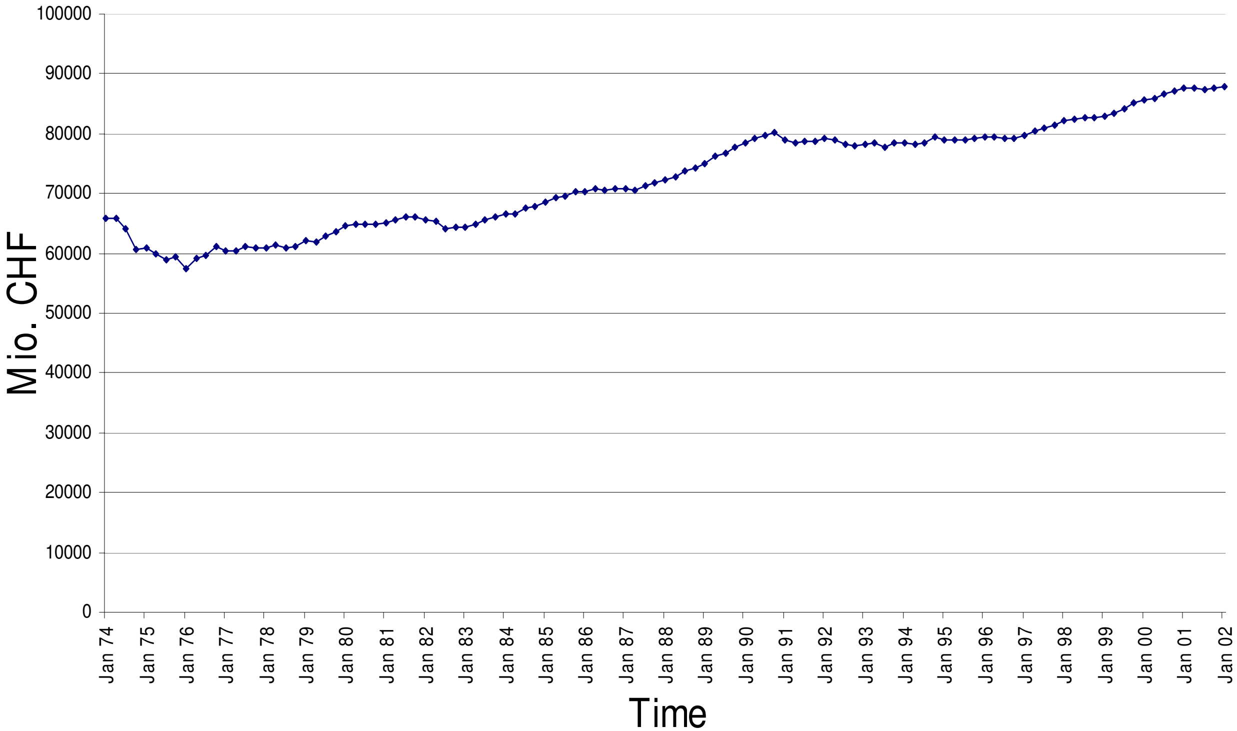


LIBOR stands for London Interbank Offered Rate. It refers to any of a number of short-term indicative interest rates compiled by the British Bankers Association (BBA) at 11:00 AM London time, each business day. LIBOR is quoted for monthly maturities out to a year for many of the world's currencies. Rates are available from Telerate news service page 3750. LIBOR rates are widely used as the underlying interest rates for derivative contracts for all currencies except the Euro.
Similar rates are quoted in other world capitals. PIBOR stands for Paris Interbank Offered Rate. TIBOR stands for Tokyo Interbank Offered Rates. Use of these rates is modest.
Euribor stands for Euro Interbank Offered Rate. These interest rates for the Euro are compiled by the European Banking Federation (FBE—*Fédération Bancaire de l'Union Européenne*) and are released at 11:00 AM Brussels time, each business day. Rates are quoted for one week and monthly maturities out to a year. They are available on Telerate page 248. Euribor is widely used as the underlying interest rate for Euro-denominated derivative contracts.

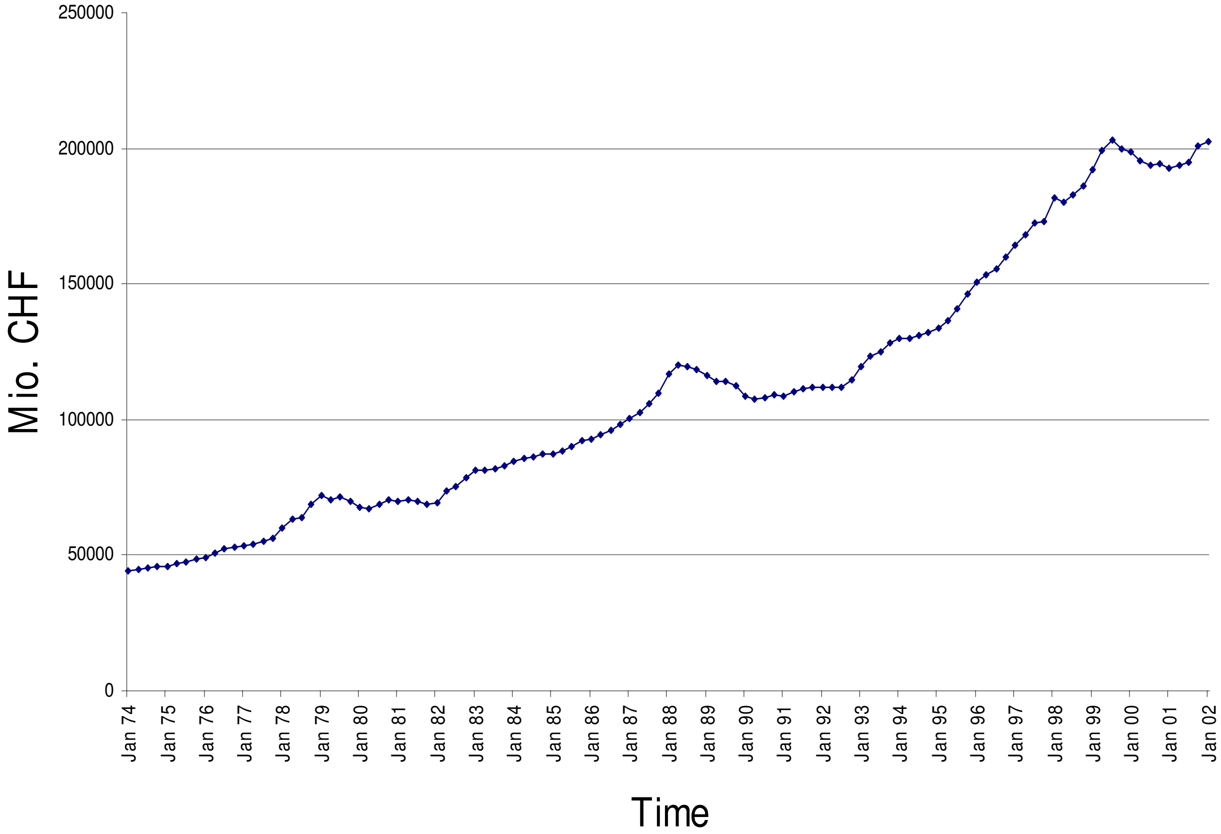
Consumer price index



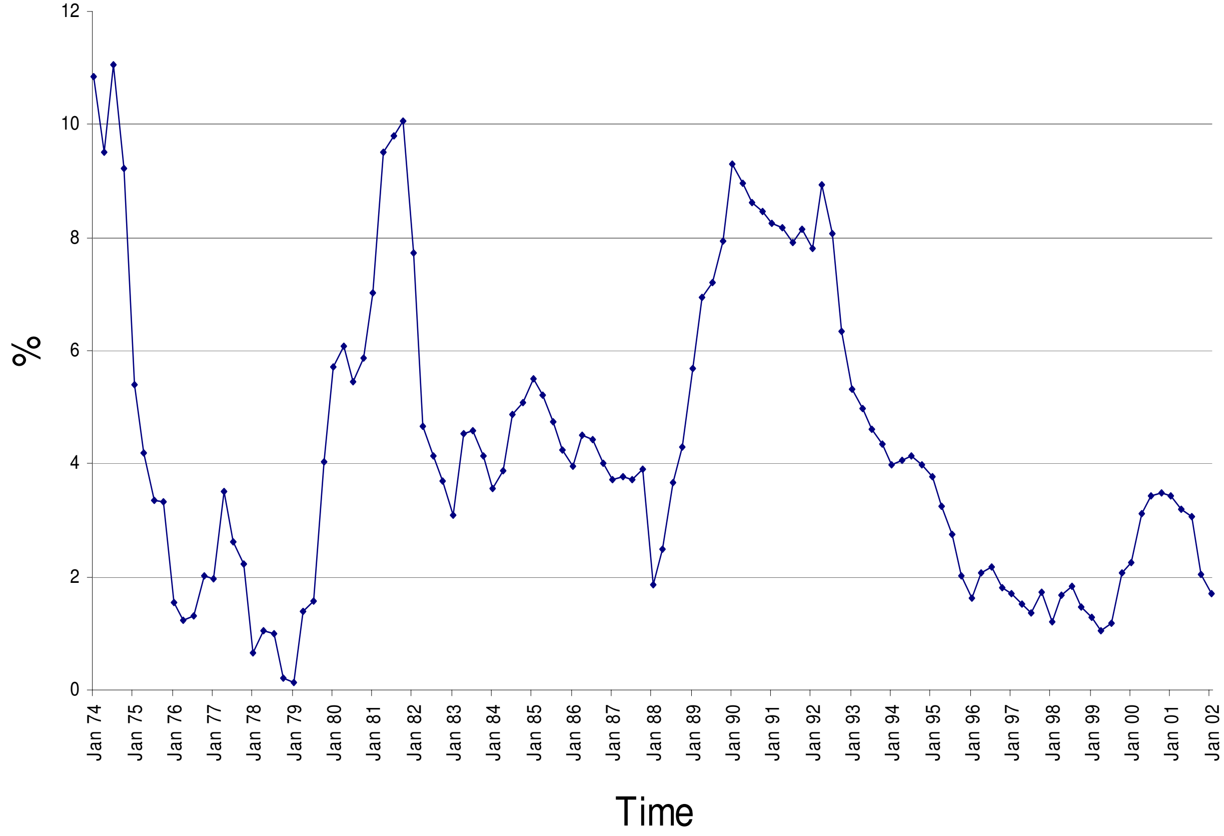
Gross Domestic Product in 1990 Swiss



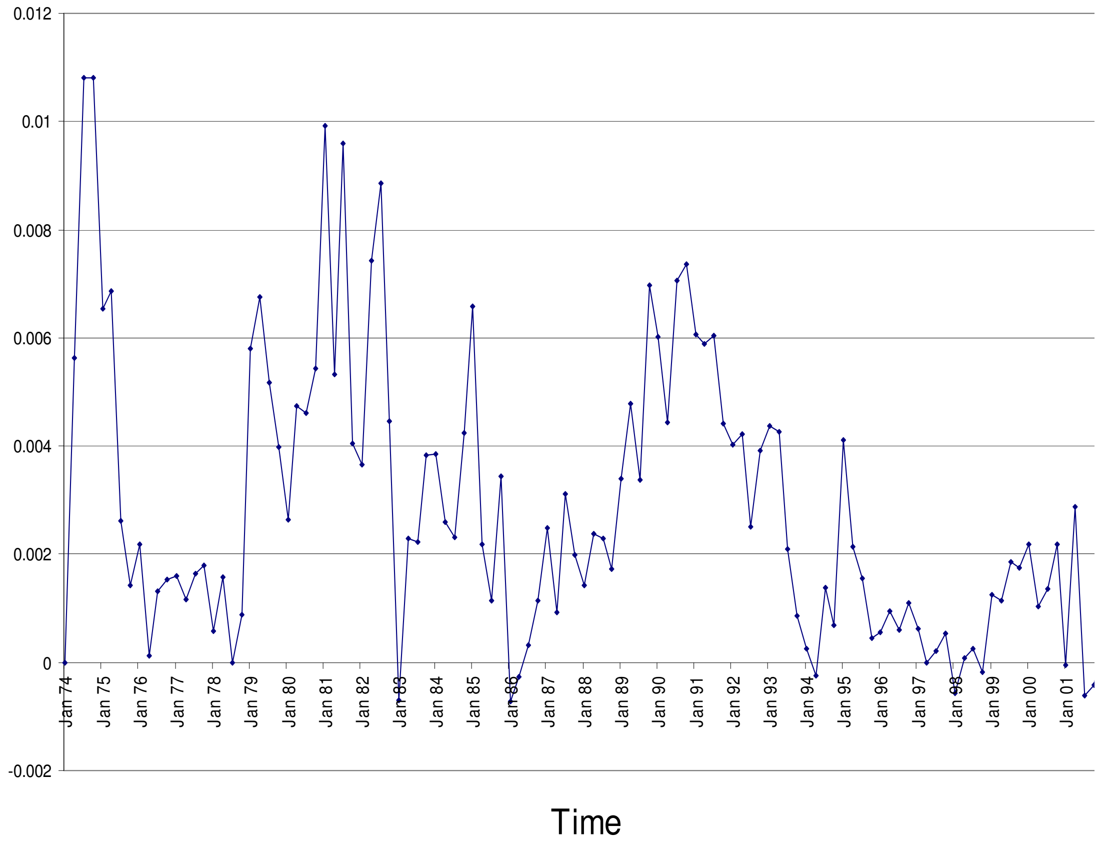
Money stock M1



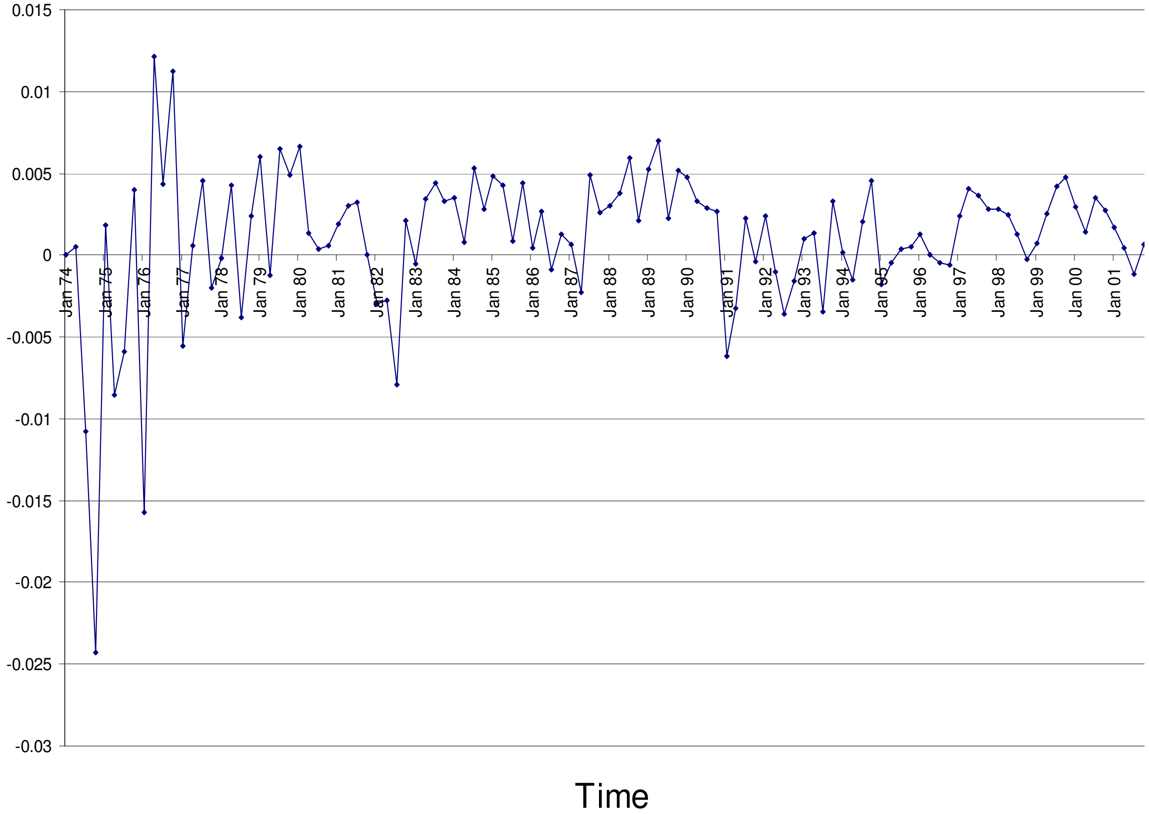
Three month Swiss LIBOR rate of interest



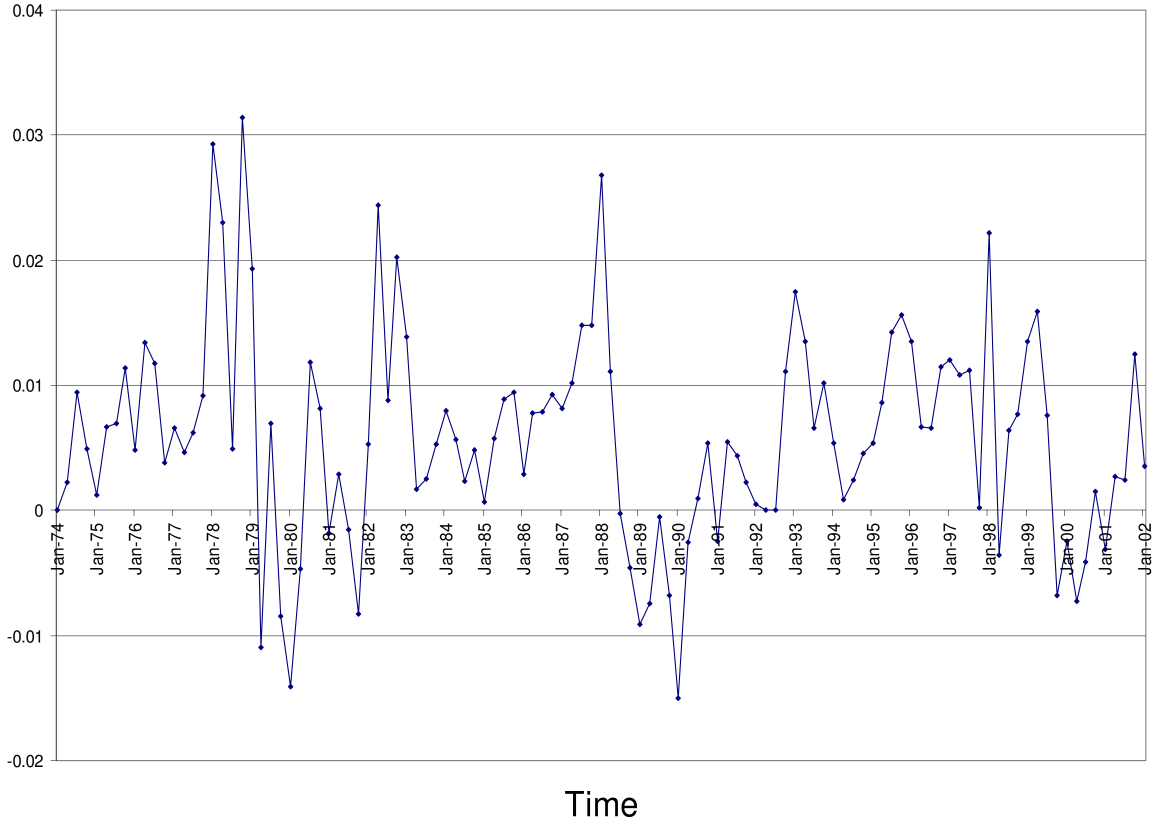
$\Delta \log p_t$


0.012
0.01
0.008
0.006
0.004
0.002
0
-0.002
Jan 74
Jan 75
Jan 76
Jan 77
Jan 78
Jan 79
Jan 80
Jan 81
Jan 82
Jan 83
Jan 84
Jan 85
Jan 86
Jan 87
Jan 88
Jan 89
Jan 90
Jan 91
Jan 92
Jan 93
Jan 94
Jan 95
Jan 96
Jan 97
Jan 98
Jan 99
Jan 00
Jan 01
Time

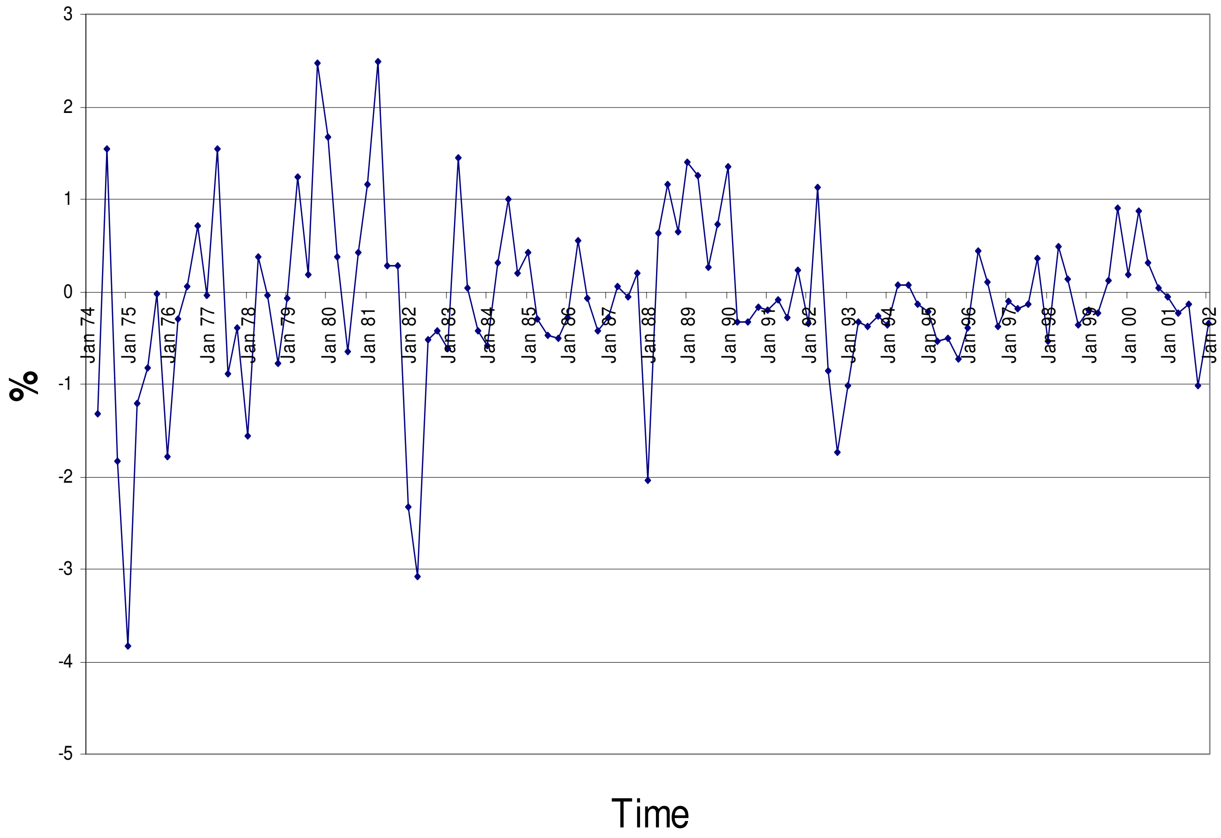
$\Delta \log y_t$


0.015
0.01
0.005
0
-0.005
-0.01
-0.015
-0.02
-0.025
-0.03
Jan 74
Jan 75
Jan 76
Jan 77
Jan 78
Jan 79
Jan 80
Jan 81
Jan 82
Jan 83
Jan 84
Jan 85
Jan 86
Jan 87
Jan 88
Jan 89
Jan 90
Jan 91
Jan 92
Jan 93
Jan 94
Jan 95
Jan 96
Jan 97
Jan 98
Jan 99
Jan 00
Jan 01
Time

$\Delta \log m_t$


0.04
0.03
0.02
0.01
0
-0.01
-0.02
Jan-74
Jan-75
Jan-76
Jan-77
Jan-78
Jan-79
Jan-80
Jan-81
Jan-82
Jan-83
Jan-84
Jan-85
Jan-86
Jan-87
Jan-88
Jan-89
Jan-90
Jan-91
Jan-92
Jan-93
Jan-94
Jan-95
Jan-96
Jan-97
Jan-98
Jan-99
Jan-00
Jan-01
Jan-02
Time

$\Delta \log r_t$



3
2
1
0
-1
-2
-3
-4
-5
%
Jan 74
Jan 75
Jan 76
Jan 77
Jan 78
Jan 79
Jan 80
Jan 81
Jan 82
Jan 83
Jan 84
Jan 85
Jan 86
Jan 87
Jan 88
Jan 89
Jan 90
Jan 91
Jan 92
Jan 93
Jan 94
Jan 95
Jan 96
Jan 97
Jan 98
Jan 99
Jan 00
Jan 01
Jan 02
Time

Structural VAR: Methodology in a bivariate system

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt}$$
$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt}$$

$$\mathbb{E}(\varepsilon_{yt}) = 0,\ \text{Var}(\varepsilon_{yt}) = \sigma_y^2,\ \text{Cov}(\varepsilon_{yt}, \varepsilon_{yt-j}) = 0$$
$$\mathbb{E}(\varepsilon_{zt}) = 0,\ \text{Var}(\varepsilon_{zt}) = \sigma_z^2,\ \text{Cov}(\varepsilon_{zt}, \varepsilon_{zt-j}) = 0,\ \text{Cov}(\varepsilon_{yt}, \varepsilon_{zt}) = 0$$

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$
$$x_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 x_{t-1} + B^{-1}\varepsilon_t$$
$$x_t = A_0 + A_1 x_{t-1} + e_t$$
$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t}$$
$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t}$$

The stochastic properties of the ‚composite‘ shocks in the standard form VAR differ from those in the primitive form: Contemporaneous Covariances! (1)

$$e_{1t} = \frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}} \qquad e_{2t} = \frac{\varepsilon_{zt} - b_{21}\varepsilon_{yt}}{1 - b_{12}b_{21}}$$

$$\mathbb{E}(e_{1t}) = E\left[\frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}}\right] = 0$$

$$\mathbb{E}(e_{2t}) = E\left[\frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}}\right] = 0$$

$$\mathsf{Var}(e_{1t}) = \mathbb{E}(e_{1t}^2) = E\left[\frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}}\right]^2 = \frac{\sigma_y^2 + b_{12}^2\sigma_z^2}{(1 - b_{12}b_{21})^2}$$

$$\mathsf{Var}(e_{2t}) = \ldots$$

The stochastic properties of the ‚composite' shocks in the standard form VAR differ from those in the primitive form: Contemporaneous Covariances! (2)

$$\text{Cov}(e_{1t}e_{1t-i}) = \mathbb{E}(e_{1t}e_{1t-i}) = \frac{E\left[(\varepsilon_{yt}-b_{12}\varepsilon_{zt})(\varepsilon_{yt-i}-b_{12}\varepsilon_{zt-i})\right]}{(1-b_{12}b_{21})^2} = 0$$

$$\text{Cov}(e_{2t}e_{2t-i}) = \mathbb{E}(e_{1t}e_{1t-i}) = \frac{E\left[(\varepsilon_{yt}-b_{12}\varepsilon_{zt})(\varepsilon_{yt-i}-b_{12}\varepsilon_{zt-i})\right]}{(1-b_{12}b_{21})^2} = 0$$

$$\text{Cov}(e_{1t}e_{2t}) \quad = \mathbb{E}(e_{1t}e_{2t}) \quad = \frac{E\left[(\varepsilon_{yt}-b_{12}\varepsilon_{zt})(\varepsilon_{zt}-b_{21}\varepsilon_{yt})\right]}{(1-b_{12}b_{21})^2} \quad \neq 0$$

$$\Sigma = \begin{bmatrix} \text{Var}(e_{1t}) & \text{Cov}(e_{1t}, e_{2t}) \\ \text{Cov}(e_{1t}, e_{2t}) & \text{Var}(e_{2t}) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \qquad \sigma_{12} = \sigma_{21}$$

$$\Sigma_\varepsilon = \begin{bmatrix} \text{Var}(\varepsilon_{yt}) & \text{Cov}(\varepsilon_{yt}, \varepsilon_{zt}) \\ \text{Cov}(\varepsilon_{yt}, \varepsilon_{zt}) & \text{Var}(\varepsilon_{zt}) \end{bmatrix} = \begin{bmatrix} \dots \end{bmatrix}$$

For Impulse Response Analysis we need to trace the effects of the shocks to the primitive VAR

Parameters of the unrestricted primitive VAR cannot be consistently estimated by OLS.
VAR in primitive form not useful for forecasting and IR analysis

$$
\begin{aligned}
y_t &= b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{yt} \\
z_t &= b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{zt} \\
B x_t &= \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t
\end{aligned}
$$

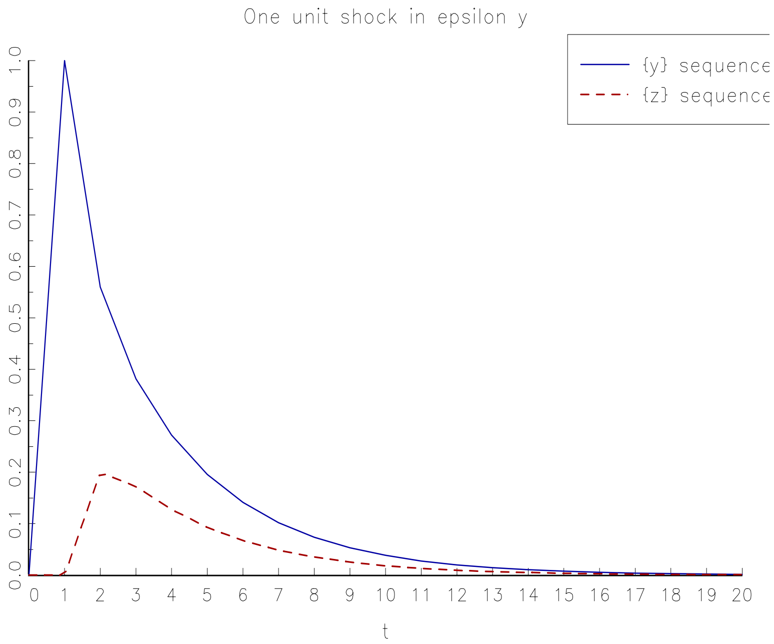
For estimation, forecasting and IRF analysis we need to use the standard form of the VAR

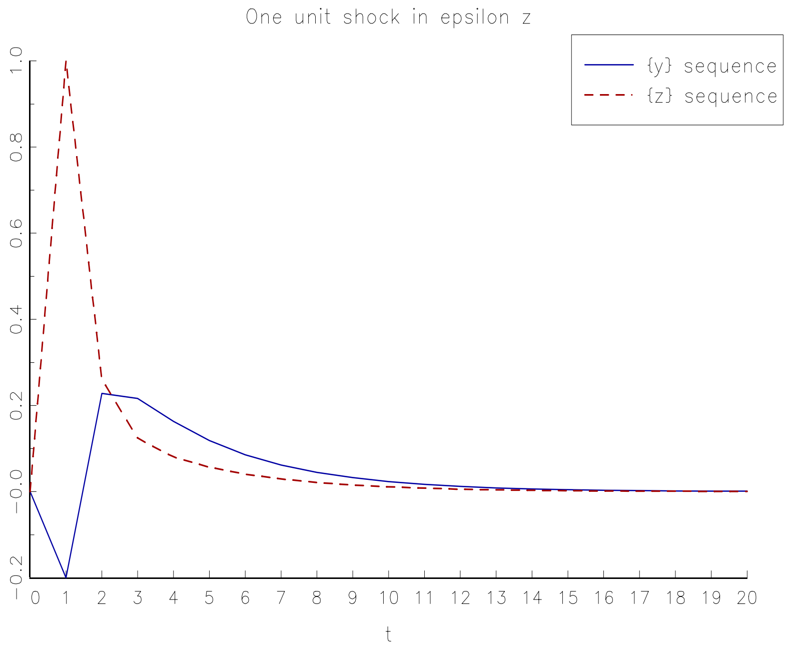
$$
\begin{aligned}
x_t &= B^{-1}\Gamma_0 + B^{-1}\Gamma_1 x_{t-1} + B^{-1}\varepsilon_t \\
y_t &= a_{10} + a_{11} y_{t-1} + a_{12} z_{t-1} + e_{1t} \\
z_t &= a_{20} + a_{21} y_{t-1} + a_{22} z_{t-1} + e_{2t}
\end{aligned}
$$

...But to trace the effects of the shocks, we need to know B.

For Impulse Response Analysis we need to trace the effects of the shocks to the primitive VAR

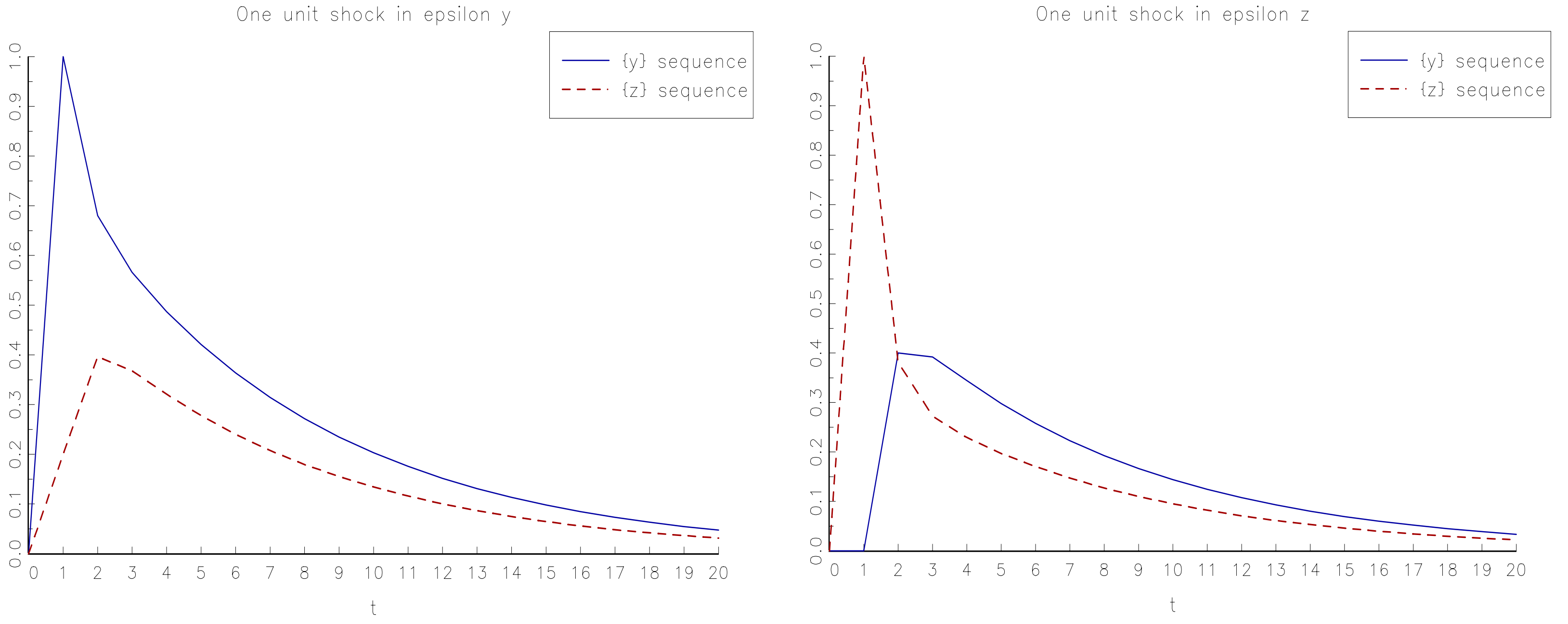
$$
\begin{aligned}
y_t &= -0.2z_t + 0.6y_{t-1} + 0.4z_{t-1} + \varepsilon_{yt} \\
z_t &= +0.2y_{t-1} + 0.3z_{t-1} + \varepsilon_{zt}
\end{aligned}
$$





For Impulse Response Analysis we need to trace the effects of the shocks to the primitive VAR

$$
\begin{aligned}
y_t &= 0.6y_{t-1} + 0.4z_{t-1} + \varepsilon_{yt} \\
z_t &= 0.2y_t + 0.2y_{t-1} + 0.3z_{t-1} + \varepsilon_{zt}
\end{aligned}
$$



How to produce Impulse Response functions when SVAR parameters of primitive form are not known? Step 1 - 2

1. Write down VAR in primitive form

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{yt}$$
$$z_t = b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{zt}$$
$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$

2. Write VAR in standard form. Estimate equation by equation using OLS

$$x_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 x_{t-1} + B^{-1}\varepsilon_t$$
$$y_t = a_{10} + a_{11} y_{t-1} + a_{12} z_{t-1} + e_{1t}$$
$$z_t = a_{20} + a_{21} y_{t-1} + a_{22} z_{t-1} + e_{2t}$$

How to produce Impulse Response functions when SVAR parameters of primitive form are not known? Step 3

3. Use estimated residual series to estimate

$$\Sigma = \begin{bmatrix} \mathsf{Var}(e_{1t}) & \mathsf{Cov}(e_{1t}, e_{2t}) \\ \mathsf{Cov}(e_{1t}, e_{2t}) & \mathsf{Var}(e_{2t}) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

$$\hat{\sigma}_1^2 = (\frac{1}{T}) \sum_{t=1}^{T} \hat{e}_{1t}^2, \qquad \hat{\sigma}_2^2 = (\frac{1}{T}) \sum_{t=1}^{T} \hat{e}_{2t}^2, \qquad \hat{\sigma}_{12} = (\frac{1}{T}) \sum_{t=1}^{T} \hat{e}_{1t}\hat{e}_{2t}$$

How to produce Impulse Response functions when SVAR parameters of primitive form are not known? Step 4

4. Exploit relationship of variances and covariances of primitive shocks and variances and covariances of composite shocks.

$$\Sigma_\varepsilon = \begin{bmatrix} \mathsf{Var}(\varepsilon_y) & 0 \\ 0 & \mathsf{Var}(\varepsilon_z) \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \quad e_t = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \quad \varepsilon_t = \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix}$$

$$e_t = B^{-1}\varepsilon_t \quad \Rightarrow \quad \varepsilon_t = Be_t$$

$$\Sigma_\varepsilon = B\Sigma B'$$

$$\begin{bmatrix} \mathsf{Var}(\varepsilon_1) & 0 \\ 0 & \mathsf{Var}(\varepsilon_2) \end{bmatrix} = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 & b_{21} \\ b_{12} & 1 \end{bmatrix}$$

How to produce Impulse Response functions when SVAR parameters of primitive form are not known? (Step 4 cont.)

And you are generally stuck!

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t \Rightarrow \quad x_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 x_{t-1} + B^{-1}\varepsilon_t$$
$$\Rightarrow \quad x_t = A_0 + A_1 x_{t-1} + e_t$$

$$\begin{bmatrix} 1 & b_{12} & b_{13} & \dots & b_{1n} \\ b_{21} & 1 & b_{23} & \dots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \dots & 1 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{nt} \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \\ \vdots \\ b_{n0} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \dots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \dots & \gamma_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \dots & \gamma_{nn} \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{nt-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}$$

$$\Sigma_\varepsilon = \begin{bmatrix} \mathsf{Var}(\varepsilon_1) & 0 & \dots & 0 \\ 0 & \mathsf{Var}(\varepsilon_1) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \mathsf{Var}(\varepsilon_n) \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix} \quad \Sigma_\varepsilon = B\Sigma B'$$

How to produce Impulse Response functions when SVAR parameters of primitive form are not known? (Step 4 cont.)

A solution: Cholesky Decomposition: $\Sigma_\varepsilon = B\Sigma B'$

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ b_{21} & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \dots & 1 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{nt} \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \\ \vdots \\ b_{n0} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \cdots & \gamma_{nn} \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{nt-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ b_{21} & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \dots & 1 \end{bmatrix} \Rightarrow B^{-1} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ c_{21} & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \dots & 1 \end{bmatrix}$$

$e_t = B^{-1}\varepsilon_t$

$e_{1t} = \varepsilon_{1t}$

$e_{2t} = c_{21}\varepsilon_{1t} + \varepsilon_{2t}$

$e_{3t} = c_{31}\varepsilon_{1t} + c_{32}\varepsilon_{2t} + \varepsilon_{3t}$

$\vdots$

$e_{nt} = c_{n1}\varepsilon_{1t} + c_{n2}\varepsilon_{2t} + \cdots + \varepsilon_{nt}$

How to produce Impulse Response functions when SVAR parameters of primitive form are not known? (Step 4 cont.)

Two variable numerical example:

$$\begin{aligned}
\Sigma &= \begin{bmatrix} 0.5 & 0.4 \\ 0.4 & 0.5 \end{bmatrix} \\
\Sigma_\varepsilon &= \begin{bmatrix} \mathsf{Var}(\varepsilon_y) & 0 \\ 0 & \mathsf{Var}(\varepsilon_z) \end{bmatrix} \\
\Sigma_\varepsilon &= B\Sigma B'
\end{aligned}$$

$$\begin{aligned}
&\mathsf{Var}(\varepsilon_y) = 0.5 \quad + 0.8b_{12} \quad + 0.5b_{12}^2 \\
&0 = 0.5b_{21} + 0.4b_{21}b_{12} + 0.4 \quad + 0.5b_{12} \\
&0 = 0.5b_{21} + 0.4b_{12}b_{21} + 0.4 \quad + 0.5b_{12} \\
&\mathsf{Var}(\varepsilon_z) = 0.5b_{21}^2 + 0.8b_{21} \quad + 0.5
\end{aligned}$$

How to produce Impulse Response functions when SVAR parameters of primitive form are not known? Step 4 cont.

Choosing a different ordering of variables in the Cholesky decomposition

$$
\begin{aligned}
\text{Var}(\varepsilon_y) &= 0.5 \quad + 0.8b_{12} \quad + 0.5b_{12}^2 \\
0 &= 0.5b_{21} + 0.4b_{21}b_{12} + 0.4 \quad + 0.5b_{12} \\
0 &= 0.5b_{21} + 0.4b_{12}b_{21} + 0.4 \quad + 0.5b_{12} \\
\text{Var}(\varepsilon_z) &= 0.5b_{21}^2 + 0.8b_{21} \quad + 0.5
\end{aligned}
$$

How to produce Impulse Response functions when SVAR parameters of primitive form are not known? Step 5

5. Use identified B matrix and estimate coefficients of standard VAR to trace a shock of one standard deviation in shocks (shocks to the primitive system).

   (Fix starting values, set future shocks equal to zero) $e_t = B^{-1}\begin{pmatrix} \sqrt{\text{Var}(\varepsilon_y)} \\ 0 \end{pmatrix}$ Iterate forward to see how shock is transferred by the system

   $x_{t+1} = \hat{A}_0 + \hat{A}_1 x_t + e_t \quad x_{t+2} = \hat{A}_0 + \hat{A}_1 x_{t+1} \quad x_{t+3} = \hat{A}_0 + \hat{A}_1 x_{t+2} \quad \cdots$

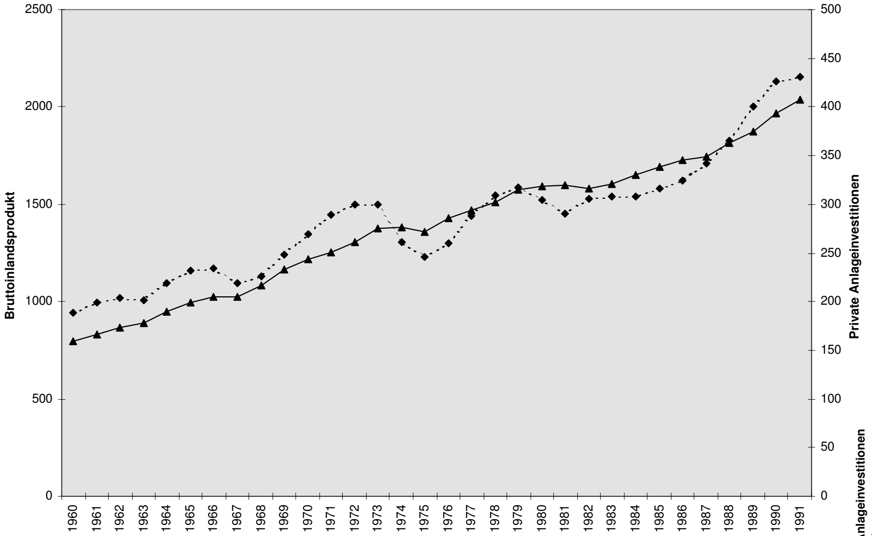
   Plot resulting series

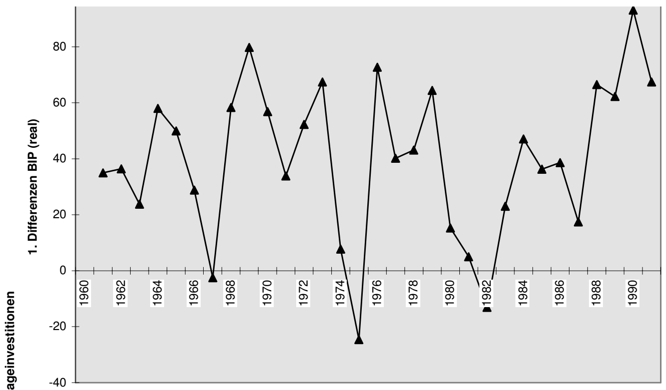
You're stuck! Why?

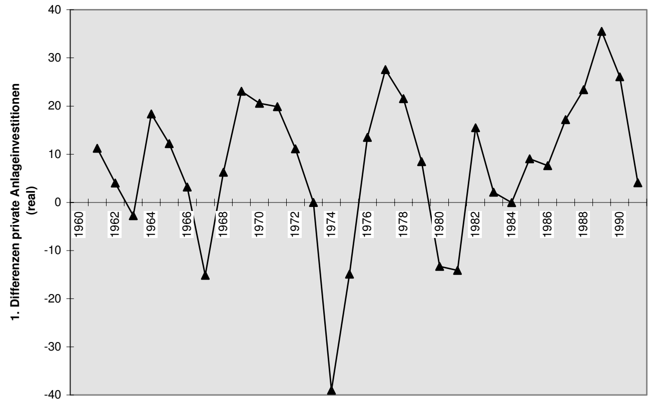
VI. Cointegration and Error Correction Models (Basics)

[Hamilton (1994), Chapter 19 (parts);
Enders (1995), Chapter 6 (parts)
Hayashi (2000), Chapter 10]

Many time series in economics and finance look like realizations of non-stationary stochastic processes







Ordinary Least Squares regression using non-stationary time series is hazardous!

Applying OLS to macro time series yields

small t-values

high R^2

positively autocorrelated residuals

Granger and Newbold

Simulation of independent random walk processes : $\begin{array}{c} y_{1t} = y_{1t-1} + u_{1t} \\ y_{2t} = y_{2t-1} + u_{2t} \\ \vdots \end{array}$,

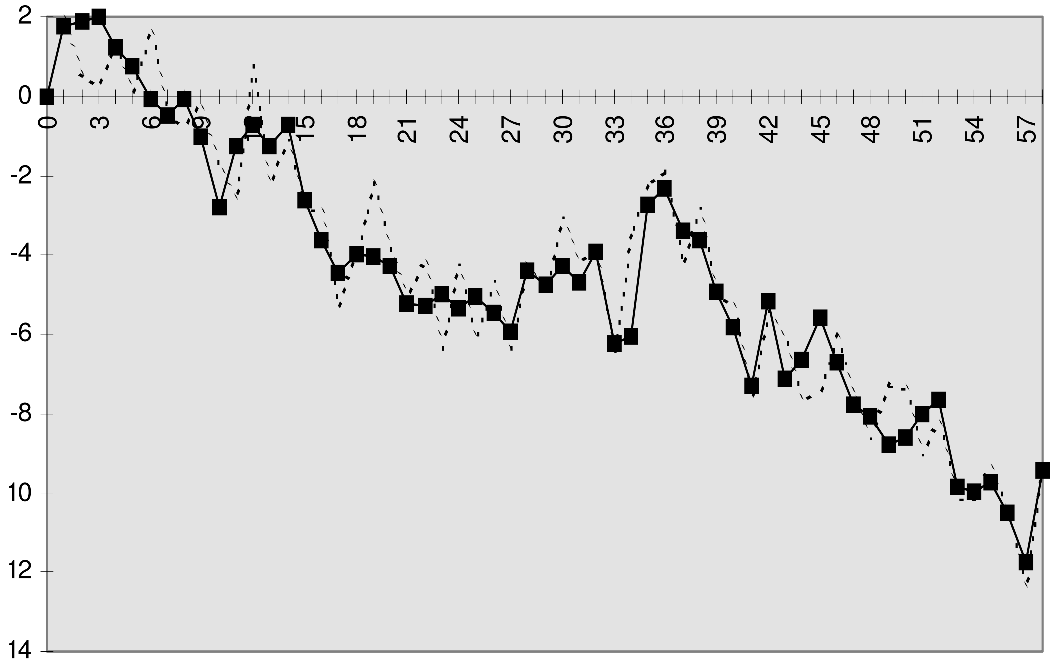
where $y_{1t} = \alpha + \gamma_1 y_{2t-1} + \gamma_2 y_{3t-1} + \ldots + e_t$

Multiple linear regression

Result: Regression yields too often to a rejection of the (correct) null hypothesis that slope parameters are zero.

Cointegration: Graphical illustration

$y_{1t} = y_{2t-1} + u_{1t} \qquad\qquad y_{2t} = y_{2t-1} + u_{2t}$



Cointegration: An economic interpretation

Long run equilibrium relation of economic time series
Possibility of short term deviations from equilibrium
Economic mechanisms move system to equilibrium

Examples:

Term structure of interest rates
Stock prices of assets traded on different markets
Purchase power parity between two countries
Consumption and Income

Cointegration: A definition

Long run equilibrium relation of economic time series

$(n \times 1)$ vector of time series $\mathbf{y}_t = (y_{1t}, y_{2t}, y_{3t}, \ldots, y_{nt})'$ is cointegrated if each series is

- ◇ non-stationary (integrated of order one)
- ◇ there exists (at least one) $\mathbf{a}'\mathbf{y}_t$ linear combination which produces a stationary process

Bivariate example:

$$
\begin{aligned}
y_{1t} &= \gamma \cdot y_{2t} + u_{1t} \\
y_{2t} &= y_{2t-1} + u_{2t} \\
y_{1t} - y_{1t-1} &= \gamma \cdot u_{2t} + u_{1t} - u_{1t-1} \\
y_{2t} - y_{2t-1} &= \Delta y_{2t} = u_{2t}
\end{aligned}
$$

Linear combination $y_{1t} - \gamma \cdot y_{2t} = u_{1t}$ is stationary

$y_{1t} - \gamma \cdot y_{2t}$: Cointegrating relation $\mathbf{a} = (1, -\gamma)'$ cointegrating vector

Cointegration: An economic example Purchase Power Parity (PPP)

No transaction costs and free trade

P_t^S: Index of price level Switzerland (CHF per good)

P_t^U: Index of price level USA ($ per good)

S_t: Exchange rate (Dollar/CHF)

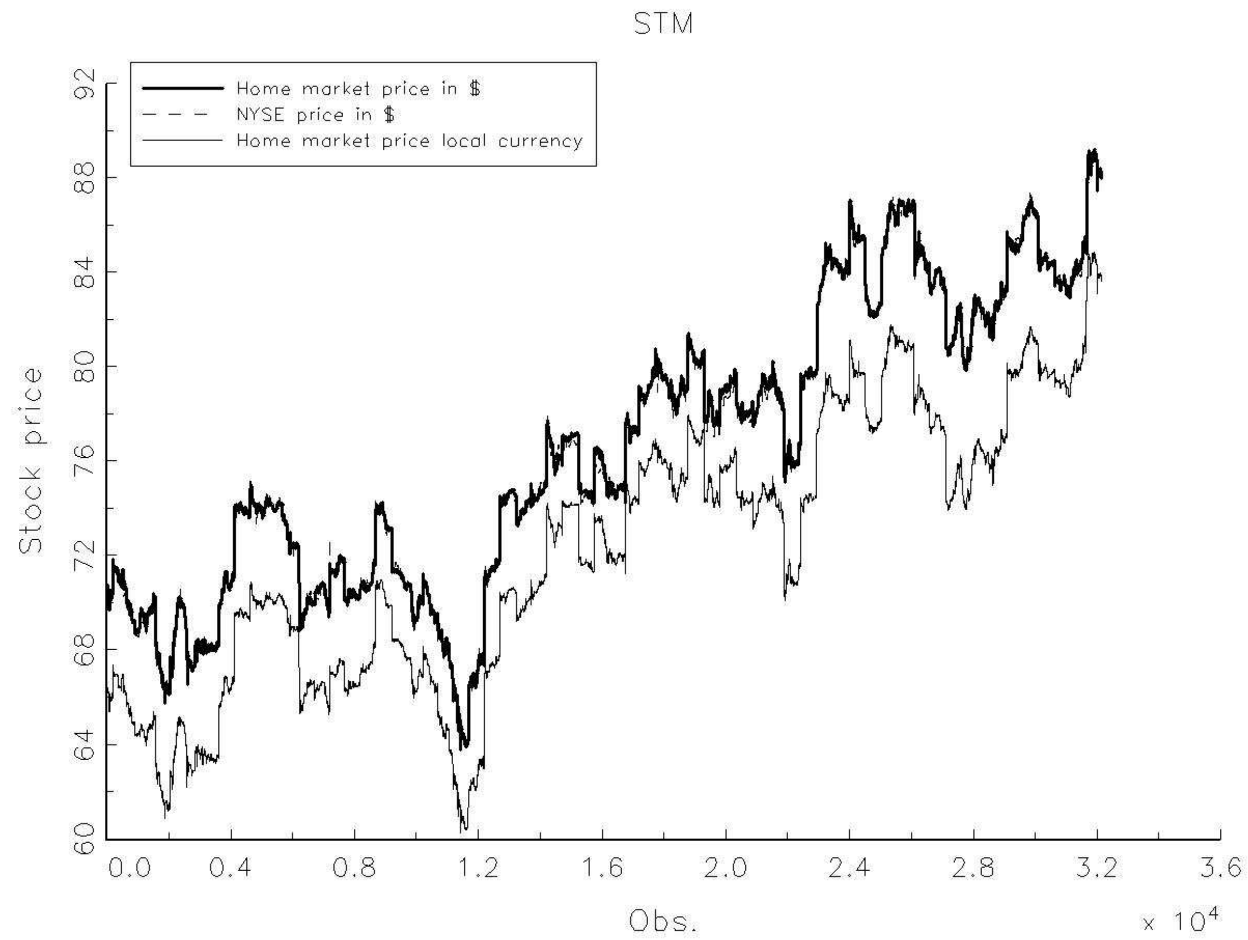
$$P_t^U = S_t \cdot P_t^S$$

in logs: $p_t^U = s_t + p_t^S \quad \Rightarrow \quad p_t^U - p_t^S - s_t = 0$

Weaker version of PPP: $z_t \equiv p_t^U - p_t^S - s_t$,

where $\{z_t\}$ is a stationary stochastic process

Purchase power parity in the real world: Assets traded on parallel markets



The appropriate econometric specification to model dynamics of cointegrated time series: The equilibrium correction model

The bivariate case

$$\Delta y_{1t} = a_0 + \gamma_1(y_{1t-1} - \beta_0 - \beta_1 y_{2t-1}) + a_{11}\Delta y_{1t-1} + a_{12}\Delta y_{2t-1} + \text{more lags} + u_{1t}$$

$$\Delta y_{2t} = b_0 + \gamma_2(y_{1t-1} - \beta_0 - \beta_1 y_{2t-1}) + a_{21}\Delta y_{1t-1} + a_{22}\Delta y_{2t-1} + \text{more lags} + u_{2t}$$

Multivariate case: Number of cointegrating relations?

Engle and Granger have proposed a method to estimate the parameters of a cointegrated system (1)

OLS estimation of ECM not feasible

$\Delta y_{1t} = a_0 + \gamma_1(y_{1t-1} - \beta_0 - \beta_1 y_{2t-1}) + a_{11}\Delta y_{1t-1} + a_{12}\Delta y_{2t-1} +$ more lags $+ u_{1t}$

$\Delta y_{2t} = b_0 + \gamma_2(y_{1t-1} - \beta_0 - \beta_1 y_{2t-1}) + a_{21}\Delta y_{1t-1} + a_{22}\Delta y_{2t-1} +$ more lags $+ u_{2t}$

Assume n variables, $h = 1$ cointegrating relation (e.g. PPP $n = 3, h = 1$)

4 step approach

Step 1:

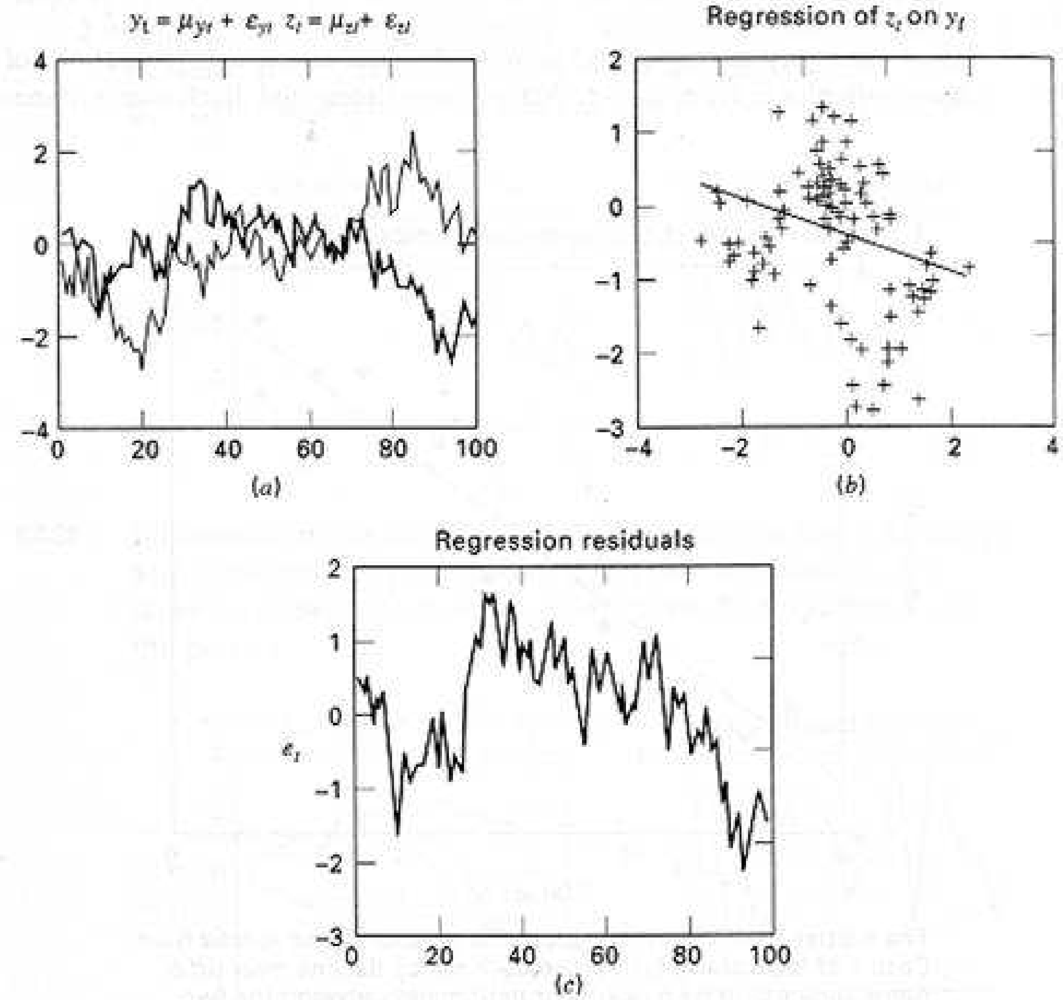
Test whether each of n variables is integrated of order one (I(1), non-stationary, unit root process, first difference yields stationary series).

Standard tests: Dickey-Fuller and Perron tests. Null hypothesis: series non-stationary. Distribution of test statistic under Null: Non-standard, obtained by simulations. Critical values (quantiles) tabulated.

If null hypothesis rejected (given α) for all series: cointegration rejected. Model with VAR. If for some variables null rejected (given α) for others not: cointegration hypothesis rejected. Exclude variables from cointegrating relation.

If null hypothesis (non-stationarity) maintained (α) proceed to step 2

Illustration: Regression when no cointegration present



Step 2

Impose normalization. Put one variable on LHS, others on RHS. Run a regression: exemplary: $n = 2$

$$y_{1t} = \beta_0 + \beta_1 y_{2t} + \varepsilon_t$$

Back out $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\varepsilon}_t = y_{1t} - \hat{\beta}_0 - \hat{\beta}_1 y_{2t}$

If y_{1t}, y_{2t} are cointegrated $\Rightarrow \quad \varepsilon_t = y_{1t} - \beta_0 - \beta_1 y_{2t}$ is stationary

Test nonstationarity of series $\hat{\varepsilon}_t = y_{1t} - \hat{\beta}_0 - \hat{\beta}_1 y_{2t}$ using stationarity tests
Residual series based on estimated parameters: Different distribution of test statistic: Use correct critical value tables!
If null hypothesis of non-stationarity of $\hat{\varepsilon}_t$ is rejected, proceed to step 3.

Step 3

Replace in ECM

$$\Delta y_{1t} = a_0 + \gamma_1(y_{1t-1} - \beta_0 - \beta_1 y_{2t-1}) + a_{11}\Delta y_{1t-1} + a_{12}\Delta y_{2t-1} + \text{more lags} + u_{1t}$$

$$\Delta y_{2t} = b_0 + \gamma_2(y_{1t-1} - \beta_0 - \beta_1 y_{2t-1}) + a_{21}\Delta y_{1t-1} + a_{22}\Delta y_{2t-1} + \text{more lags} + u_{2t}$$

$$y_{1t} - \beta_0 - \beta_1 y_{2t}$$

by

$$\hat{\varepsilon}_t = y_{1t} - \hat{\beta}_0 - \hat{\beta}_1 y_{2t}$$

$$\Delta y_{1t} = a_0 + \gamma_1 \hat{\varepsilon}_{t-1} + a_{11}\Delta y_{1t-1} + a_{12}\Delta y_{2t-1} + \text{more lags} + u_{1t}$$

$$\Delta y_{2t} = b_0 + \gamma_1 \hat{\varepsilon}_{t-1} + a_{21}\Delta y_{1t-1} + a_{22}\Delta y_{2t-1} + \text{more lags} + u_{2t}$$

Estimate parameters by OLS. Regression with only stationary variables on both sides.

Step 4

Innovation accounting

Plot impulse-response functions iterating on

$$\Delta y_{1t} = \hat{a}_0 + \hat{\gamma}_1(y_{1t-1} - \hat{\beta}_0 - \hat{\beta}_1 y_{2t-1}) + \hat{a}_{11}\Delta y_{1t-1} + \hat{a}_{12}\Delta y_{2t-1} + \text{more lags} + u_{1t}$$
$$\Delta y_{2t} = \hat{b}_0 + \hat{\gamma}_2(y_{1t-1} - \hat{\beta}_0 - \hat{\beta}_1 y_{2t-1}) + \hat{a}_{21}\Delta y_{1t-1} + \hat{a}_{22}\Delta y_{2t-1} + \text{more lags} + u_{2t}$$

As for SVAR: Possible contemporaneous correlation of u_{1t}, u_{2t}

Plotting the impulse-response function

Cholesky Decomposition

Ordering of variables impacts results

Problems of E&G method

With n variables up to $n - 1$ cointegrating relations may exist

Conclusion step 2 may depend on ordering

Johansen method for ML estimation of Gaussian cointegrated systems.

Hypothesis regarding price discovery in international equity trading and empirical tests based on high frequency data

◇ Simultaneous trading of same asset at different trading venues

◇ Worldwide competition for liquidity. Viability of securities markets depends on performance of trading mechanisms. Efficient capital market: Value-relevant information flows quickly into prices.

◇ Q1: Price discovery in home market or at the world‘s leading trading venue?

◇ Bacidore/Sofianos (2000): ”Price discovery takes place at home and NYSE market participants take those prices as given”

Hypothesis regarding price discovery in international equity trading and empirical tests based on high frequency data (cont.)

- ♢ $\Rightarrow$ "Winner market takes all"-hypothesis (Chowdry and Nanda, RFS 1991): In case of international parallel trading one market will dominate price discovery.

- ♢ Kim/Szakmary/Mathur (JBF 2000): Home market dominates price discovery. Problem: Aggregation of price dynamics in daily data. Non-simultaneous trading (time zones).

- ♢ Q2: Symmetric reaction of stock prices to exchange rate movements?

Starting point: 100 % Price discovery in home market

P_t^h : Stock price home market at time t in Euro (log)
P_t^u : Stock price US market in \$ (log)
E_t : \$/ Euro exchange rate (log)

E_t and P_t^h follow random walks

$$\begin{aligned} E_t &= E_{t-1} + \varepsilon_t^e \\ P_t^h &= P_{t-1}^h + \varepsilon_t^h \end{aligned}$$

The US price tracks the home market price:

$$P_t^u = P_{t-1}^h + E_{t-1} + \varepsilon_t^u$$

Cointegration between home market price, US price and exchange rate

Arbitrage prevents long run deviations from equilibrium ⇒ log-exchange rate, log-€-Kurs und log-$-Kurs are cointegrated

$$\begin{aligned} P_t^h - P_t^u + E_t &= \\ \left[P_{t-1}^h + \varepsilon_t^h - P_{t-1}^h - E_{t-1} - \varepsilon_t^u + E_{t-1} + \varepsilon_t^e\right] &= \\ \varepsilon_t^h - \varepsilon_t^u + \varepsilon_t^e \end{aligned}$$

with cointegrating vector $(1 - 11)$.

- ♢ Only own innovations ε_t^h exert permanent impact on €-price. (100% information share)
- ♢ Only own innovations ε_t^e exert permanent impact on exchange rate. (100% information share)
- ♢ $-price: Merely transitory influence of own market innovations ε_t^u . Only home market and exchange rate innovations permanently impounded in US price.

In a general model the innovations of all three price series contribute to the long run dynamics of the system

One cointegrating relation between €-price, $-price and exchange rate but...
... innovations $\varepsilon_t^h, \varepsilon_t^e$, and ε_t^u may exert permanent effects on all three price series
... their importance (the information share) is determined empirically.

Non-stationary VAR using €-price, $-price and exchange rate.
Cointegration between €-price, $-price and exchange rate.
Granger representation theorem $\Rightarrow$ VECM

Write VECM in VMA representation and simulate VMA parameters
Decompose variance of long run effect of each price series into the effects caused by the innovations of each series.
Variance Share = Information Share

In a general model the innovations of all three price series contribute to the long run dynamics of the system

Assumptions for a general model:

ONE cointegrating relation between €-price, $-price and exchange rate but...

... Innovations $\varepsilon_t^h, \varepsilon_t^e$, and ε_t^u may exert permanent effects on all three price series.

... their importance (the information share) is determined empirically.

Estimation of the information shares is based on a VECM

Non-stationary VAR using €-price, $-price and exchange rate.

Cointegration between €-price, $-price and exchange rate.

Granger representation theorem $\Rightarrow$ VECM

$$\begin{aligned}
\Delta E_t &= \beta_1(\alpha_1 P^h_{t-1} - \alpha_2 P^u_{t-1} - \alpha_3 E_t) + \delta_{11}\Delta P^h_{t-1} + \delta_{12}\Delta P^u_{t-1} + \delta_{13}\Delta E_{t-1} + \varepsilon^e_t \\
\Delta P^h_t &= \beta_2(\alpha_1 P^h_{t-1} - \alpha_2 P^u_{t-1} - \alpha_3 E_t) + \delta_{21}\Delta P^h_{t-1} + \delta_{22}\Delta P^u_{t-1} + \delta_{23}\Delta E_{t-1} + \varepsilon^h_t \\
\Delta P^u_t &= \beta_3(\alpha_1 P^h_{t-1} - \alpha_2 P^u_{t-1} - \alpha_3 E_t) + \delta_{31}\Delta P^h_{t-1} + \delta_{32}\Delta P^u_{t-1} + \delta_{33}\Delta E_{t-1} + \varepsilon^u_t
\end{aligned}$$

By simulating the VECM we obtain the weight matrix from which the information shares can be computed

Write VECM in VMA representation: $\begin{bmatrix} \Delta E_t \\ \Delta P_t^h \\ \Delta P_t^u \end{bmatrix} = \begin{bmatrix} \varepsilon_t^e \\ \varepsilon_t^h \\ \varepsilon_t^u \end{bmatrix} + \mathbf{\Psi}_1 \begin{bmatrix} \varepsilon_{t-1}^e \\ \varepsilon_{t-1}^h \\ \varepsilon_{t-1}^u \end{bmatrix} + \mathbf{\Psi}_2 \begin{bmatrix} \varepsilon_{t-2}^e \\ \varepsilon_{t-2}^h \\ \varepsilon_{t-2}^u \end{bmatrix} + \dots$

$$\mathbf{\Psi} = \begin{bmatrix} \psi_{11} & \psi_{12} & \psi_{13} \\ \psi_{21} & \psi_{22} & \psi_{23} \\ \psi_{31} & \psi_{32} & \psi_{33} \end{bmatrix} = \mathbf{I} + \mathbf{\Psi}_1 + \mathbf{\Psi}_2 \quad + \quad \dots$$

$$\begin{pmatrix} \text{permanent impact on exchange rate} \\ \text{permanent impact on -€Price} \\ \text{permanent impact on \$-Price} \end{pmatrix} = \begin{bmatrix} \psi_{11} & \psi_{12} & \psi_{13} \\ \psi_{21} & \psi_{22} & \psi_{23} \\ \psi_{31} & \psi_{32} & \psi_{33} \end{bmatrix} \times \begin{bmatrix} \varepsilon_t^e \\ \varepsilon_t^h \\ \varepsilon_t^u \end{bmatrix}$$

(follows from Stock/Watson's common trends representation of cointegrated systems) Economic common sense: $\psi_{12} = 0$, $\psi_{13} = 0$: Stock prices do not affect exchange rate.

Cointegration implies $\psi_{22} = \psi_{32}$ and $\psi_{23} = \psi_{33}$.

Hasbrouck (1995): Defines the information share of a market as its contribution to the variance of the permanent component of a given price series

$$\text{Var}(\text{perm. impact on exchange rate}) = \psi_{11}^2 \text{Var}(\varepsilon_t^e) + \psi_{12}^2 \text{Var}(\varepsilon_t^h) + \psi_{13}^2 \text{Var}(\varepsilon_t^u)$$

(neglecting contemporaneous correlations)

$$\frac{\psi_{11}^2 \text{Var}(\varepsilon_t^e)}{\psi_{11}^2 \text{Var}(\varepsilon_t^e) + \psi_{12}^2 \text{Var}(\varepsilon_t^h) + \psi_{13}^2 \text{Var}(\varepsilon_t^u)} \equiv \text{Information Share}$$

Hypothesized $\psi_{12} = 0, \psi_{13} = 0$

100% of relevant information is generated in exchange rate series itself (Empirically testable) Information shares for home market and US market? "Winner market takes all"-hypothesis: One market dominates! Sofianos' "home market hypothesis".

The empirical analysis is based on high frequency data for three NYSE traded German stocks and US/Euro exchange rate data

XETRA (electronic trading system of German Stock Exchange) and NYSE (TAQ) bid-ask prices for SAP, Deutsche Telekom (DT) and DaimlerChrysler (DCX).

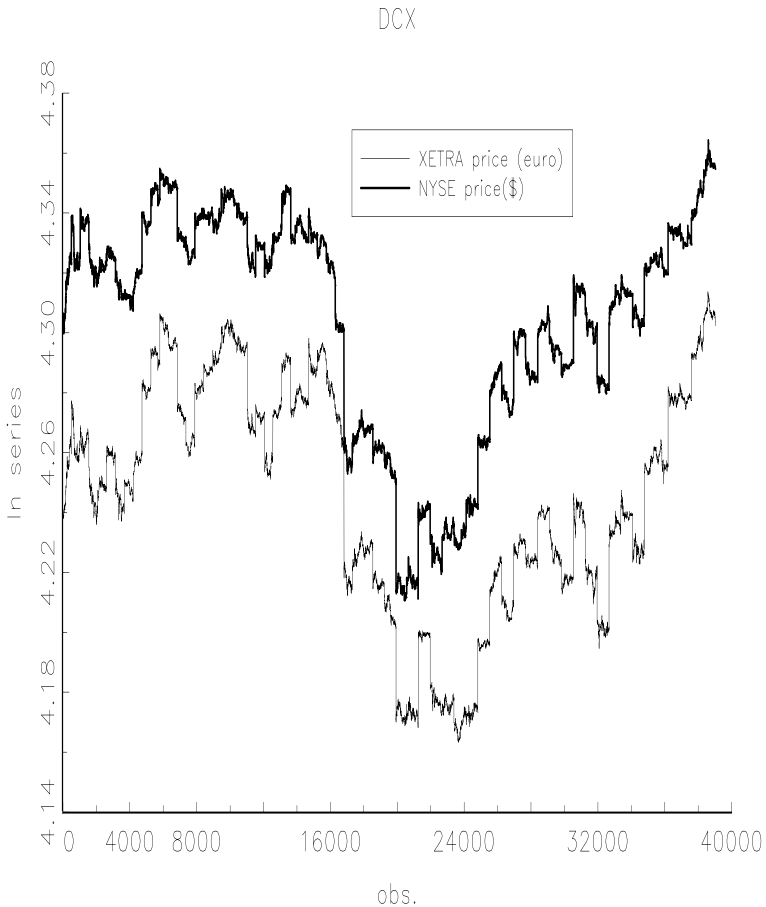
US/Euro indicative quotes: Olsen & Associates, Zürich

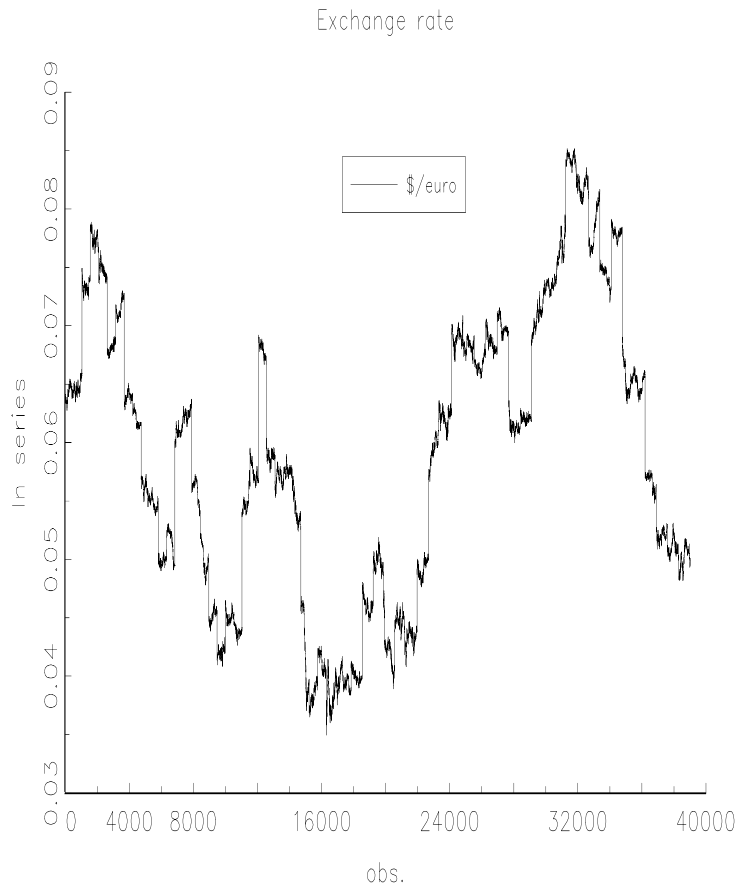
August-Oktober 1999

Mid-quotes from overlapping trading period NYSE-XETRA [GMT 14:30-16(:30)]

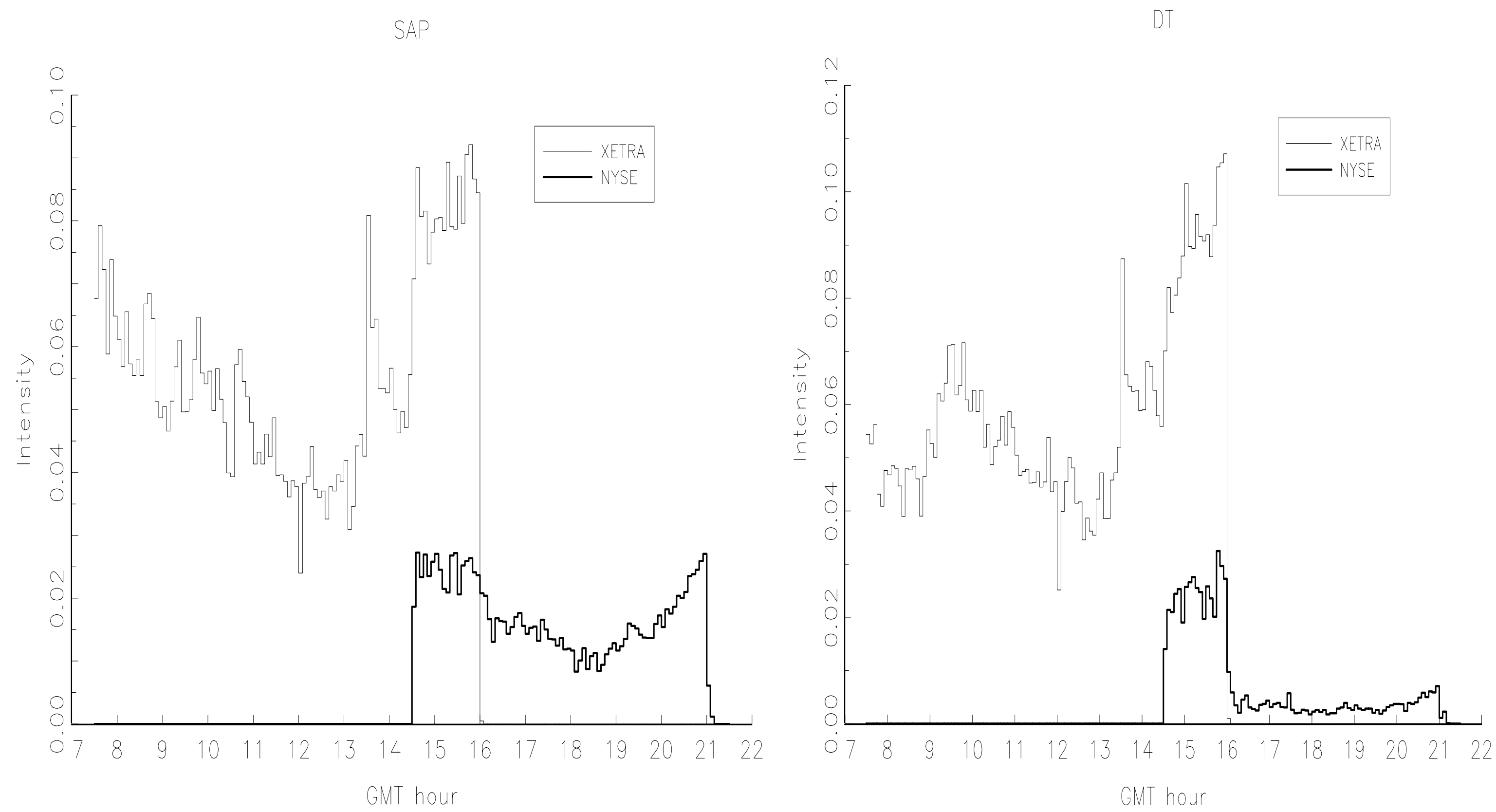
Equally spaced 10 seconds data generated from transactions data.

A look at the data





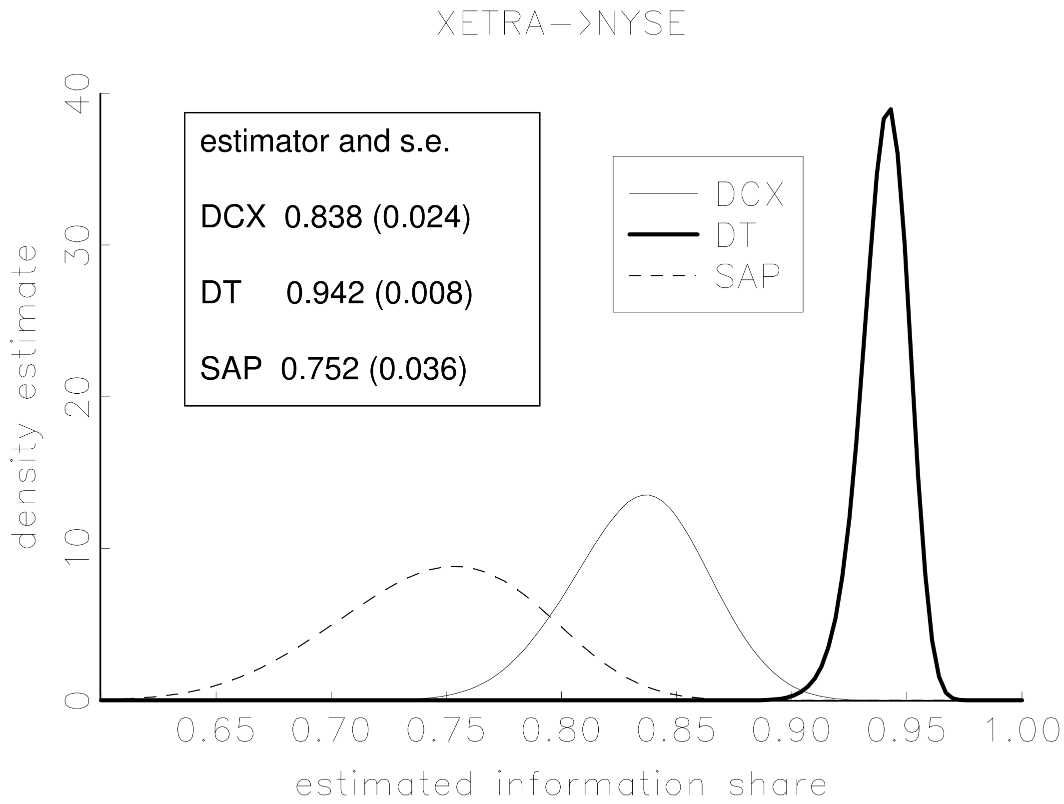
Comparing Deutsche Telekom and SAP one finds significant differences in intra day quoting intensity patterns



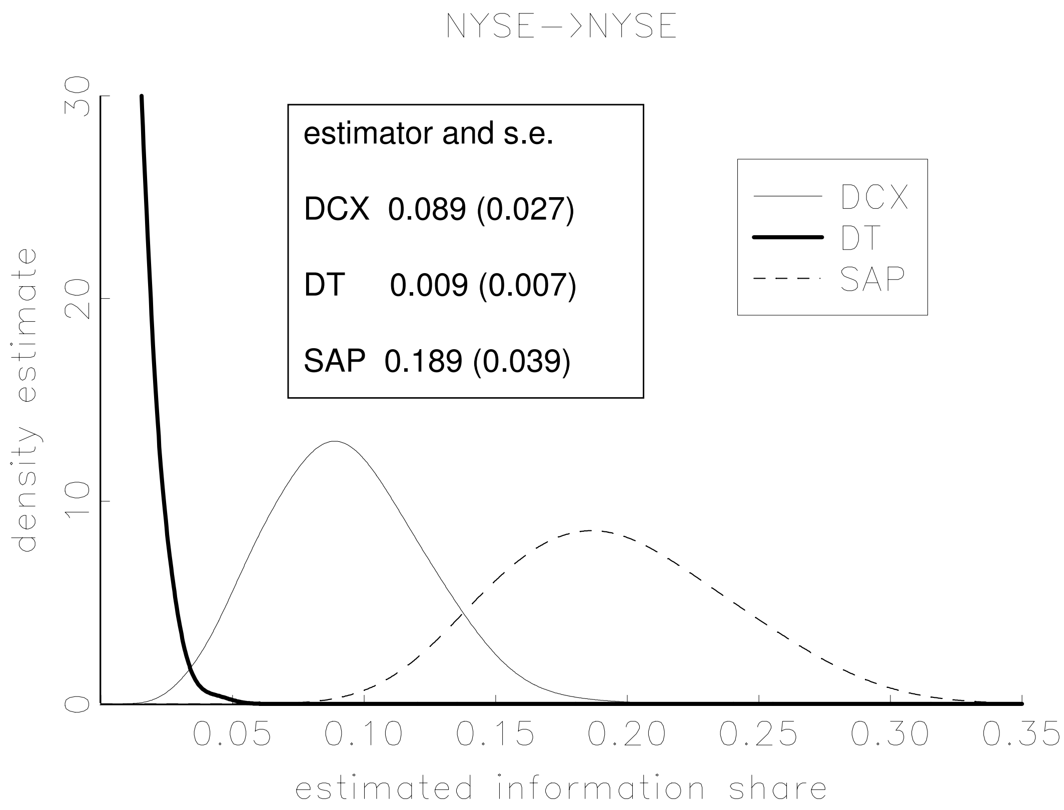
The empirical results

- ♢ Johansen's method confirms the existence of ONE cointegrating relation between stock prices and exchange rate.
- ♢ Implies two stochastic trends (efficient stock price and exchange rate).
- ♢ As expected, no permanent impact of stock prices on exchange rates.
- ♢ Only the US price incorporates exchange rate shocks. The home market does not react. Unexpected (?) asymmetric effect.
- ♢ Support for "winner market takes all"-hypothesis.
- ♢ Support for home market hypothesis, but qualitative differences are obvious: Deutsche Telekom as "national" stock: Price discovery exclusively in Germany DaimlerChrysler: The larger information share is generated in the German market SAP ("New Economy", significant US-sales): Largest US information share

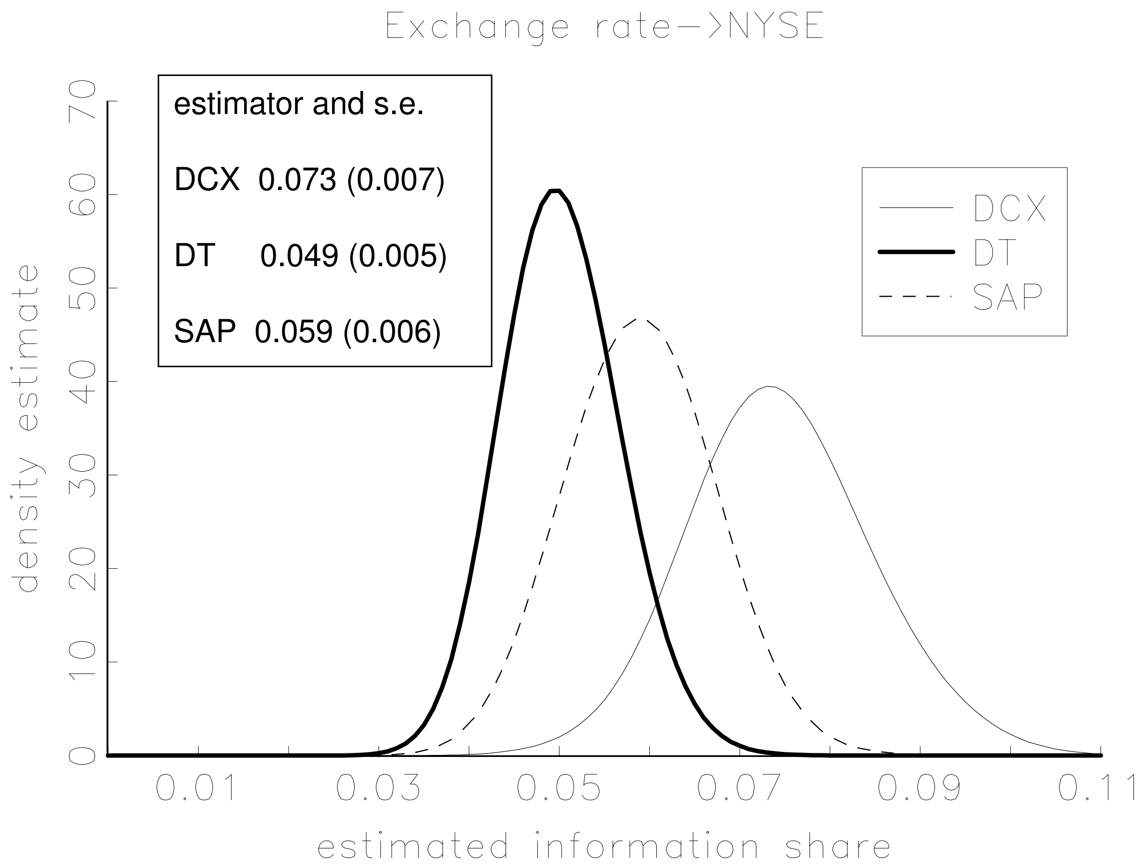
Information of share XETRA innovations w.r.t NYSE price [Kernel density estimates based on 1000 Bootstrap replications (Li/Maddala, 1997)]



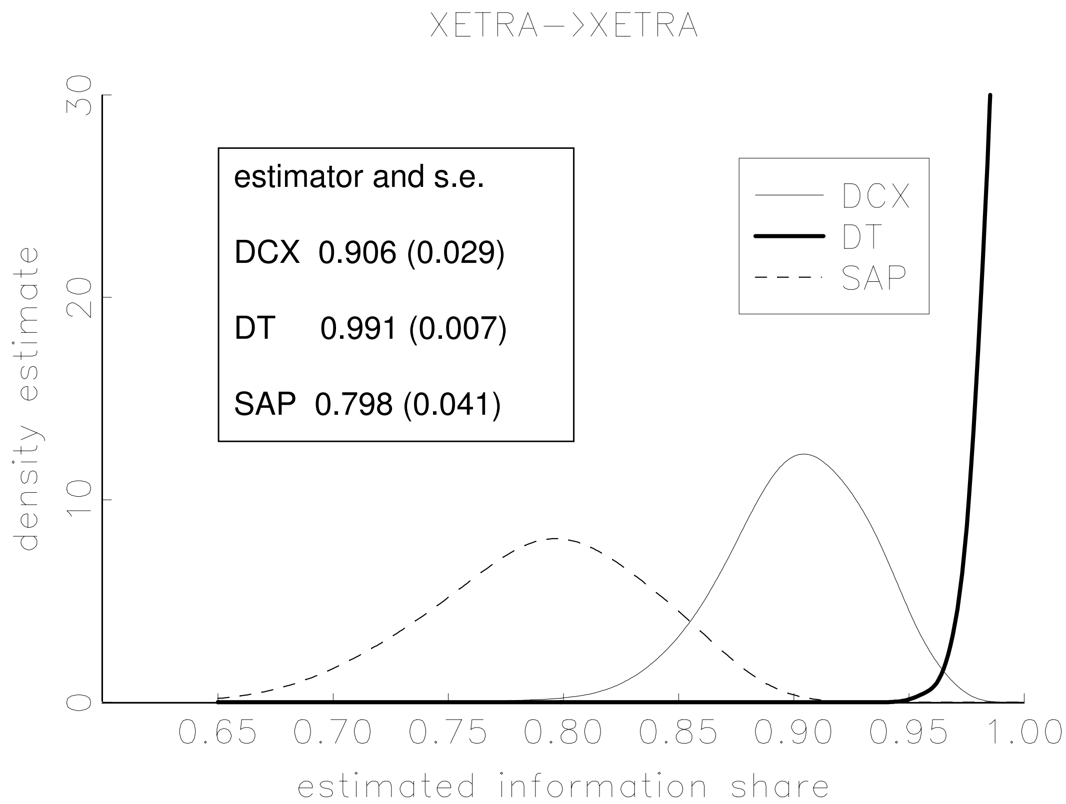
Information share of NYSE innovations w.r.t. NYSE price



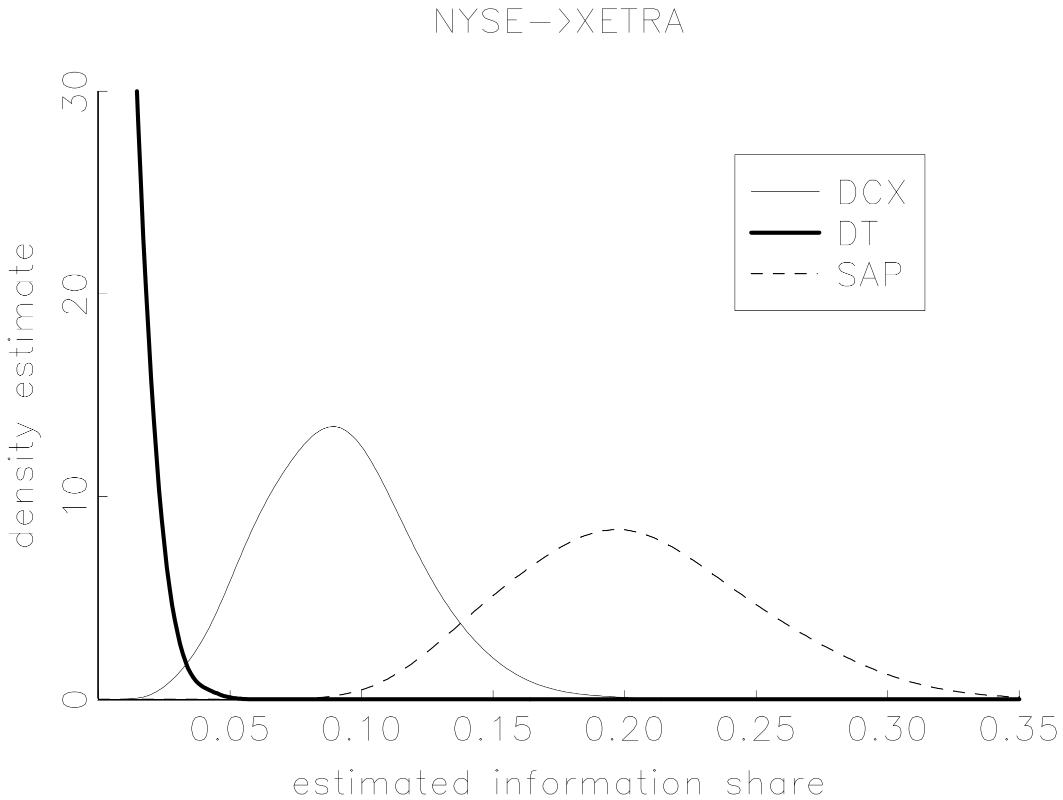
Information share of exchange rate innovations w.r.t NYSE price



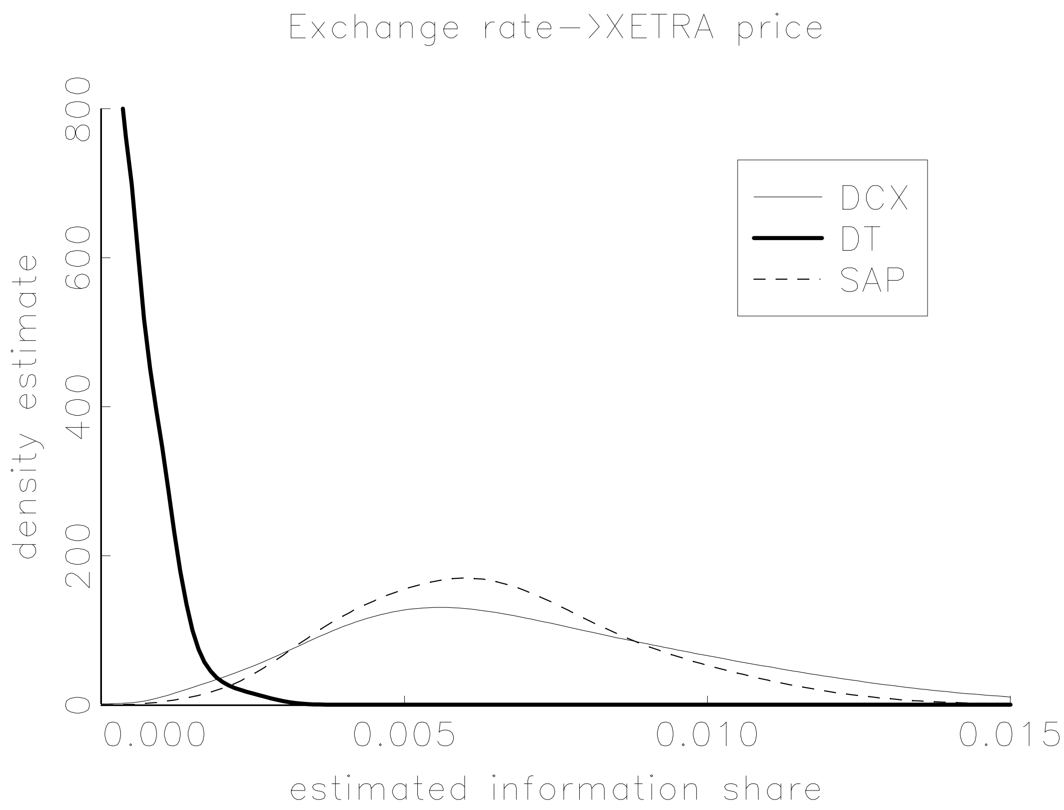
Information share of XETRA innovations w.r.t XETRA price



Information share of NYSE innovations w.r.t XETRA price



Information share of exchange rate innovations w.r.t XETRA price



Summary

◇ One cointegrating relation between exchange rate and \$ and €-prices found in high frequency data.

◇ Asymmetric price reactions in response to exchange rate shocks.

◇ Support for "winner market takes all"-hypothesis: One market dominates price discovery.

◇ Support for home market hypothesis.

◇ Qualitative differences between stocks. Truly national stocks vs. stocks with larger international focus.

◇ DaimlerChrysler: Takeover or merger among equals?

The following quote from "A Blueprint for Success", TSE, October 1998, illustrates the competitive threat from U.S. exchanges perceived by the non-U.S. exchanges.

"The TSE cannot afford to have the U.S. markets become the price discovery mechanism for Canadian interlisted stocks."

$$\begin{pmatrix} \text{permanent impact on exchange rate} \\ \text{permanent impact on -€Price} \\ \text{permanent impact on \$-Price} \end{pmatrix} = \begin{bmatrix} \psi_{11} & \psi_{12} & \psi_{13} \\ \psi_{21} & \psi_{22} & \psi_{23} \\ \psi_{31} & \psi_{32} & \psi_{33} \end{bmatrix} \times \begin{bmatrix} \varepsilon_t^e \\ \varepsilon_t^h \\ \varepsilon_t^u \end{bmatrix}$$

$$\textbf{DCX} \begin{bmatrix} 0.567\ (0.010) & 0.005\ (0.011) & 0.011\ (0.012) \\ -0.132\ (0.025) & 0.822\ (0.031) & 0.250\ (0.033) \\ 0.435\ (0.027) & 0.818\ (0.032) & 0.261\ (0.034) \end{bmatrix}$$

$$\textbf{DT} \begin{bmatrix} 0.594\ (0.006) & 0.004\ (0.007) & 0.004\ (0.008) \\ -0.046\ (0.026) & 0.879\ (0.030) & 0.081\ (0.031) \\ 0.539\ (0.027) & 0.875\ (0.030) & 0.085\ (0.031) \end{bmatrix}$$

$$\textbf{SAP} \begin{bmatrix} 0.596\ (0.007) & 0.005\ (0.008) & 0.001\ (0.008) \\ -0.149\ (0.021) & 0.689\ (0.024) & 0.287\ (0.026) \\ 0.444\ (0.023) & 0.685\ (0.025) & 0.288\ (0.026) \end{bmatrix}$$

VII. Structural Vector Autoregressive Models (Advanced)

[Hamilton (1994), Chapter 11; Hayashi (2000), Chapter 6.3 / 6.4]

To analyze the interdependence of three East Asian stock markets, (Tokyo, Singapore and South Korea) we set up a Structural VAR (SVAR)

$$
\begin{aligned}
r_t^T &= k^T && && +\beta_{12}^{(0)} r_t^S + && \beta_{13}^{(0)} r_t^K + && \beta_{11}^{(1)} r_{t-1}^T + \beta_{12}^{(1)} r_{t-1}^S + \beta_{13}^{(1)} r_{t-1}^K + u_t^T \\
r_t^S &= k^S && +\beta_{21}^{(0)} r_t^T && + && \beta_{23}^{(0)} r_t^K + && \beta_{21}^{(1)} r_{t-1}^T + \beta_{22}^{(1)} r_{t-1}^S + \beta_{23}^{(1)} r_{t-1}^K + u_t^S \\
r_t^K &= k^K && +\beta_{31}^{(0)} r_t^T && +\beta_{32}^{(0)} r_t^S + && && \beta_{31}^{(1)} r_{t-1}^T + \beta_{32}^{(1)} r_{t-1}^S + \beta_{33}^{(1)} r_{t-1}^K + u_t^K
\end{aligned}
$$

$$
\underset{(3\times 1)}{\mathbf{y}_t} = \begin{bmatrix} r_t^T \\ r_t^S \\ r_t^K \end{bmatrix} \quad \underset{(3\times 1)}{\mathbf{k}} = \begin{bmatrix} k^T \\ k^S \\ k^K \end{bmatrix} \quad \underset{(3\times 3)}{\mathbf{B_0}} = \begin{bmatrix} 1 & -\beta_{12}^{(0)} & -\beta_{13}^{(0)} \\ -\beta_{21}^{(0)} & 1 & -\beta_{23}^{(0)} \\ -\beta_{31}^{(0)} & -\beta_{32}^{(0)} & 1 \end{bmatrix} \quad \underset{(3\times 1)}{\mathbf{u}_t} = \begin{bmatrix} u_t^T \\ u_t^S \\ u_t^K \end{bmatrix}
$$

$$
\mathbf{B}_0 \mathbf{y}_t = \mathbf{k} + \mathbf{B}_1 \mathbf{y}_{t-1} + \mathbf{u}_t
$$

The innovations of a VAR in primitive form are assumed to be both serially and cross-sectionally uncorrelated (orthogonal/pure/idiosyncratic innovations/shocks)

$$\mathbf{B}_0\mathbf{y}_t = \mathbf{k} + \mathbf{B}_1\mathbf{y}_{t-1} + \mathbf{B}_2\mathbf{y}_{t-2} + \ldots + \mathbf{B}_p\mathbf{y}_{t-p} + \mathbf{u}_t$$

$$\begin{aligned} \mathbb{E}(\mathbf{u}_t) &= \mathbf{0} \\ \mathbb{E}(\mathbf{u}_t\mathbf{u}_\tau') &= \begin{cases} \mathbf{D} & \text{for} \quad t = \tau \\ \mathbf{0} & \text{otherwise.} \end{cases} \end{aligned}$$

$\mathbf{D}$ diagonal matrix

Writing the VAR in standard form „solves“ the system

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \ldots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

$\mathbf{c} = \mathbf{B}_0^{-1}\mathbf{k}$ $(n \times 1)$ vector of constants

$\mathbf{\Phi}_s = \mathbf{B}_0^{-1}\mathbf{B}_s$ $(n \times n)$ matrix of AR coefficients for $s = 1, ..., p$

$\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\mathbf{u}_t$ $(n \times 1)$ vector generalization of White Noise.

The innovations of a VAR in standard form are, by construction, contemporaneously correlated (composite innovations/shocks)

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \ldots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

$$\begin{aligned}
\mathbb{E}(\boldsymbol{\varepsilon}_t) &= \mathbb{E}(\mathbf{B}_0^{-1}\mathbf{u}_t) = \mathbf{B}_0^{-1}\mathbb{E}(\mathbf{u}_t) = \mathbf{0} \\
\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') &= \mathbb{E}(\mathbf{B}_0^{-1}\mathbf{u}_t\mathbf{u}_t'[\mathbf{B}_0^{-1}]') \equiv \mathbf{\Omega} \\
\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_\tau') &= \begin{cases} \mathbf{\Omega} & \text{for} \quad t = \tau \\ \mathbf{0} & \text{otherwise.} \end{cases}
\end{aligned}$$

The lag operator provides notational convenience

Lag operator:

$$L(y_t) = y_{t-1},\ L^2(y_t) = y_{t-2},\ \ldots$$

VAR(p) written with lag operator

$$[\mathbf{I}_n - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - \ldots - \mathbf{\Phi}_p L^p]\mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t$$

or

$$\mathbf{\Phi}(L)\mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t$$

We take expectations of the endogenous variables

Assuming stationarity: $\mathbb{E}(\mathbf{y}_t) = \boldsymbol{\mu}$

$$\begin{aligned}
\mathbb{E}(\mathbf{y}_t) &= \mathbf{c} + \mathbf{\Phi}_1 \mathbb{E}(\mathbf{y}_{t-1}) + \ldots + \mathbf{\Phi}_p \mathbb{E}(\mathbf{y}_{t-p}) + \mathbb{E}(\boldsymbol{\varepsilon}_t) \\
\boldsymbol{\mu} &= \mathbf{c} + \mathbf{\Phi}_1 \boldsymbol{\mu} + \mathbf{\Phi}_2 \boldsymbol{\mu} + \ldots + \mathbf{\Phi}_p \boldsymbol{\mu} \\
\boldsymbol{\mu} &= \mathbf{c} + [\mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \ldots + \mathbf{\Phi}_p] \boldsymbol{\mu} \\
[\mathbf{I}_n - \mathbf{\Phi}_1 - \mathbf{\Phi}_2 - \ldots - \mathbf{\Phi}_p]\, \boldsymbol{\mu} &= \mathbf{c} \\
[\mathbf{I}_n - \mathbf{\Phi}_1 L - \ldots - \mathbf{\Phi}_p L^p]\, \boldsymbol{\mu} &= \mathbf{c} \\
\mathbf{\Phi}(L) \boldsymbol{\mu} &= \mathbf{c}
\end{aligned}$$

It is convenient to express a VAR in terms of deviations from the means

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \ldots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \\
\mathbf{\Phi}(L)\boldsymbol{\mu} &= \mathbf{c} \\
(\mathbf{y}_t - \boldsymbol{\mu}) &= \mathbf{\Phi}_1(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \mathbf{\Phi}_2(\mathbf{y}_{t-2} - \boldsymbol{\mu}) + \ldots + \mathbf{\Phi}_p(\mathbf{y}_{t-p} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t
\end{aligned}$$

With some additional notation a VAR(p) can be rewritten as a VAR(1)

$$(\mathbf{y}_t - \boldsymbol{\mu}) = \mathbf{\Phi}_1(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \mathbf{\Phi}_2(\mathbf{y}_{t-2} - \boldsymbol{\mu}) + \ldots + \mathbf{\Phi}_p(\mathbf{y}_{t-p} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t$$

Define:

$$\underset{(np\times 1)}{\boldsymbol{\xi}_t} \equiv \begin{bmatrix} \mathbf{y}_t - \boldsymbol{\mu} \\ \mathbf{y}_{t-1} - \boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-p+1} - \boldsymbol{\mu} \end{bmatrix} \quad \underset{(np\times np)}{\mathbf{F}} \equiv \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \mathbf{\Phi}_3 & \ldots & \mathbf{\Phi}_{p-1} & \mathbf{\Phi}_p \\ \mathbf{I}_n & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_n & \mathbf{0} \end{bmatrix} \quad \underset{(np\times 1)}{\mathbf{v}_t} \equiv \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

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

Consider a forward iteration of the VAR(1) system

$$
\begin{array}{lllll}
\boldsymbol{\xi}_t & = & \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t & & \\
\boldsymbol{\xi}_{t+1} & = & \mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1} & & \\
\boldsymbol{\xi}_{t+2} & = & \mathbf{F}\boldsymbol{\xi}_{t+1} + \mathbf{v}_{t+2} & & \\
\boldsymbol{\xi}_{t+3} & = & \mathbf{F}\boldsymbol{\xi}_{t+2} + \mathbf{v}_{t+3} & = & \mathbf{v}_{t+3} + \mathbf{F}(\mathbf{F}\boldsymbol{\xi}_{t+1} + \mathbf{v}_{t+2}) \\
\vdots & & & = & \mathbf{v}_{t+3} + \mathbf{F}\mathbf{v}_{t+2} + \mathbf{F}^2\boldsymbol{\xi}_{t+1} \\
 & & & = & \mathbf{v}_{t+3} + \mathbf{F}\mathbf{v}_{t+2} + \mathbf{F}^2(\mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1}) \\
 & & & = & \mathbf{v}_{t+3} + \mathbf{F}\mathbf{v}_{t+2} + \mathbf{F}^2\mathbf{v}_{t+1} + \mathbf{F}^3\boldsymbol{\xi}_t
\end{array}
$$

iterating s times yields:

$$
\boldsymbol{\xi}_{t+s} = \mathbf{v}_{t+s} + \mathbf{F}\mathbf{v}_{t+s-1} + \mathbf{F}^2\mathbf{v}_{t+s-2} + \ldots + \mathbf{F}^{s-1}\mathbf{v}_{t+1} + \mathbf{F}^s\boldsymbol{\xi}_t
$$

To obtain the Vector Moving Average (VMA) representation we focus on the first rows of the system

the first n rows of the system

$$\boldsymbol{\xi}_{t+s} = \mathbf{v}_{t+s} + \mathbf{F}\mathbf{v}_{t+s-1} + \mathbf{F}^2\mathbf{v}_{t+s-2} + \ldots + \mathbf{F}^{s-1}\mathbf{v}_{t+1} + \mathbf{F}^s\boldsymbol{\xi}_t$$

are:

$$\begin{aligned}\mathbf{y}_{t+s} &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_{t+s} + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t+s-1} + \boldsymbol{\Psi}_2\boldsymbol{\varepsilon}_{t+s-2} + \ldots + \boldsymbol{\Psi}_{s-1}\boldsymbol{\varepsilon}_{t+1} \\ &+ \mathbf{F}_{11}^{(s)}(\mathbf{y}_t - \boldsymbol{\mu}) + \mathbf{F}_{12}^{(s)}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \ldots + \mathbf{F}_{1p}^{(s)}(\mathbf{y}_{t-p+1} - \boldsymbol{\mu})\end{aligned}$$

$\mathbf{F}^{(j)}$: $\mathbf{F}$ raised to the j^{th} power
$\mathbf{F}_{11}^{(j)} = \boldsymbol{\Psi}_j$: first n rows and columns 1 through n
$\mathbf{F}_{1p}^{(j)}$: first n rows and columns $(n(p-1)+1)$ through np

Forecast of $\mathbf{y}_{t+s}$ on the basis of $\mathbf{y}_t, \mathbf{y}_{t-1}, ...$

$$\hat{\mathbf{y}}_{t+s|t} = \boldsymbol{\mu} + \mathbf{F}_{11}^{(s)}(\mathbf{y}_t - \boldsymbol{\mu}) + \mathbf{F}_{12}^{(s)}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \ldots + \mathbf{F}_{1p}^{(s)}(\mathbf{y}_{t-p+1} - \boldsymbol{\mu})$$

Forecast error:

$$\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t} = \boldsymbol{\varepsilon}_{t+s} + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t+s-1} + \boldsymbol{\Psi}_2\boldsymbol{\varepsilon}_{t+s-2} + \ldots + \boldsymbol{\Psi}_{s-1}\boldsymbol{\varepsilon}_{t+1}$$

Vector MA(∞) Representation

Eigenvalues of $\mathbf{F}$ inside the unit circle $\Rightarrow$ stationarity of $\{\mathbf{y}_t\}$

$\Rightarrow$ Vector MA(∞) Representation

$$\boldsymbol{\xi}_t = \sum_{i=0}^{\infty} \mathbf{F}^i \mathbf{v}_{t-i}$$

First n rows:

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \boldsymbol{\Psi}_3 \boldsymbol{\varepsilon}_{t-3} + \ldots \\
\mathbf{y}_t &= \boldsymbol{\mu} + \left[\mathbf{I}_n + \boldsymbol{\Psi}_1 L + \boldsymbol{\Psi}_2 L^2 \ldots\right] \boldsymbol{\varepsilon}_t \\
\mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\Psi}(L) \boldsymbol{\varepsilon}_t
\end{aligned}$$

Combining results shows how VAR and MA coefficients are related

$$\mathbf{\Phi}(L)\mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t \qquad \mathbf{\Phi}(L)\boldsymbol{\mu} = \mathbf{c} \qquad \mathbf{y}_t = \boldsymbol{\mu} + \mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t$$

$$\begin{aligned}
\mathbf{\Phi}(L)[\boldsymbol{\mu} + \mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t] &= \mathbf{c} + \boldsymbol{\varepsilon}_t \\
\mathbf{\Phi}(L)\boldsymbol{\mu} + \mathbf{\Phi}(L)\mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t &= \mathbf{c} + \boldsymbol{\varepsilon}_t \\
\mathbf{c} + \mathbf{\Phi}(L)\mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t &= \mathbf{c} + \boldsymbol{\varepsilon}_t \\
\underbrace{[\mathbf{\Phi}(L)\mathbf{\Psi}(L)]}_{\mathbf{I}_n}\boldsymbol{\varepsilon}_t &= \boldsymbol{\varepsilon}_t
\end{aligned}$$

The VMA coefficients can be recursively computed from the VAR coefficients

$$
\begin{aligned}
\mathbf{I}_n &= \mathbf{\Psi}(L)\mathbf{\Phi}(L) \\
\mathbf{I}_n &= (\mathbf{I}_n + \mathbf{\Psi}_1 L + \mathbf{\Psi}_2 L^2 + \ldots)(\mathbf{I}_n - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - \ldots - \mathbf{\Phi}_p L^p) \\
\mathbf{I}_n &= \mathbf{I}_n + (\mathbf{\Psi}_1 - \mathbf{\Phi}_1)L + (\mathbf{\Psi}_2 - \mathbf{\Phi}_1\mathbf{\Psi}_1 - \mathbf{\Phi}_2)L^2 + \ldots
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \qquad \mathbf{\Psi_1} &= \mathbf{\Phi_1} \\
\mathbf{\Psi_2} &= \mathbf{\Phi_1}\mathbf{\Psi_1} + \mathbf{\Phi_2}
\end{aligned}
$$

general for L^s $s = 1, 2, \ldots :$ $\qquad \mathbf{\Psi_s} = \mathbf{\Phi_1}\mathbf{\Psi_{s-1}} + \mathbf{\Phi_2}\mathbf{\Psi_{s-2}} + \ldots + \mathbf{\Phi_p}\mathbf{\Psi_{s-p}}$

The Impulse-Response Function gives the response of the system to one unit shocks in the ε

$$
\begin{aligned}
\mathbf{y}_{t+s} &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_{t+s} + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t+s-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t+s-2} + \ldots + \boldsymbol{\Psi}_s \boldsymbol{\varepsilon}_t + \ldots + \ldots \\
\frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}'_t} &= \boldsymbol{\Psi}_{\mathbf{s}}
\end{aligned}
$$

Sequence of $\boldsymbol{\Psi}_1$, $\boldsymbol{\Psi}_2$,...: Impulse-Response Function

e.g. response of $y_{i,t+s}$ to a one-time impulse in $\boldsymbol{\varepsilon}_{j,t}$ with all other variables dated t or earlier held constant: $\frac{\partial y_{i,t+s}}{\partial \boldsymbol{\varepsilon}_{jt}} = \psi_s[i,j]$

This numerical example shows how to obtain the VMA coefficients from VAR(2) parameters

s		$\mathbf{\Phi}_s$			$\mathbf{\Psi}_s$	
	-0.029	0.034	0.035	-0.029	0.034	0.035
1	0.007	0.195	0.044	0.007	0.195	0.044
	0.027	0.090	0.060	0.027	0.090	0.060
	-0.071	-0.024	0.020	-0.069	-0.015	0.022
2	-0.050	-0.062	0.016	-0.047	-0.020	0.028
	0.005	-0.016	0.004	0.006	0.008	0.013
				0.003	-0.005	-0.002
3				-0.008	-0.016	0.003
				-0.006	-0.004	0.004
⋮		⋮			⋮	
				0.000	0.000	0.000
10				0.000	0.000	0.000
				0.000	0.000	0.000

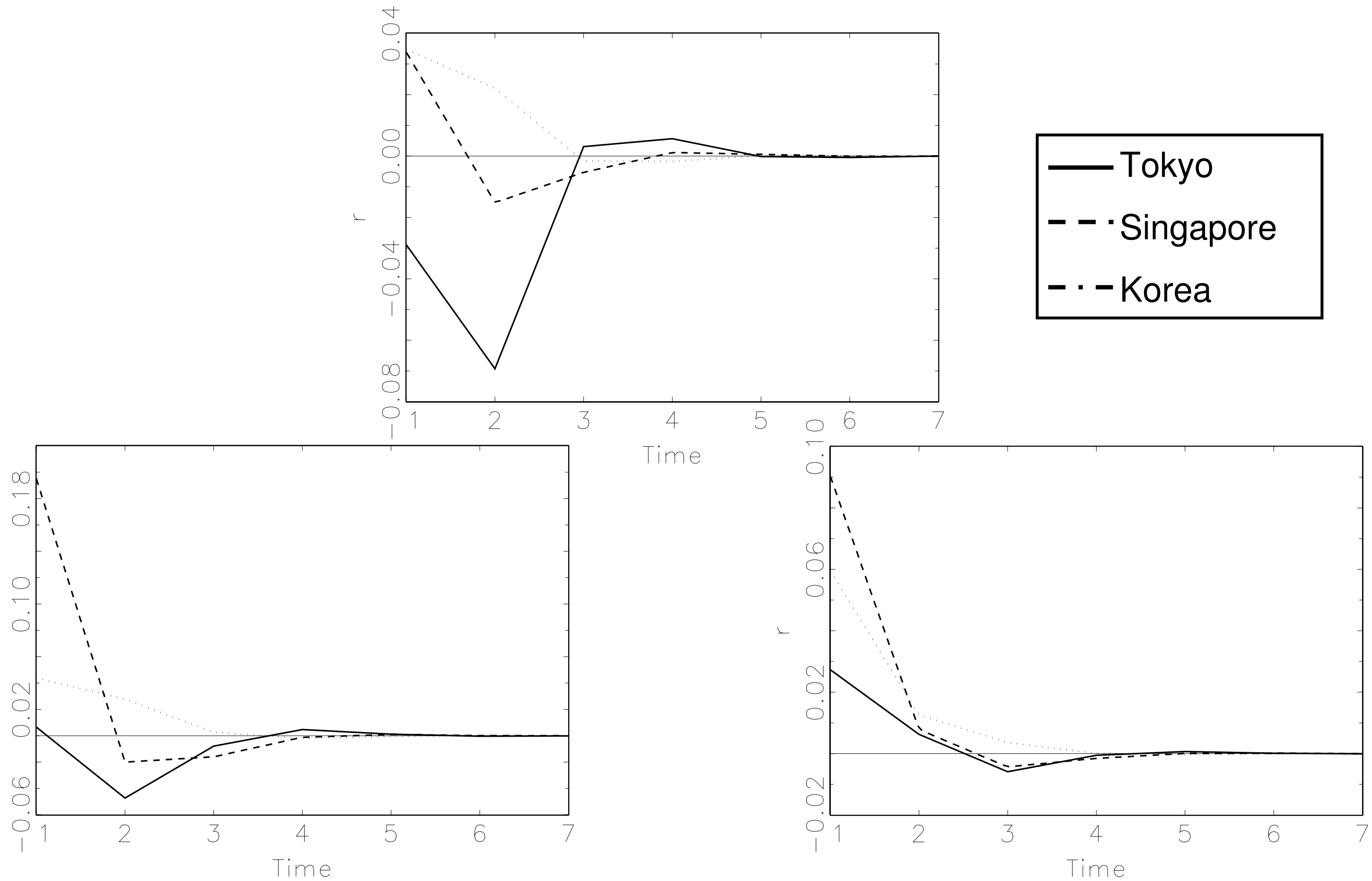
This numerical example shows how to obtain the VMA coefficients from the VAR(2) parameters

$$
\begin{aligned}
\boldsymbol{\Psi}_1 &= \boldsymbol{\Phi}_1 \\
&= \begin{bmatrix} -0.029 & 0.034 & 0.035 \\ 0.007 & 0.195 & 0.044 \\ 0.027 & 0.090 & 0.060 \end{bmatrix} \\
\boldsymbol{\Psi}_2 &= \boldsymbol{\Phi_1}\boldsymbol{\Psi_1} + \boldsymbol{\Phi_2} \\
&= \begin{bmatrix} -0.029 & 0.034 & 0.035 \\ 0.007 & 0.195 & 0.044 \\ 0.027 & 0.090 & 0.060 \end{bmatrix} \cdot \begin{bmatrix} -0.029 & 0.034 & 0.035 \\ 0.007 & 0.195 & 0.044 \\ 0.027 & 0.090 & 0.060 \end{bmatrix} \\
&+ \begin{bmatrix} -0.071 & -0.024 & 0.020 \\ -0.050 & -0.062 & 0.016 \\ 0.005 & -0.016 & 0.004 \end{bmatrix} = \underline{\underline{\begin{bmatrix} -0.069 & -0.015 & 0.022 \\ -0.047 & -0.020 & 0.028 \\ 0.006 & 0.008 & 0.013 \end{bmatrix}}}
\end{aligned}
$$

This numerical example shows how to obtain the VMA coefficients from the VAR(2) parameters

$$
\begin{aligned}
\Psi_3 &= \boldsymbol{\Phi_1 \Psi_2 + \Phi_2 \Psi_1} \\
&= \begin{bmatrix} -0.029 & 0.034 & 0.035 \\ 0.007 & 0.195 & 0.044 \\ 0.027 & 0.090 & 0.060 \end{bmatrix} \cdot \begin{bmatrix} -0.069 & -0.015 & 0.022 \\ -0.047 & -0.020 & 0.028 \\ 0.006 & 0.008 & 0.013 \end{bmatrix} \\
&+ \begin{bmatrix} -0.071 & -0.024 & 0.020 \\ -0.050 & -0.062 & 0.016 \\ 0.005 & -0.016 & 0.004 \end{bmatrix} \cdot \begin{bmatrix} -0.029 & 0.034 & 0.035 \\ 0.007 & 0.195 & 0.044 \\ 0.027 & 0.090 & 0.060 \end{bmatrix} \\
&= \begin{bmatrix} 0.003 & -0.005 & -0.002 \\ -0.008 & -0.016 & 0.003 \\ -0.006 & -0.004 & 0.004 \end{bmatrix}
\end{aligned}
$$

The plots show a graphical representation of the VMA coefficients



To obtain the idiosyncratic shocks from the composite shocks we need the structural parameters, the matrix B_0

covariance matrix of $\boldsymbol{\varepsilon}_t$:

$$\mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega}$$

relation between shocks in VAR and SVAR: $\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\mathbf{u}_t$

$$\begin{aligned}\mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') &= \mathbf{B}_0^{-1}\mathbb{E}(\mathbf{u}_t\mathbf{u}_t')\left[\mathbf{B}_0^{-1}\right]' \\ &= \mathbf{B}_0^{-1}\mathbf{D}\left[\mathbf{B}_0^{-1}\right]'\end{aligned}$$

To identify the structural parameters B_0, we decompose the variance covariance matrix of composite innovations (Cholesky-Decomposition)

$$\begin{aligned}\mathbf{\Omega} &= \mathbf{ADA}' \\ &= \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{21} & 1 & 0 & \dots & 0 \\ a_{31} & a_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & 1 \end{bmatrix} \cdot \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & d_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & d_n \end{bmatrix} \cdot \begin{bmatrix} 1 & a_{21} & a_{31} & \dots & a_{n1} \\ 0 & 1 & a_{32} & \dots & a_{n2} \\ 0 & 0 & 1 & \dots & a_{n3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}\end{aligned}$$

$\mathbf{\Omega}$: real symmetric positive definite matrix
$\mathbf{A}$: lower triangular matrix with ones along the principal diagonal
$\mathbf{D}$: diagonal matrix with positive elements

The idiosyncratic innovations can then be backed out from the composite innovations

Define $\mathbf{A} = \mathbf{B}_0^{-1}$

$$\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega} = \mathbf{A}\mathbf{D}\mathbf{A}'$$

Construct from $\mathbf{A}\mathbf{u}_t = \boldsymbol{\varepsilon}_t$: $\mathbf{u}_t \equiv \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$ with variance

$$\begin{aligned}
\mathbb{E}(\mathbf{u}_t \mathbf{u}_t') &= [\mathbf{A}^{-1}]\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')[\mathbf{A}^{-1}]' \\
&= [\mathbf{A}^{-1}]\boldsymbol{\Omega}[\mathbf{A}']^{-1} \\
&= [\mathbf{A}^{-1}]\mathbf{A}\mathbf{D}\mathbf{A}'[\mathbf{A}']^{-1} \\
&= \mathbf{D}
\end{aligned}$$

This implies: $\mathbb{E}(\mathbf{u}_{it}\mathbf{u}_{jt}') = 0 \; i \neq j$

The numerical example shows the decomposition of the variance covariance matrix in the present application

Example:

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

$$\begin{bmatrix} 1.79 & 0.62 & 0.16 \\ 0.62 & 1.99 & 0.28 \\ 0.16 & 0.28 & 2.67 \end{bmatrix} = \begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.34 & 1.00 & 0.00 \\ 0.09 & 0.13 & 1.00 \end{bmatrix} \begin{bmatrix} 1.79 & 0.00 & 0.00 \\ 0.00 & 1.78 & 0.00 \\ 0.00 & 0.00 & 2.63 \end{bmatrix} \begin{bmatrix} 1.00 & 0.34 & 0.09 \\ 0.00 & 1.00 & 0.13 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$$

$\mathbf{\Omega}$ and $\mathbf{D}$ multiplied by 10000.

The composite shocks are generated as linear combinations of the pure innovations

$$\mathbf{A} \cdot \mathbf{u}_t = \boldsymbol{\varepsilon}_t$$

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{21} & 1 & 0 & \dots & 0 \\ a_{31} & a_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ \vdots \\ u_{nt} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varepsilon}_{1t} \\ \boldsymbol{\varepsilon}_{2t} \\ \boldsymbol{\varepsilon}_{3t} \\ \vdots \\ \boldsymbol{\varepsilon}_{nt} \end{bmatrix}$$

Thus, $u_{1t} = \boldsymbol{\varepsilon}_{1t}$ and $u_{jt} = \boldsymbol{\varepsilon}_{jt} - a_{j1}u_{1t} - a_{j2}u_{2t} - \ldots - a_{j,j-1}u_{j-1,t}$

$\Rightarrow$ variable ORDERING matters!

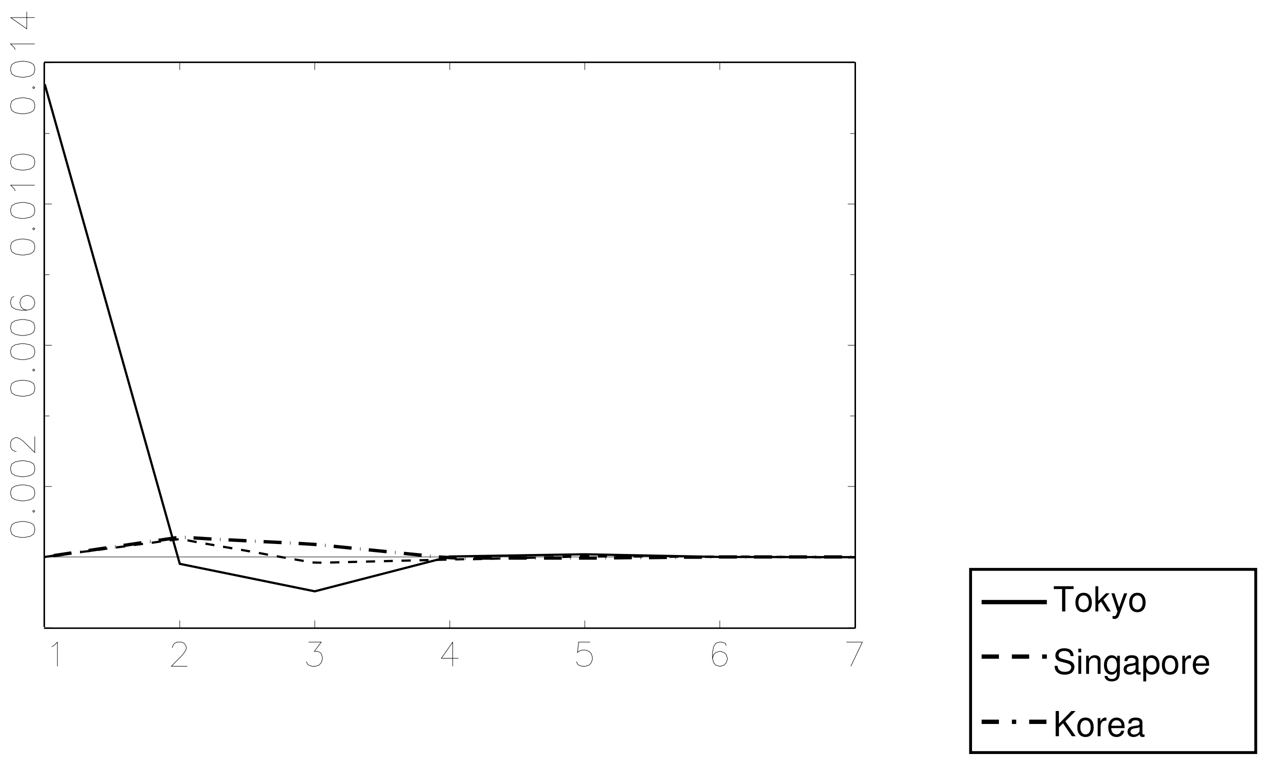
In most applications in economics and finance you want to trace a shock in the pure innovation

$$\frac{\partial \hat{\mathbb{E}}(\mathbf{y}_{t+s}|y_{jt}, y_{j-1,t}, ..., y_{1t}, \mathbf{x}_{t-1})}{\partial u_{jt}} = \mathbf{\Psi}_s \mathbf{a}_j$$

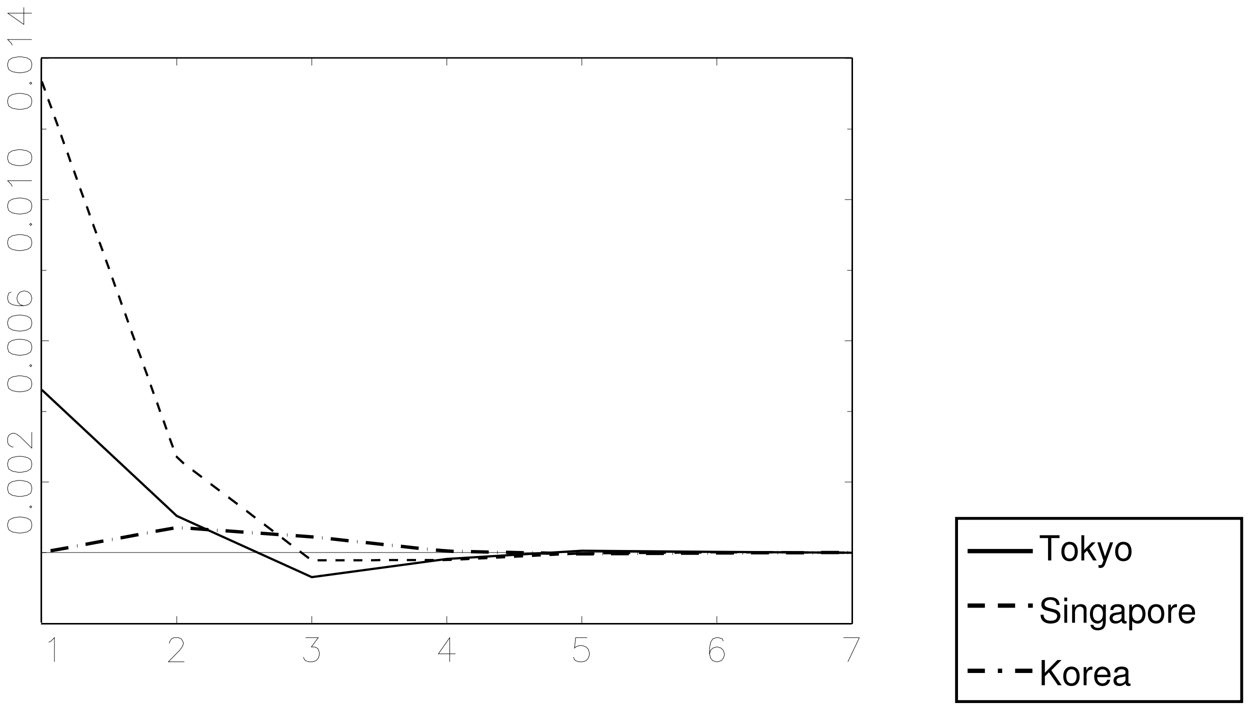
with $\mathbf{a}_j$ as the jth column of $\mathbf{A}$

$\Rightarrow$ orthogonalized impulse-response function

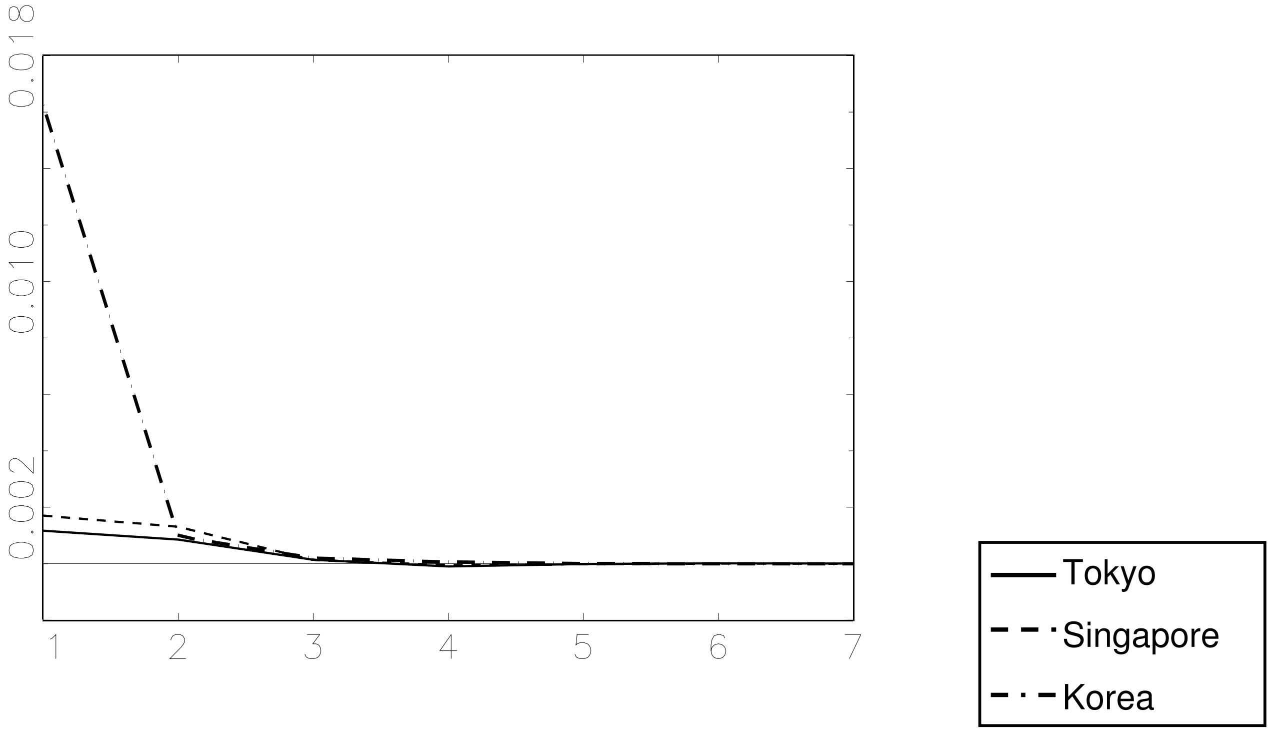
Orthogonalized impulse-response function of Tokyo to one standard deviation shock in the SVAR(2) with Cholesky Ordering: Tokyo Singapore Korea



Orthogonalized impulse-response function of Singapore to one standard deviation shock in the SVAR(2) with Cholesky Ordering: Tokyo Singapore Korea



Orthogonalized impulse-response function of Singapore to one standard deviation shock in the SVAR(2) with Cholesky Ordering: Tokyo Singapore Korea



To attribute information shares to the markets we consider a decomposition of the Mean Squared Forecast Error

$$\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t} = \boldsymbol{\varepsilon}_{t+s} + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t+s-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t+s-2} + \ldots + \boldsymbol{\Psi}_{s-1} \boldsymbol{\varepsilon}_{t+1}$$

$$\begin{aligned} MSE(\hat{\mathbf{y}}_{t+s|t}) &= \mathbb{E}[(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})'] \\ &= \boldsymbol{\Omega} + \boldsymbol{\Psi}_1 \boldsymbol{\Omega} \boldsymbol{\Psi}_1' + \boldsymbol{\Psi}_2 \boldsymbol{\Omega} \boldsymbol{\Psi}_2' + \ldots + \boldsymbol{\Psi}_{s-1} \boldsymbol{\Omega} \boldsymbol{\Psi}_{s-1}' \end{aligned}$$

The Cholesky ordering allows a decomposition of the variance of the composite innovations into the contributions of the pure innovations

$$\boldsymbol{\varepsilon}_t = \mathbf{A}\mathbf{u}_t = \mathbf{a}_1 u_{1t} + \mathbf{a}_2 u_{2t} + \ldots + \mathbf{a}_n u_{nt}$$

$$\begin{aligned} \boldsymbol{\Omega} &= \mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \mathbf{A} \cdot \mathbb{E}(\mathbf{u}_t \mathbf{u}_t') \cdot \mathbf{A}' = \mathbf{A}\mathbf{D}\mathbf{A}' \\ &= \mathbf{a}_1 \mathbf{a}_1' \cdot \mathsf{Var}(u_{1t}) + \mathbf{a}_2 \mathbf{a}_2' \cdot \mathsf{Var}(u_{2t}) + \ldots + \mathbf{a}_n \mathbf{a}_n' \cdot \mathsf{Var}(u_{nt}) \end{aligned}$$

We can also decompose the MSE of the s-step ahead forecast

$$
\begin{aligned}
MSE(\hat{\mathbf{y}}_{t+s|t}) &= \mathbb{E}[(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})'] \\
&= \mathbf{\Omega} + \mathbf{\Psi}_1\mathbf{\Omega}\mathbf{\Psi}_1' + \mathbf{\Psi}_2\mathbf{\Omega}\mathbf{\Psi}_2' + \ldots + \mathbf{\Psi}_{s-1}\mathbf{\Omega}\mathbf{\Psi}_{s-1}'
\end{aligned}
$$

$$
\begin{aligned}
\underset{(n\times n)}{MSE(\hat{\mathbf{y}}_{t+s|t})} &= \sum_{j=1}^{n}\{\mathsf{Var}(u_{jt}) \cdot [\mathbf{a}_j\mathbf{a}_j' + \mathbf{\Psi}_1\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_1' \\
&\quad + \mathbf{\Psi}_2\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_2' + \ldots + \mathbf{\Psi}_{s-1}\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_{s-1}']\}
\end{aligned}
$$

contribution of the jth orthogonalized innovation to the MSE of the s-period-ahead forecast:

$$
\mathsf{Var}(u_{jt}) \cdot [\mathbf{a}_j\mathbf{a}_j' + \mathbf{\Psi}_1\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_1' + \mathbf{\Psi}_2\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_2' + \ldots + \mathbf{\Psi}_{s-1}\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_{s-1}']
$$

The numerical example illustrates the decomposition of the variance covariance matrix of the composite shocks (MSE 1 step forecast)

$$\underset{(n\times n)}{MSE(\hat{\mathbf{y}}_{t+1|t})} = \text{Var}(u_t^T)\cdot[\mathbf{a}_1\mathbf{a}_1'] + \text{Var}(u_t^S)\cdot[\mathbf{a}_2\mathbf{a}_2'] + \text{Var}(u_t^K)\cdot[\mathbf{a}_3\mathbf{a}_3']$$

$$\begin{aligned}\underset{(n\times n)}{MSE(\hat{\mathbf{y}}_{t+1|t})} &= 1.79\cdot\begin{bmatrix}1.000 & 0.344 & 0.087\\ 0.344 & 0.119 & 0.030\\ 0.087 & 0.030 & 0.008\end{bmatrix} + 1.78\cdot\begin{bmatrix}0.000 & 0.000 & 0.000\\ 0.000 & 1.000 & 0.127\\ 0.000 & 0.127 & 0.016\end{bmatrix}\\ &+ 2.63\cdot\begin{bmatrix}0.000 & 0.000 & 0.000\\ 0.000 & 0.000 & 0.000\\ 0.000 & 0.000 & 1.000\end{bmatrix}\\ &= \begin{bmatrix}1.793 & 0.618 & 0.157\\ 0.618 & 1.994 & 0.281\\ 0.157 & 0.281 & 2.674\end{bmatrix}\end{aligned}$$

$\text{Var}(u_t^T), \text{Var}(u_t^S), \text{Var}(u_t^K)$ and $MSE(\hat{\mathbf{y}}_{t+1|t})$ taken times 10000

The numerical example illustrates the decomposition of the MSE of the two step forecast

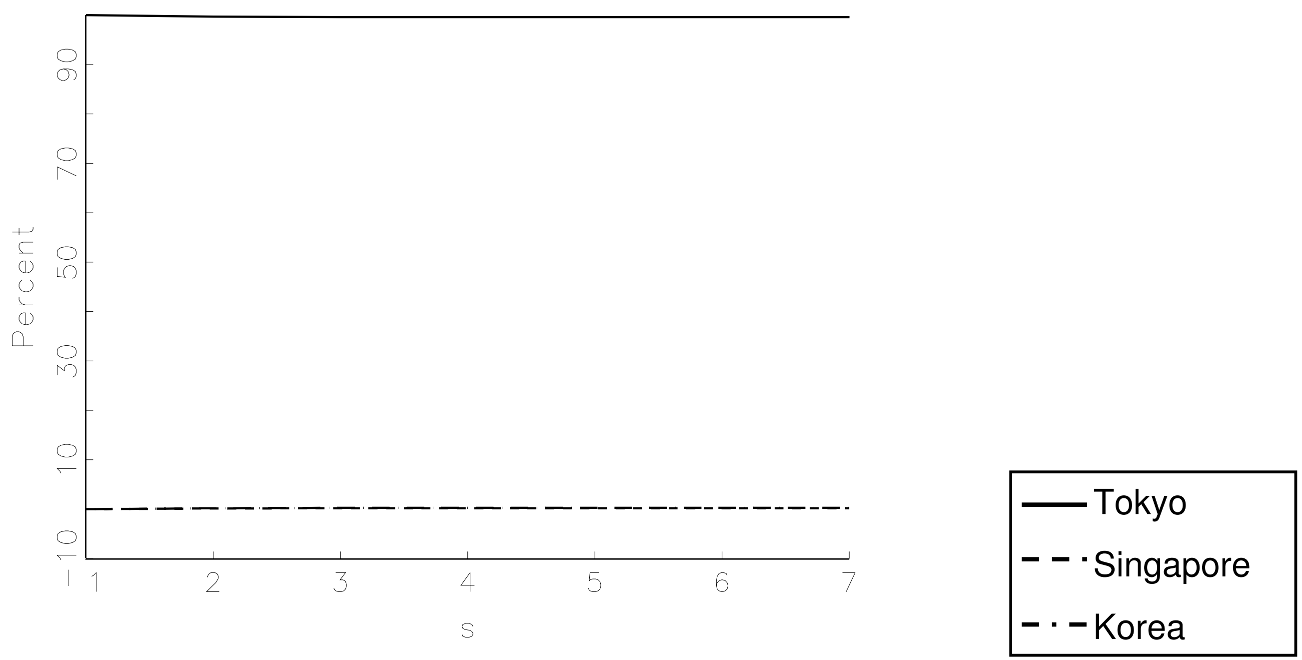
$$\underset{(n\times n)}{MSE(\hat{\mathbf{y}}_{t+2|t})} = \text{Var}(u_t^T)[\mathbf{a}_1\mathbf{a}_1' + \mathbf{\Psi}_1\mathbf{a}_1\mathbf{a}_1'\mathbf{\Psi}_1'] + \text{Var}(u_t^S)[\mathbf{a}_2\mathbf{a}_2' + \mathbf{\Psi}_1\mathbf{a}_2\mathbf{a}_2'\mathbf{\Psi}_1'] + \text{Var}(u_t^K)[\mathbf{a}_3\mathbf{a}_3' + \mathbf{\Psi}_1\mathbf{a}_3\mathbf{a}_3'\mathbf{\Psi}_1']$$

$$\underset{(n\times n)}{MSE(\hat{\mathbf{y}}_{t+2|t})} = 1.79 \cdot \begin{bmatrix} 1.000 & 0.343 & 0.086 \\ 0.343 & 0.125 & 0.035 \\ 0.086 & 0.035 & 0.012 \end{bmatrix} + 1.78 \cdot \begin{bmatrix} 0.001 & 0.008 & 0.004 \\ 0.008 & 1.040 & 0.147 \\ 0.004 & 0.147 & 0.026 \end{bmatrix} + 2.63 \cdot \begin{bmatrix} 0.001 & 0.002 & 0.002 \\ 0.002 & 0.002 & 0.003 \\ 0.002 & 0.003 & 1.004 \end{bmatrix} = \underline{\underline{\begin{bmatrix} 1.799 & 0.633 & 0.167 \\ 0.633 & 2.082 & 0.332 \\ 0.167 & 0.332 & 2.707 \end{bmatrix}}}$$

$\text{Var}(u_t^T), \text{Var}(u_t^S), \text{Var}(u_t^K)$ and $MSE(\hat{\mathbf{y}}_{t+2|t})$ taken times 10000

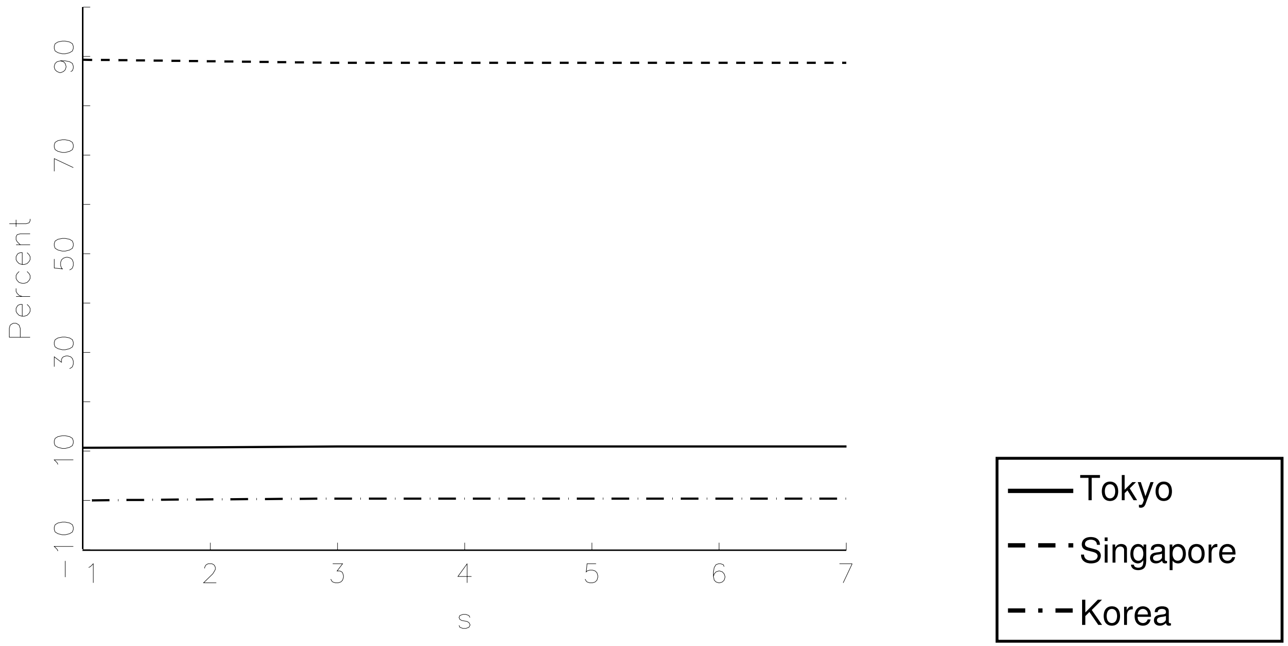
Variance Decomposition of Tokyo

Cholesky Ordering: Tokyo Singapore Korea



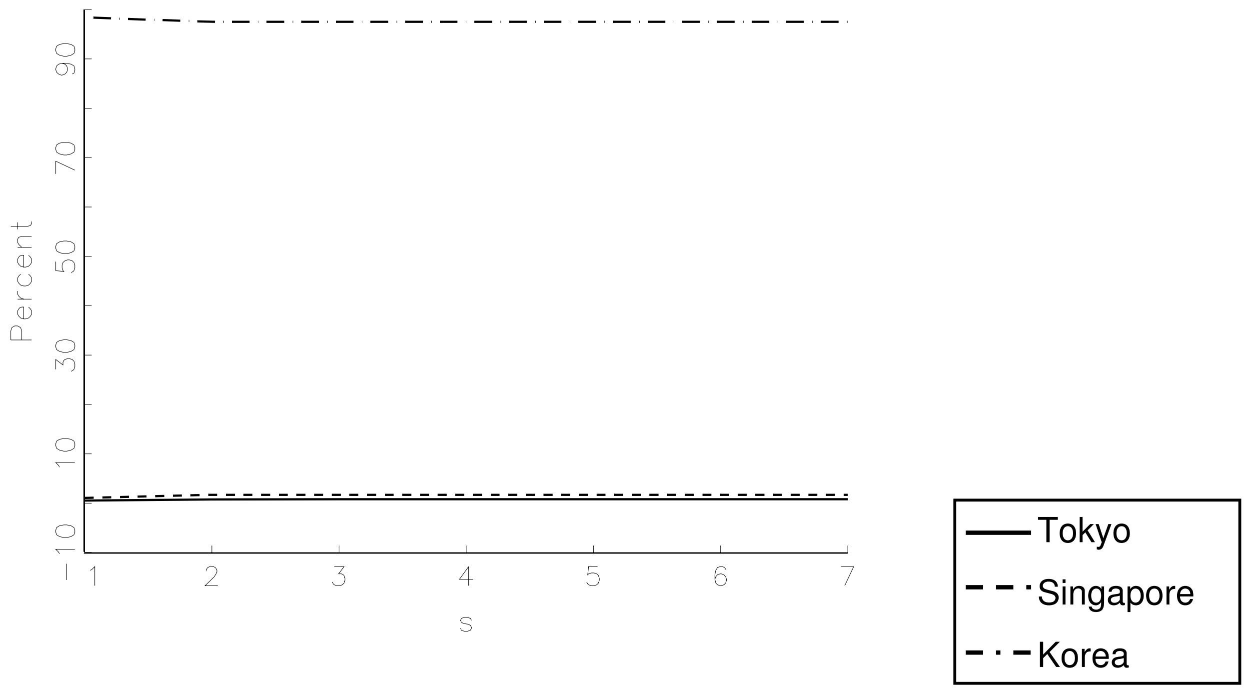
Variance Decomposition of Singapore

Cholesky Ordering: Tokyo Singapore Korea



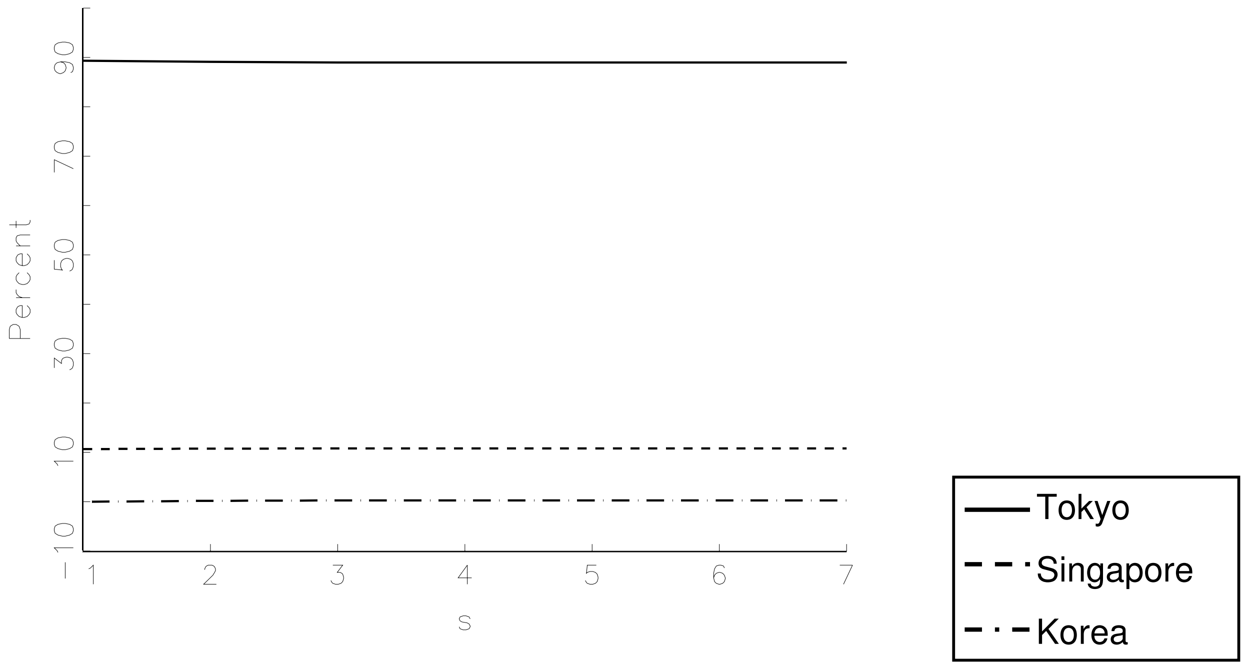
Variance Decomposition of Korea

Cholesky Ordering: Tokyo Singapore Korea



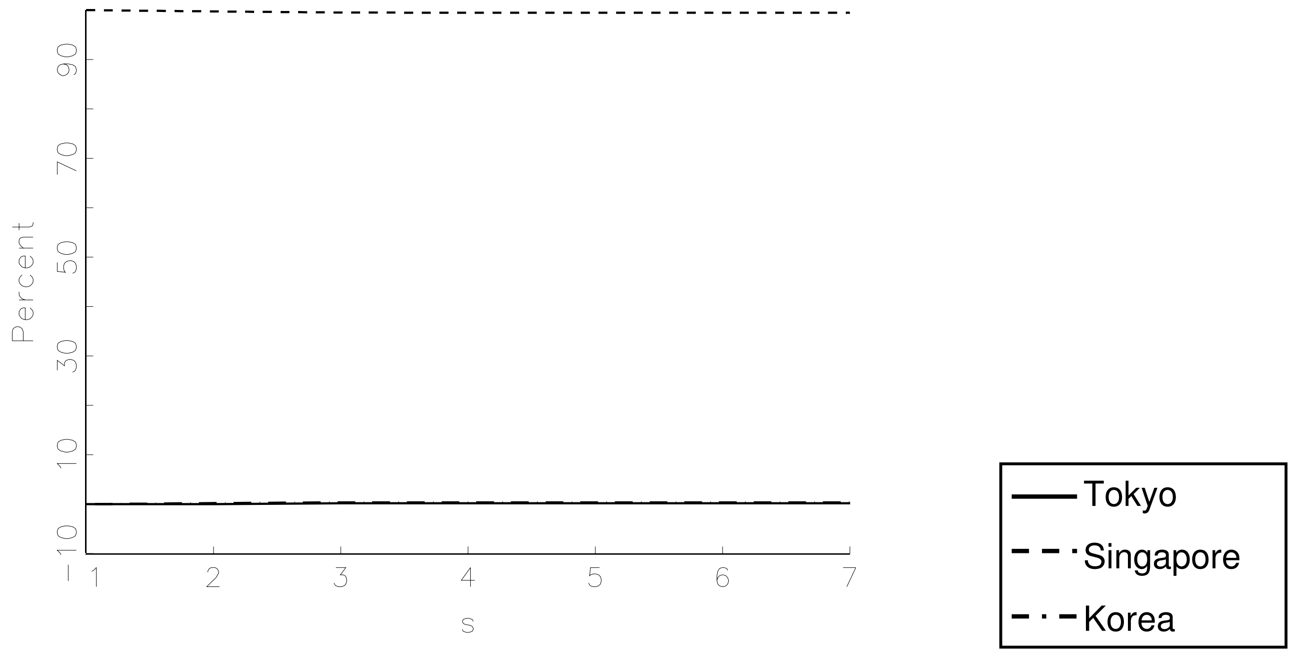
Variance Decomposition of Tokyo

Cholesky Ordering: Singapore Tokyo Korea



Variance Decomposition of Singapore

Cholesky Ordering: Singapore Tokyo Korea



Variance Decomposition of Korea

Cholesky Ordering: Singapore Tokyo Korea