Multivariate Time Series Analysis

Yannick Hoga

University of Duisburg-Essen

Winter 2019

Outline

Preleminaries

- Contact:

Yannick Hoga	Thilo Reinschlüssel
0201 18-34365	0201 18-34913
yannick.hoga@vwl.uni-due.de	thilo.reinschluessel@vwl.uni-due.de

- Slides, exercises and announcements will be provided on

  https://moodle.uni-due.de/course/view.php?id=18068.

  The course password is “MTSA”.

- To illustrate the theory, we will use the programming language R in the lectures and tutorials.

Resources

- **Books**
  - Hamilton (1994) *Time Series Analysis*. Princeton University Press, 1st ed.
  - Lütkepohl (2005) *New Introduction to Multiple Time Series Analysis*. Springer, 1st ed.
  - Tsay (2010) *Analysis of Financial Time Series*. Wiley, 3rd ed.
  - Tsay (2014) *Multivariate Time Series Analysis: With R and Financial Applications*. Wiley, 1st ed.
- **R packages**: `MTS`, `fGarch`, `forecast`, `tseries`, `fUnitRoots`, `urca` and many more
- If you load the above packages, you can replicate the figures in the slides in R

- In this lecture we consider K-dimensional time series

$$\{\boldsymbol{z}_t = (z_{1t}, \dots, z_{Kt})\}_{t=1,\dots,T}$$

- The goals of analyzing $\boldsymbol{z}_t$ include ...

  1. studying the relationships between component series z_{kt} and
  2. improving the accuracy in predicting $\boldsymbol{z}_{T+h}$ for $h > 0$.

- To achieve these, we shall ...

  1. explore the basic properties of $\boldsymbol{z}_t$ and
  2. study econometric models for $\boldsymbol{z}_t$.

- We shall introduce some basic concepts of multivariate time series in this Introduction.

- These are mostly straightforward extensions of the univariate concepts.

Example I

Quarterly U.S. GDP Growth and Unemployment in R

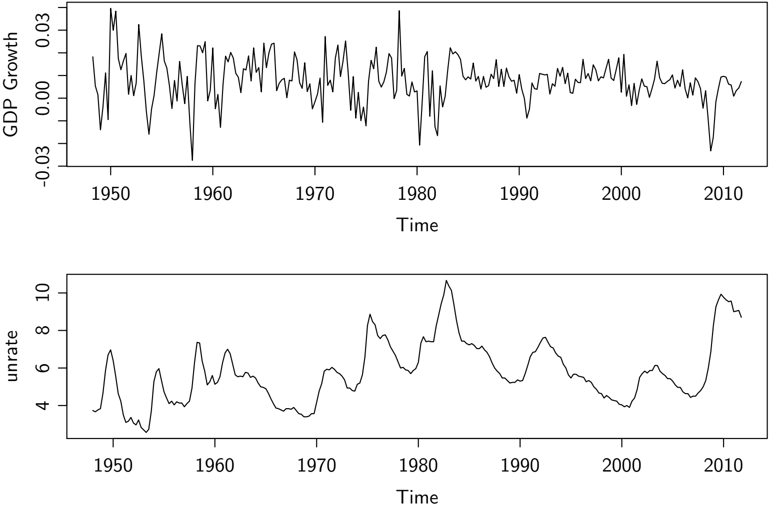
```r
da     <- read.table("Resources/Data/q-gdpunemp.txt", header = T)  # load data
head(da)
```

```
##   year mon  gdp  rate
## 1 1948   1 1822 3.733
## 2 1948   4 1855 3.667
## 3 1948   7 1865 3.767
## 4 1948  10 1868 3.833
## 5 1949   1 1842 4.667
## 6 1949   4 1836 5.867
```

```r
gdp    <- ts(da$gdp,  start = c(1948, 1), frequency = 4) # GDP from 1948Q1-2011Q4
unrate <- ts(da$rate, start = c(1948, 1), frequency = 4)
plot(diff(log(gdp)), ylab="GDP Growth")     # approximate GDP percentage growth
plot(unrate)                                # Unemployment in %
```

Example II

Quarterly U.S. GDP Growth and Unemployment in R



Example I

Weekly U.S. Gasoline and Crude Oil prices (Dollars/Gallon) in R

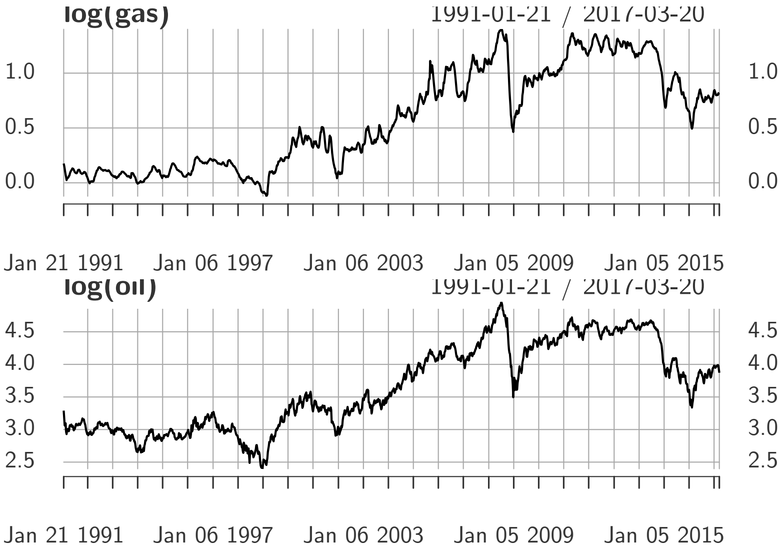
```
GAS.data <- read.csv("Resources/Data/w-gasreg.csv")
head(GAS.data)
```

```
##        DATE GASREGCOVW
## 1 1/21/1991      1.192
## 2 1/28/1991      1.168
## 3  2/4/1991      1.139
## 4 2/11/1991      1.106
## 5 2/18/1991      1.078
## 6 2/25/1991      1.054
```

```
OIL.data <- read.csv("Resources/Data/w-wti.csv")
Date <- as.Date(GAS.data$DATE, "%m/%d/%Y")          # Create date variable
gas  <- xts(x = GAS.data$GASREGCOVW, order.by = Date)
oil  <- xts(x = OIL.data$WCOILWTICO, order.by = Date)
plot(log(gas))
plot(log(oil))
```

Example II

Weekly U.S. Gasoline and Crude Oil prices (Dollars/Gallon) in R



Example I

Monthly IBM and S&P Composite Index Returns in R

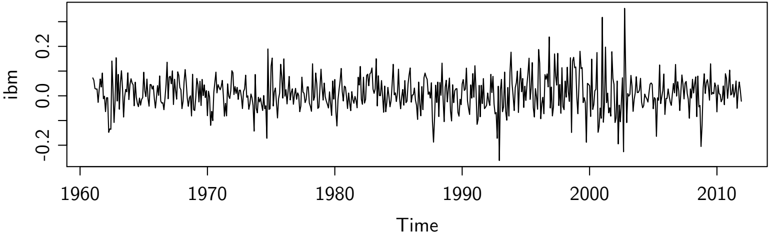
```r
data("mts-examples", package="MTS")
head(ibmspko)
```

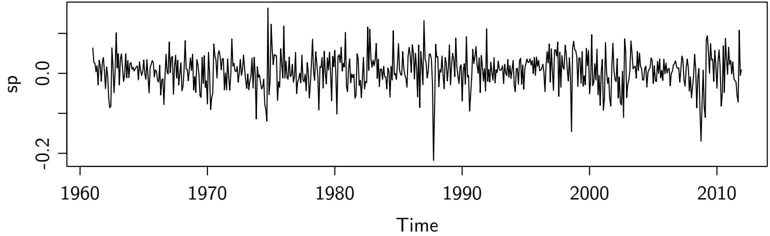
```
##       date      ibm        sp        ko
## 1 19610131  0.07251  0.063156  0.009331
## 2 19610228  0.06250  0.026870  0.103236
## 3 19610330  0.02963  0.025536  0.012291
## 4 19610428  0.02734  0.003843 -0.050000
## 5 19610531  0.02752  0.019139  0.087719
## 6 19610630 -0.02661 -0.028846 -0.058065
```

```r
ibm <- ts(ibmspko$ibm, start = c(1961, 1), frequency = 12) # from 1961M1-2011M12
sp  <- ts(ibmspko$sp,  start = c(1961, 1), frequency = 12)
plot(ibm)     # IBM simple returns
plot(sp)      # S&P Composite simple returns
```

Example II

Monthly IBM and S&P Composite Index Returns in R





Review of Matrix Operations

Multivariate time series analysis requires some additional matrix operations:

Definition 1.1:

Let $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_q)$ be a $(p \times q)$-matrix with columns $\boldsymbol{a}_i$. Then, **vectorization** gives a pq-dimensional column vector:

$$\text{vec}(\boldsymbol{A}) = (\boldsymbol{a}_1', \ldots, \boldsymbol{a}_q')'.$$

If $\boldsymbol{C}$ is a $(m \times n)$-matrix, then the **Kronecker product** $\boldsymbol{A} \otimes \boldsymbol{C}$ is an $(pm) \times (qn)$-matrix given by

$$\boldsymbol{A} \otimes \boldsymbol{C} = \begin{pmatrix} a_{11}\boldsymbol{C} & a_{12}\boldsymbol{C} & \ldots & a_{1q}\boldsymbol{C} \\ a_{21}\boldsymbol{C} & a_{22}\boldsymbol{C} & \ldots & a_{2q}\boldsymbol{C} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1}\boldsymbol{C} & a_{p2}\boldsymbol{C} & \ldots & a_{pq}\boldsymbol{C} \end{pmatrix},$$

where a_{ij} are the scalar elements of $\boldsymbol{A}$.

UNIVERSITÄT DUISBURG ESSEN

Proposition 1.2:

Assuming dimensions are proper, the following hold:

1. $(\boldsymbol{A} \otimes \boldsymbol{C})' = \boldsymbol{A}' \otimes \boldsymbol{C}'$.
2. $\boldsymbol{A} \otimes (\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{A} \otimes \boldsymbol{D}$.
3. $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{F} \otimes \boldsymbol{G}) = (\boldsymbol{AF}) \otimes (\boldsymbol{BG})$.
4. If $\boldsymbol{A}$ and $\boldsymbol{C}$ are invertible, then $(\boldsymbol{A} \otimes \boldsymbol{C})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{C}^{-1}$.
5. For square matrices $\boldsymbol{A}$ and $\boldsymbol{C}$, $\operatorname{tr}(A \otimes \boldsymbol{C}) = \operatorname{tr}(\boldsymbol{A}) \operatorname{tr}(\boldsymbol{C})$.
6. $\operatorname{vec}(\boldsymbol{A} + \boldsymbol{C}) = \operatorname{vec}(\boldsymbol{A}) + \operatorname{vec}(\boldsymbol{C})$,
7. $\operatorname{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{B})$.
8. $\operatorname{tr}(\boldsymbol{AC}) = \operatorname{vec}(\boldsymbol{C}')' \operatorname{vec}(\boldsymbol{A}) = \operatorname{vec}(\boldsymbol{A}')' \operatorname{vec}(\boldsymbol{C})$.
9. $\operatorname{tr}(\boldsymbol{ABC}) = \operatorname{vec}(\boldsymbol{A}')'(\boldsymbol{C}' \otimes \boldsymbol{I}) \operatorname{vec}(\boldsymbol{B})$, where $\boldsymbol{I}$ is the identity matrix.
10. $\operatorname{tr}(\boldsymbol{AC}) = \operatorname{tr}(\boldsymbol{CA})$.
11. $\operatorname{tr}(\boldsymbol{A} + \boldsymbol{C}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{C})$.

Definition 1.3:

If $f(\boldsymbol{\beta})$ is a scalar function of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$, then

$$\frac{\partial f}{\partial \boldsymbol{\beta}} = \left(\frac{\partial f}{\partial \beta_1}, \ldots, \frac{\partial f}{\partial \beta_m}\right)' \quad \text{and} \quad \frac{\partial f}{\partial \boldsymbol{\beta}'} = \frac{\partial f}{\partial \boldsymbol{\beta}}'.$$

The **Hessian matrix** is

$$\frac{\partial^2 f}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \begin{pmatrix} \frac{\partial^2 f}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 f}{\partial \beta_1 \partial \beta_m} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial \beta_m \partial \beta_1} & \cdots & \frac{\partial^2 f}{\partial \beta_m \partial \beta_m} \end{pmatrix}.$$

If $f(\boldsymbol{A})$ is a scalar function of the $(m \times n)$-matrix $\boldsymbol{A} = (a_{ij})$, then

$$\frac{\partial f}{\partial \boldsymbol{A}} = \left(\frac{\partial f}{\partial a_{ij}}\right).$$

Proposition 1.4:

Assuming dimensions are proper, the following hold:

1. $\frac{\partial \boldsymbol{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}'} = \boldsymbol{A} \qquad \frac{\partial \boldsymbol{\beta}'\boldsymbol{A}'}{\partial \boldsymbol{\beta}} = \boldsymbol{A}'$.
2. $\frac{\partial \boldsymbol{\beta}'\boldsymbol{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = (\boldsymbol{A} + \boldsymbol{A}')\beta$.
3. $\frac{\partial^2 \boldsymbol{\beta}'\boldsymbol{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \boldsymbol{A} + \boldsymbol{A}'$.
4. $\frac{\partial \log |\boldsymbol{A}|}{\partial \boldsymbol{A}} = (\boldsymbol{A}')^{-1}$.
5. For square matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$, we have $\frac{\partial \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C})}{\partial \boldsymbol{A}} = -(\boldsymbol{A}^{-1}\boldsymbol{C}\boldsymbol{B}\boldsymbol{A}^{-1})'$.

Definition 1.5:

If there is a $(K \times 1)$-vector $\boldsymbol{v}$ such that

$$\underset{(K\times K)}{\boldsymbol{A}}\boldsymbol{v} = \lambda \boldsymbol{v} \tag{1.1}$$

for some scalar λ, then λ is called the **eigenvalue** of $\boldsymbol{A}$ with corresponding **eigenvector** $\boldsymbol{v}$.

Remark 1.6:

Rewrite (1.1) as $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = 0$. Recall that a linear system of equations has nontrivial solutions if and only if the determinant vanishes, so the solutions of this equation are given by the solutions of the **characteristic equation**

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0.$$

Lemma 1.7: .

For a symmetric positive definite $(m \times m)$ matrix $\boldsymbol{A}$ there exists a unique lower triangular $(m \times m)$ matrix $\boldsymbol{B}$ with real positive diagonal elements such that $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}^\top$.

Example 1.8:

For $\boldsymbol{A}$ (2×2) we have

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{21} \\ 0 & b_{22} \end{pmatrix} \Rightarrow$$

$$\begin{aligned} b_{11} &= \sqrt{a_{11}}, \\ b_{21} &= a_{12}/b_{11}, \\ b_{22} &= \sqrt{a_{22} - b_{21}^2} \end{aligned}$$

Stationarity

Definition 1.9:

1. If for any $t_1, \ldots, t_m$ ($m \geq 1$) and $l > 0$

$$(\boldsymbol{z}'_{t_1}, \ldots, \boldsymbol{z}'_{t_m})' \overset{\mathcal{D}}{=} (\boldsymbol{z}'_{t_1+l}, \ldots, \boldsymbol{z}'_{t_m+l})',$$

then $\boldsymbol{z}_t$ is **strictly stationary**.
2. Assume that $\boldsymbol{\mu}_t = E(\boldsymbol{z}_t)$ and $\boldsymbol{\Gamma}_{t,l} = \operatorname{cov}(\boldsymbol{z}_t, \boldsymbol{z}_{t-l})$ exist for all t and l. Then, $\boldsymbol{z}_t$ is **weakly stationary** if $\boldsymbol{\mu}_t = \boldsymbol{\mu}$ and $\boldsymbol{\Gamma}_{t,l} = \boldsymbol{\Gamma}_l$.

Remark 1.10:

1. Strict stationarity means the joint distributions are time invariant, whereas weak stationarity only requires the first two (cross-)moments to be time invariant.
2. In general, strict stationarity does not imply weak stationarity and vice versa.
3. We take *stationary* to mean *weakly stationary* in the following.

UNIVERSITÄT DUISBURG ESSEN

Definition 1.11:

We call $\boldsymbol{z}_t$ **linear**, if

$$\boldsymbol{z}_t = \underset{(K\times 1)}{\boldsymbol{\mu}} + \sum_{i=-\infty}^{\infty} \underset{(K\times K)}{\boldsymbol{\Psi}_i} \underset{(K\times 1)}{\boldsymbol{a}_{t-i}},$$

where $\boldsymbol{a}_t \overset{\text{i.i.d.}}{\sim} (\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}})$ are the **innovations**. $\boldsymbol{z}_t$ is **causal**, if $\boldsymbol{\Psi}_i = \boldsymbol{0}$ for $i < 0$.

Remark 1.12:

1. We often assume $\boldsymbol{\Sigma}_{\boldsymbol{a}}$ to be positive definite; otherwise, the dimension of $\boldsymbol{z}_t$ can be reduced.
2. For the linear time series in the above definition to be meaningful, we assume $\sum_{i=-\infty}^{\infty} \|\boldsymbol{\Psi}_i\| < \infty$, where $\|\cdot\|$ denotes any matrix norm (e.g., the Frobenius norm $\|\boldsymbol{A}\| = \sqrt{\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}')}$). This implies in particular that $\|\boldsymbol{\Psi}_i\| \underset{(i\to\pm\infty)}{\longrightarrow} 0$, i.e., there is negligible influence of the infinite past on $\boldsymbol{z}_t$.

Linear Models

- In the first part of this course we focus on causal linear time series

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i \boldsymbol{a}_{t-i}. \tag{1.2}$$

- If $\boldsymbol{a}_t$ is multivariate Gaussian, $\boldsymbol{z}_t$ is called a **multivariate Gaussian process**.
- For the model in (1.2) to be **identifiable**, we assume that
  1. $\boldsymbol{\Psi}_0 = \boldsymbol{I}$;
  2. $\operatorname{cov}(\boldsymbol{a}_t) = \boldsymbol{\Sigma}_{\boldsymbol{a}}$ is positive definite.
- An alternative representation is

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i \boldsymbol{b}_{t-i}, \qquad \boldsymbol{b}_t \overset{\text{i.i.d.}}{\sim} (\boldsymbol{0}, \boldsymbol{I}_K),$$

where $\boldsymbol{\Psi}_0$ is lower triangular with 1's on the diagonal. Why?

Invertibility

Definition 1.13:

We call $\boldsymbol{z}_t$ **invertible**, if

$$\boldsymbol{z}_t = \phi_0 + \sum_{i=1}^{\infty} \phi_i \boldsymbol{z}_{t-i} + \boldsymbol{a}_t$$

for $(K \times K)$-matrices ϕ_i satsifying $\sum_{i=1}^{\infty} \|\phi_i\| < \infty$.

- Invertibility means that we can reconstruct the innovations $\boldsymbol{a}_t$ from the history of observed data $\boldsymbol{z}_t, \boldsymbol{z}_{t-1}, \ldots$.
- Non-invertibility means that structural shocks cannot be recovered from a history of observed variables
- Why is invertibility important? Forecasting.

Assume that $\boldsymbol{z}_t$ is (weakly) stationary.

Definition 1.14:

The **lag-l autocovariance matrix** of $\boldsymbol{z}_t$ is defined by

$$\boldsymbol{\Gamma}_l = (\Gamma_{ij}(l))_{i,j=1}^{k} = \mathrm{cov}(\boldsymbol{z}_t, \boldsymbol{z}_{t-l}) = \mathrm{E}[(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_{t-l} - \boldsymbol{\mu})'].$$

Remark 1.15:

- $\Gamma_{ij}(l)$ quantifies the linear dependence of z_{it} and $z_{j,t-l}$. Hence, except for $l = 0$, $\boldsymbol{\Gamma}_l$ is not symmetric, because $\Gamma_{ij}(l)$ is the covariance between z_{it} and $z_{j,t-l}$, and $\Gamma_{ji}(l)$ is the covariance between z_{jt} and $z_{i,t-l}$
- It is left as an Exercise to show that $\boldsymbol{\Gamma}_l = \boldsymbol{\Gamma}'_{-l}$. Therefore, we only need to consider $\boldsymbol{\Gamma}_l$ for $l \geq 0$.

Definition 1.16:

The **lag-/ cross-correlation matrix (CCM)** of $\boldsymbol{z}_t$ is defined by

$$\boldsymbol{\rho}_l = (\rho_{ij}(l))_{i,j=1}^{K} = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_l\boldsymbol{D}^{-1},$$

where $\boldsymbol{D} = \operatorname{diag}\{\sqrt{\Gamma_{11}(0)}, \ldots, \sqrt{\Gamma_{KK}(0)}\}$. The function $l \mapsto \boldsymbol{\rho}_l$ is the **CCM function**.

Remark 1.17:

- $\rho_{ij}(l) = \operatorname{cov}(z_{it}, z_{j,t-l})/\sqrt{\operatorname{Var}(z_{it})\operatorname{Var}(z_{j,t-l})}$ is the well-known correlation coefficient.
- Again, $\boldsymbol{\rho}_{-l} = \boldsymbol{\rho}_l'$.
- Since $\boldsymbol{\rho}_l$ is a matrix, we typcially plot the CCMs element-by-element. For instance, plotting $l \mapsto \rho_{11}(l)$ gives the univariate autocorrelation function (ACF) of z_{1t}, plotting $l \mapsto \rho_{12}(l)$ provides the dependence of z_{1t} on lagged z_{2t}, etc.

Using the CCM function, the linear relationship between $\{z_{it}\}$ and $\{z_{jt}\}$ can be summarized as follows:

1. z_{it} and z_{jt} have *no linear relationship* if $\rho_{ij}(l) = \rho_{ji}(l) = 0$ for all $l \geq 0$.
2. z_{it} and z_{jt} are *concurrently correlated* if $\rho_{ij}(0) \neq 0$.
3. z_{it} and z_{jt} are *uncoupled* if $\rho_{ij}(l) = 0$ and $\rho_{ji}(l) = 0$ for all $l > 0$.
4. There is a *unidirectional relationship* from z_{it} to z_{jt} if $\rho_{ij}(l) = 0$ for all $l > 0$, but $\rho_{ji}(v) \neq 0$ for some $v > 0$. In this case, z_{it} does not depend on past z_{jt}, but z_{jt} depends on some past values of z_{it}.
5. There is a *feedback relationship* between z_{it} and z_{jt} if $\rho_{ij}(l) \neq 0$ for some $l > 0$ and $\rho_{ji}(v) \neq 0$ for some $v > 0$.

UNIVERSITÄT DUISBURG ESSEN

Definition 1.18:

Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T$ be a sample of a stationary time series. Then, the **lag-l sample cross-covariance matrix** is

$$\widehat{\boldsymbol{\Gamma}}_l = \frac{1}{T} \sum_{t=l+1}^{T} (\boldsymbol{z}_t - \overline{\boldsymbol{z}})(\boldsymbol{z}_{t-l} - \overline{\boldsymbol{z}})', \qquad l \geq 0,$$

where $\overline{\boldsymbol{z}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{z}_t$ is the sample mean. The **lag-l sample CCM** is

$$\widehat{\boldsymbol{\rho}}_l = \widehat{\boldsymbol{D}}^{-1} \widehat{\boldsymbol{\Gamma}}_l \widehat{\boldsymbol{D}}^{-1}, \qquad l \geq 0,$$

where $\widehat{\boldsymbol{D}} = \mathrm{diag}\{\widehat{\Gamma}_{11}^{1/2}(0), \ldots, \widehat{\Gamma}_{KK}^{1/2}(0)\}$ is the diagonal matrix of the sample standard deviations of the component series.

Testing for Serial Independence

- Clearly, if there is no dependence over time in $\boldsymbol{z}_t$, we do not have to model anything and our job is done.
- So in practice, we would like to test for some integer m

$$H_0 \; : \; \boldsymbol{\rho}_1 = \ldots = \boldsymbol{\rho}_m = \mathbf{0} \qquad \text{vs.} \qquad H_1 \; : \; \boldsymbol{\rho}_i \neq 0 \text{ for some } 1 \leq i \leq m.$$

- We shall discuss a *multivariate Portmanteau test* based on

Theorem 1.19:

If $\boldsymbol{z}_t \overset{\text{i.i.d.}}{\sim} (\mathbf{0}, \boldsymbol{\Sigma})$, then for $l > 0$,

$$\sqrt{T}\widehat{\rho}_{ij}(l) \xrightarrow{d} N(0,1), \qquad \text{as } T \to \infty.$$

- Hence, the cross-correlation is insignificant at the 5%-level if $|\widehat{\rho}_{ij}(l)| \geq 2/\sqrt{T}$.
- When K is moderate or large, it is hard to digest many CCM functions simultaneously.

Testing for Serial Independence

Multivariate Portmanteau Test

- To formally test H_0, consider the multivariate extension of the Ljung–Box statistic:

$$Q_K(m) = T^2 \sum_{i=1}^{m} \frac{1}{T-i} \widehat{\boldsymbol{b}}_i' (\widehat{\boldsymbol{\rho}_0}^{-1} \otimes \widehat{\boldsymbol{\rho}_0}^{-1}) \widehat{\boldsymbol{b}}_i,$$

  where

  - $\otimes$ is the Kronecker product of matrices and
  - $\widehat{\boldsymbol{b}}_i = \mathrm{vec}(\widehat{\boldsymbol{\rho}}_i')$, where $\mathrm{vec}(\boldsymbol{A})$ is the column-stacking vector of the matrix $\boldsymbol{A}$.

- The asymptotic distribution is given by the following

Corollary 1.20: Tsay (2014), Sec. 1.5.

Under the assumptions of Theorem 1.19,

$$Q_K(m) \xrightarrow{d} \chi^2_{mK^2}, \qquad \text{as } T \to \infty.$$

- For $K = 1$, the well-known univariate result obtains.

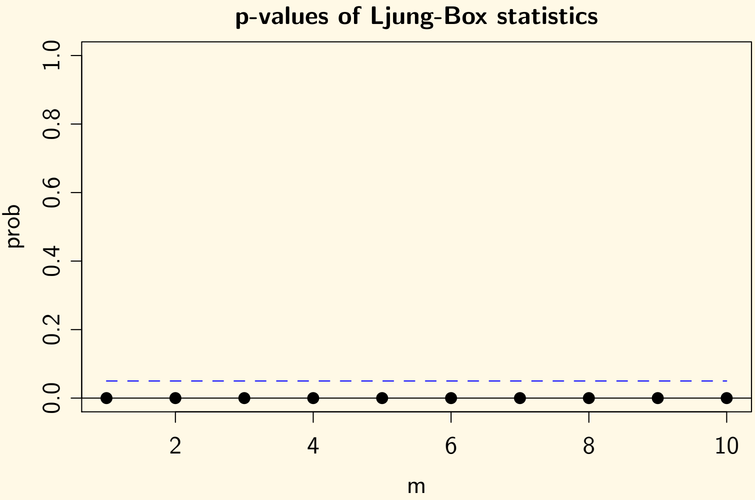
Example 1.21:

We apply the multivariate Portmanteau test to the GDP growth and unemployment rates. Of course, we expect a rejection of H_0.

```
x <- cbind(diff(log(gdp)), unrate[-1])  # combine GDP growth and unemployment
mq(x, lag = 10)                         # Compute Q_k(m) statistic up to m = 10
```

```
## Ljung-Box Statistics:
##          m       Q(m)     df    p-value
##  [1,]    1        286      4          0
##  [2,]    2        533      8          0
##  [3,]    3        737     12          0
##  [4,]    4        902     16          0
##  [5,]    5       1031     20          0
##  [6,]    6       1131     24          0
##  [7,]    7       1210     28          0
##  [8,]    8       1271     32          0
##  [9,]    9       1321     36          0
## [10,]   10       1362     40          0
```

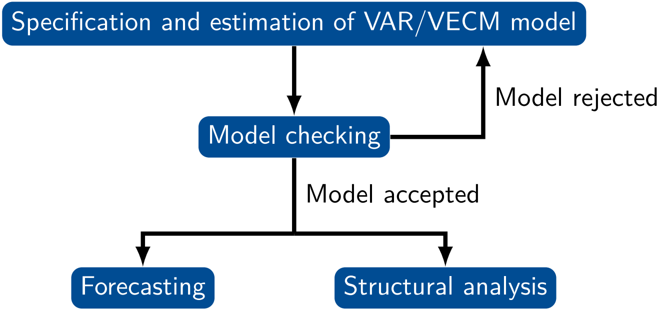
Example 1.21: Continued.



Outline

- The outline of this lecture is as follows:

  1. VAR models (Chapters 2–6)
     - This part is mostly based on Tsay (2014)
  2. vector error correction models (Chapters 7–9)
     - This part is mostly based on Lütkepohl (2005, Part II)


Specification and estimation of VAR/VECM model
Model rejected
Model checking
Model accepted
Forecasting
Structural analysis

Outline

VAR Models

For good reasons, the most commonly used model in analyzing multivariate time series is the vector autoregressive (VAR) model:

1. Simplicity and ease of interpretation.
2. Easy to estimate via maximum likelihood (ML), ordinary least squares (OLS) or generalized least squares (GLS)
   - OLS = GLS, OLS asympt. equiv. to ML
3. Properties studied extensively in the literature.

- As in univariate time series, the building blocks of dependent multivariate times series are white noise processes:

Definition 2.1: White Noise.

The process $\boldsymbol{a}_t$ is a **white noise** if it is independent, identically distributed (i.i.d.) with mean $\mathbf{0}$ and positive definite (p.d.) variance-covariance matrix $\boldsymbol{\Sigma_a}$ ($\boldsymbol{a}_t \overset{\text{i.i.d.}}{\sim} (\mathbf{0}, \boldsymbol{\Sigma_a})$).

Definition 2.2:

A multivariate time series $\boldsymbol{z}_t$ follows a VAR model of order p, VAR(p), if

$$\boldsymbol{z}_t = \phi_0 + \sum_{i=1}^{p} \phi_i \boldsymbol{z}_{t-i} + \boldsymbol{a}_t$$

for a white noise $\boldsymbol{a}_t$.

Remark 2.3:

Using the lag operator L, a VAR(p) model can be written concisely as

$$\phi(L)\boldsymbol{z}_t = \phi_0 + \boldsymbol{a}_t,$$

where $\phi(L) = \boldsymbol{I}_K - \sum_{i=1}^{p} \phi_i L^i$. The lag operator L is defined as $L^i \boldsymbol{z}_t = \boldsymbol{z}_{t-i}$.

VAR(1) Models

Interpretation

- Consider the bivariate VAR(1) Model $\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t$ with $\boldsymbol{\phi}_0 = (\phi_{10}, \phi_{20})'$ and $\boldsymbol{\phi}_1 = (\phi_{1,ij})_{i,j=1,2}$, i.e.

$$z_{1,t} = \phi_{10} + \phi_{1,11} z_{1,t-1} + \phi_{1,12} z_{2,t-1} + a_{1t},$$
$$z_{2,t} = \phi_{20} + \phi_{1,21} z_{1,t-1} + \phi_{1,22} z_{2,t-1} + a_{2t}.$$

- Thus,
  - $\phi_{1,12}$ measures the linear dependence of $z_{2,t-1}$ and $z_{1,t}$ in the presence of $z_{1,t-1}$. In other words, $\phi_{1,12}$ is the conditional effect of $z_{2,t-1}$ on $z_{1,t}$ given $z_{1,t-1}$.
  - $\phi_{1,21}$ measures the linear dependence of $z_{1,t-1}$ and $z_{2,t}$ in the presence of $z_{2,t-1}$.
  - ...

UNIVERSITÄT DUISBURG ESSEN

Example 2.4:

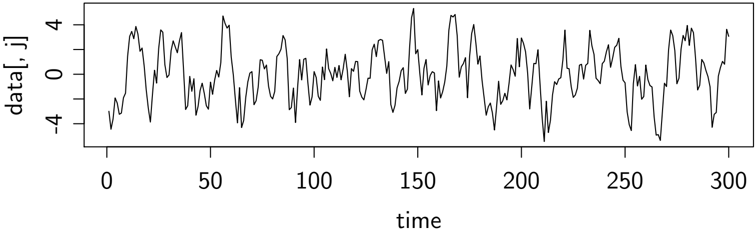
We generate a trajectory of length $T = 300$ of the bivariate VAR(1) $\boldsymbol{z}_t = \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t$ with normally distributed $\boldsymbol{a}_t$

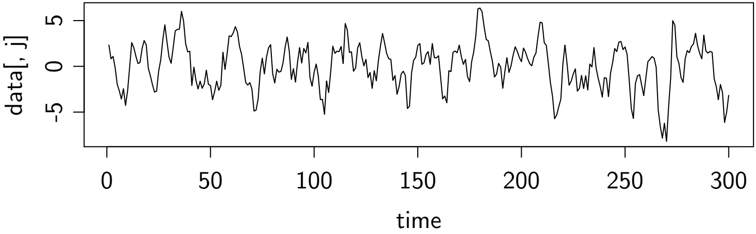
$$\boldsymbol{\phi}_1 = \begin{pmatrix} 0.8 & 0.4 \\ -0.3 & 0.6 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{a}} = \begin{pmatrix} 2.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}.$$

```
m1 <- VARMAsim(300, arlags=c(1), phi   = matrix(c(0.8, 0.4, -0.3, 0.6), 2, 2),
                                 sigma = matrix(c(2.0, 0.5,  0.5, 2.0), 2, 2) )
zt <- m1$series          # extract the trajectory ...
MTSplot(zt)              # ... and plot them
```

Example II

Trajectory in R





Stationarity

- Use the definition of VAR(1) and iterate backwards j times to obtain

$$\begin{aligned}
\boldsymbol{z}_t &= \phi_0 + \phi_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t \\
&= \phi_0 + \phi_1(\phi_0 + \phi_1 \boldsymbol{z}_{t-2} + \boldsymbol{a}_{t-1}) + \boldsymbol{a}_t \\
&= (I_K + \phi_1)\phi_0 + \phi_1^2 \boldsymbol{z}_{t-2} + \phi_1 \boldsymbol{a}_{t-1} + \boldsymbol{a}_t \\
&= \dots \\
&= \underbrace{(I_K + \phi_1 + \dots \phi_1^j)\phi_0}_{=(1)} + \underbrace{\phi_1^{j+1} \boldsymbol{z}_{t-j-1}}_{=(2)} + \underbrace{\sum_{i=0}^{j} \phi_1^i \boldsymbol{a}_{t-i}}_{=(3)} .
\end{aligned}$$

- Denote the eigenvalues of ϕ_1 by $\lambda_1, \dots, \lambda_K$.

Stationarity

- For (2): It is easily verified that the eigenvalues of ϕ_1^{j+1} are $\lambda_1^{j+1}, \ldots, \lambda_K^{j+1}$. If all eigenvalues of a matrix are 0, then the matrix must be $\mathbf{0}$. Consequently, if $\lambda_k^{j+1} \underset{(j\to\infty)}{\longrightarrow} 0$ for all $k = 1, \ldots, K$, then $\phi_1^{j+1} \underset{(j\to\infty)}{\longrightarrow} 0$. Thus,

$$\phi_1^{j+1} \boldsymbol{z}_{t-j-1} \underset{(j\to\infty)}{\longrightarrow} 0.$$

  Hence, we require $|\lambda_k| < 1$ $(k = 1, \ldots, K)$.

- For (1) and (3): If $|\lambda_k| < 1$ $(k = 1, \ldots, K)$, then from Lütkepohl (2005, p. 14) $\sum_{i=0}^{\infty} \phi_1^i \boldsymbol{a}_{t-i}$ exists in a mean-square sense and

$$(I_K + \phi_1 + \ldots + \phi_1^j)\phi_0 \underset{(j\to\infty)}{\longrightarrow} (I_K - \phi_1)^{-1}\phi_0.$$

- Combining these results suggests the stationary solution

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \phi_1^i \boldsymbol{a}_{t-i}, \qquad \text{where} \quad \boldsymbol{\mu} = (I_K - \phi_1)^{-1}\phi_0. \tag{2.1}$$

- The above shows that a necessary condition for a VAR(1) to be stationary is that all eigenvalues of ϕ_1 must be less than 1 in absolute value.
- We even have the following stronger result:

Theorem 2.5:

A VAR(1) is stationary with stationary solution (2.1) if and only if all eigenvalues of ϕ_1 are less than 1 in absolute value.

Remark 2.6:

The eigenvalues of ϕ_1 are less than 1 in absolute value if and only if

$$|\boldsymbol{I}_K - \phi_1 z| \neq 0 \qquad \text{for } |z| \leq 1, \tag{2.2}$$

that is, the polynomial $|\boldsymbol{I}_K - \phi_1 z|$ has no roots in and on the complex unit circle.

Example 2.7:

Consider the bivariate VAR(1) model

$$\begin{pmatrix} z_{1,t} \\ z_{2,t} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{pmatrix} \begin{pmatrix} z_{1,t-1} \\ z_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{1,t} \\ a_{2,t} \end{pmatrix}.$$

To see if this VAR model is stationary we could use (2.2) or directly calculate the eigenvalues, which are the roots of

$$\left| \left(\begin{pmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{pmatrix} - \lambda \boldsymbol{I}_2 \right) \right| = (0.2 - \lambda)(1.1 - \lambda) + 0.3 \cdot 0.6.$$

The roots of this polynomial are 0.5 and 0.8. Hence, the VAR is stationary. Note that the element $\phi_{1,22} = 1.1$ is greater than 1 in absolute value, but the series is stationary. This demonstrates that the eigenvalues determine stationarity, not the individual elements of $\boldsymbol{\phi}_1$.

Invertibility

- A VAR(1) model is invertible by definition.

Mean of a Stationary VAR(1)

- The mean of a stationary VAR(1) can easily be calculated from (2.1):

$$\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{z}_t) = (\boldsymbol{I}_K - \phi_1)^{-1}\phi_0.$$

- Alternatively, since a stationary VAR(1) has constant first moments through time, we apply $\mathrm{E}(\cdot)$ to the model equation to obtain

$$\begin{aligned}
& \mathrm{E}(\boldsymbol{z}_t) = \phi_0 + \phi_1 \mathrm{E}(\boldsymbol{z}_{t-1}) + \mathrm{E}(\boldsymbol{a}_t) \\
\iff \quad & \boldsymbol{\mu} = \phi_0 + \phi_1 \boldsymbol{\mu} \\
\iff \quad & \boldsymbol{\mu} = (\boldsymbol{I}_K - \phi_1)^{-1}\phi_0.
\end{aligned}$$

- We can again use (2.1) to get

$$\begin{aligned}
\boldsymbol{\Gamma}_h &= \mathrm{E}(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_{t-h} - \boldsymbol{\mu})' \\
&= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \boldsymbol{\phi}_1^i \boldsymbol{a}_{t-i}\right)\left(\sum_{j=0}^{\infty} \boldsymbol{\phi}_1^j \boldsymbol{a}_{t-h-j}\right)'\right] \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \boldsymbol{\phi}_1^i \,\mathrm{E}(\boldsymbol{a}_{t-i}\boldsymbol{a}'_{t-h-j})(\boldsymbol{\phi}_1^j)' \\
&= \sum_{i=0}^{\infty} \boldsymbol{\phi}_1^{h+i} \boldsymbol{\Sigma}_{\boldsymbol{a}} (\boldsymbol{\phi}_1^i)'.
\end{aligned}$$

- Alternatively, we can use the model equation $\boldsymbol{z}_t - \boldsymbol{\mu} = \boldsymbol{\phi}_1(\boldsymbol{z}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{a}_t$, which implies

$$\begin{aligned}\boldsymbol{\Gamma}_0 &= \mathsf{E}(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_t - \boldsymbol{\mu})' \\ &= \mathsf{E}[\boldsymbol{\phi}_1(\boldsymbol{z}_{t-1} - \boldsymbol{\mu})(\boldsymbol{z}_{t-1} - \boldsymbol{\mu})'\boldsymbol{\phi}_1'] + \mathsf{E}[\boldsymbol{a}_t\boldsymbol{a}_t'] \\ &= \boldsymbol{\phi}_1\boldsymbol{\Gamma}_0\boldsymbol{\phi}_1' + \boldsymbol{\Sigma}_{\boldsymbol{a}},\end{aligned}$$

where we used that $\boldsymbol{a}_t$ is uncorrelated with $(\boldsymbol{z}_{t-1} - \boldsymbol{\mu})$; see (2.1)

- Using $\operatorname{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\operatorname{vec}(\boldsymbol{B})$, we rewrite this as

$$\operatorname{vec}(\boldsymbol{\Gamma}_0) = (\boldsymbol{\phi}_1 \otimes \boldsymbol{\phi}_1)\operatorname{vec}(\boldsymbol{\Gamma}_0) + \operatorname{vec}(\boldsymbol{\Sigma}_{\boldsymbol{a}})$$

or, equivalently,

$$(\boldsymbol{I}_{K^2} - \boldsymbol{\phi}_1 \otimes \boldsymbol{\phi}_1)\operatorname{vec}(\boldsymbol{\Gamma}_0) = \operatorname{vec}(\boldsymbol{\Sigma}_{\boldsymbol{a}}) \tag{2.3}$$

- Given $\boldsymbol{\phi}_1$ and $\boldsymbol{\Sigma}_{\boldsymbol{a}}$, we can calculate $\boldsymbol{\Gamma}_0$ from (2.3).

- Post-multiplying $(\boldsymbol{z}_{t-l} - \boldsymbol{\mu})'$ to $\boldsymbol{z}_t - \boldsymbol{\mu} = \boldsymbol{\phi}_1(\boldsymbol{z}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{a}_t$, we get

$$(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_{t-l} - \boldsymbol{\mu})' = \boldsymbol{\phi}_1(\boldsymbol{z}_{t-1} - \boldsymbol{\mu})(\boldsymbol{z}_{t-l} - \boldsymbol{\mu})' + \boldsymbol{a}_t(\boldsymbol{z}_{t-l} - \boldsymbol{\mu})'.$$

- Taking expectations and using that $\boldsymbol{z}_{t-l}$ is uncorrelated with $\boldsymbol{a}_t$ gives

$$\boldsymbol{\Gamma}_l = \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{l-1}, \qquad l > 0. \tag{2.4}$$

- Using (2.3) in conjunction with (2.4), we can calculate the cross-covariance matrices of $\boldsymbol{z}_t$ and the CCMs.

Example 2.8:

Using (2.3) and (2.4) we can calculate the the autocovariance matrices and CCMs of the bivariate VAR(1) of Example 2.7 with

$$\boldsymbol{\Sigma}_{\boldsymbol{a}} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 2 \end{pmatrix}$$

as

$$\boldsymbol{\Gamma}_0 = \begin{pmatrix} 2.29 & 3.51 \\ 3.51 & 8.62 \end{pmatrix}, \qquad \boldsymbol{\Gamma}_1 = \begin{pmatrix} 1.51 & 3.29 \\ 2.49 & 7.38 \end{pmatrix},$$

and

$$\boldsymbol{\rho}_0 = \begin{pmatrix} 1.00 & 0.79 \\ 0.79 & 1.0 \end{pmatrix}, \qquad \boldsymbol{\rho}_1 = \begin{pmatrix} 0.66 & 0.74 \\ 0.56 & 0.86 \end{pmatrix}.$$

CCMs of a Stationary VAR(1) I

Demonstration in R

```
phi1 <- matrix(c(0.2, -0.6, 0.3, 1.1), 2, 2)   # input phi_1
sig  <- matrix(c(1, 0.8, 0.8, 2), 2, 2)         # input Sigma_a
m1   <- eigen(phi1)                             # obtain eigenvalues & vectors
m1
```

```
## eigen() decomposition
## $values
## [1] 0.8 0.5
##
## $vectors
##            [,1]    [,2]
## [1,] -0.4472 -0.7071
## [2,] -0.8944 -0.7071
```

```
I4   <- diag(4)                                 # create 4-by-4 identity matrix
pp   <- kronecker(phi1, phi1)                   # Kronecker product
c1   <- c(sig)                                  # calculate vec(Sigma_a)
dd   <- I4 - pp
ddinv<- solve(dd)                               # Obtain inverse
gam0 <- ddinv %*% matrix(c1, 4, 1)              # Obtain Gamma_0
gam0
```

CCMs of a Stationary VAR(1) II

Demonstration in R

```
##        [,1]
## [1,] 2.289
## [2,] 3.511
## [3,] 3.511
## [4,] 8.622
```

```r
g0   <- matrix(gam0, 2, 2)
g0
```

```
##        [,1]  [,2]
## [1,] 2.289 3.511
## [2,] 3.511 8.622
```

```r
g1   <- phi1 %*% g0                          # Obtain Gamma_1
g1
```

```
##        [,1]  [,2]
## [1,] 1.511 3.289
## [2,] 2.489 7.378
```

```r
D    <- diag( sqrt(diag(g0)) )               # To compute CCMs
D
```

CCMs of a Stationary VAR(1) III

Demonstration in R

```
##       [,1]  [,2]
## [1,] 1.513 0.000
## [2,] 0.000 2.936
```

```r
Di   <- solve(D)
Di %*% g0 %*% Di                                          # obtain rho_0
```

```
##        [,1]   [,2]
## [1,] 1.0000 0.7904
## [2,] 0.7904 1.0000
```

```r
Di %*% g1 %*% Di                                          # obtain rho_1
```

```
##        [,1]   [,2]
## [1,] 0.6602 0.7403
## [2,] 0.5603 0.8557
```

Implied Models for the Components

- What type of univariate dynamics does a VAR(1) imply for its components?

Example 2.9:

For the bivariate VAR(1) of Example 2.7 we have

$$\phi(L) = \begin{pmatrix} 1 - 0.2L & -0.3L \\ 0.6L & 1 - 1.1L \end{pmatrix}.$$

The **adjoint matrix** of $\phi(L)$ is (see Tsay, 2014, Appendix A)

$$\text{adj}(\phi(L)) = \begin{pmatrix} 1 - 1.1L & 0.3L \\ -0.6L & 1 - 0.2L \end{pmatrix}.$$

The product of these two matrices is

$$\text{adj}(\phi(L))\phi(L) = |\phi(L)|\boldsymbol{I}_2.$$

This is a diagonal matrix.

Implied Models for the Components

Example 2.9: Continued.

Hence, post-multiplying $\phi(L)\boldsymbol{z}_t = \boldsymbol{a}_t$ with $\text{adj}(\phi(L))$ gives

$$|\phi(L)|\boldsymbol{z}_t = \text{adj}(\phi(L))\boldsymbol{a}_t.$$

Plugging in gives

$$\begin{aligned} z_{1,t} - 1.3z_{1,t-1} + 0.4z_{1,t-2} &= a_{1,t} - 1.1a_{1,t-1} + 0.3a_{2,t-1}, \\ z_{2,t} - 1.3z_{2,t-1} + 0.4z_{2,t-2} &= -0.6a_{1,t} + a_{2,t} - 0.2a_{2,t-1}. \end{aligned}$$

Thus, for both components we obtain an ARMA(2,1) models, as the MA part has serial correlation at lag 1 only and can be rewritten as $e_t - \theta e_{t-1}$.

Implied Models for the Components

The above example illustrates the following more general result:

Theorem 2.10:

If $\boldsymbol{z}_t$ is a K-dimensional VAR(1), then each component z_{it} follows a univariate ARMA$(K, K-1)$ model.

Remark 2.11:

The orders K and $K-1$ are the maximum orders. The actual ARMA order for the individual z_{it} can be smaller.

VAR(p) Models

VAR(p) Model

- Consider the K-dimensional VAR(p) model

$$\boldsymbol{\phi}(L)\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t, \tag{2.5}$$

  where $\phi(L) = \boldsymbol{I}_K - \sum_{i=1}^{p} \boldsymbol{\phi}_i L^i$, $\boldsymbol{\phi}_p \neq 0$.

- If the VAR model is stationary, then taking expectations in (2.5) gives

$$\begin{aligned} & (\boldsymbol{I}_K - \boldsymbol{\phi}_1 - \ldots - \boldsymbol{\phi}_p)\boldsymbol{\mu} = \boldsymbol{\phi}_0 \\ \Longrightarrow \qquad & \boldsymbol{\mu} = (\boldsymbol{I}_K - \boldsymbol{\phi}_1 - \ldots - \boldsymbol{\phi}_p)^{-1}\boldsymbol{\phi}_0. \end{aligned}$$

- To derive conditions for stationarity, we express (2.5) as a VAR(1) to be able to apply previous results...

A VAR(1) Representation

- A K-dimensional VAR(p) can always be written as a Kp-dimensional VAR(1):

$$\boldsymbol{Z}_t = \phi_Z + \boldsymbol{\Phi}_1 \boldsymbol{Z}_{t-1} + \boldsymbol{A}_t, \tag{2.6}$$

where

$$\underset{(Kp\times 1)}{\boldsymbol{Z}_t} = (\boldsymbol{z}_t, \ldots, \boldsymbol{z}_{t-p+1})', \quad \underset{(Kp\times 1)}{\phi_Z} = (\phi_1, 0 \ldots, 0)',$$

$$\underset{(Kp\times Kp)}{\boldsymbol{\Phi}_1} = \begin{pmatrix} \phi_1 & \phi_1 & \ldots & \ldots & \phi_p \\ \boldsymbol{I}_K & \boldsymbol{0} & \ldots & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_K & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{0} & \ldots & \ldots & \boldsymbol{I}_K & \boldsymbol{0} \end{pmatrix}, \quad \underset{(Kp\times 1)}{\boldsymbol{A}_t} = \begin{pmatrix} \boldsymbol{a}_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Stationarity Condition

- Following the discussion for VAR(1) models, (2.6) is stationary if and only if

$$|\boldsymbol{I}_{Kp} - \boldsymbol{\Phi}_1 z| \neq 0 \qquad \text{for } |z| \leq 1.$$

- Since $|\boldsymbol{I}_{Kp} - \boldsymbol{\Phi}_1 z| = |\boldsymbol{I}_K - \phi_1 z - \ldots - \phi_p z^p|$ by Problem 2.1 in Lütkepohl (2005), (2.6) is stationary if and only if the **stability condition** holds:

$$|\boldsymbol{I}_K - \phi_1 z - \ldots - \phi_p z^p| \neq 0 \qquad \text{for } |z| \leq 1.$$

Theorem 2.12:
A VAR(p) models is stationary if and only if the stability condition holds.

Autocovariances of a Stationary VAR(p)

Yule–Walker Equations

- As in the VAR(1) case, the CCMs can be obtained either from moment equations or the causal representation.
- We first deal with the former possibility.
- Write (2.5) as

$$\boldsymbol{z}_t - \boldsymbol{\mu} = \boldsymbol{\phi}_1(\boldsymbol{z}_{t-1} - \boldsymbol{\mu}) + \ldots + \boldsymbol{\phi}_p(\boldsymbol{z}_{t-p} - \boldsymbol{\mu}) + \boldsymbol{a}_t.$$

- Post-multiply this by $(\boldsymbol{z}_{t-l} - \boldsymbol{\mu})'$ and take expectations to obtain the **Yule–Walker equations**

$$\boldsymbol{\Gamma}_l - \boldsymbol{\phi}_1\boldsymbol{\Gamma}_{l-1} - \ldots - \boldsymbol{\phi}_p\boldsymbol{\Gamma}_{l-p} = \begin{cases} \boldsymbol{\Sigma}_{\boldsymbol{a}}, & \text{if } l = 0; \\ \mathbf{0}, & \text{if } l > 0. \end{cases}$$

- From (2.6) we obtain similarly as in the VAR(1) case that

$$\left(\boldsymbol{I}_{(Kp)^2} - \boldsymbol{\Phi}_1 \otimes \boldsymbol{\Phi}_1\right) \mathrm{vec}(\boldsymbol{\Gamma}_0^*) = \mathrm{vec}(\boldsymbol{\Sigma}_{\boldsymbol{A}}),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{A}} = \mathrm{Var}(\boldsymbol{A}_t)$ and

$$\boldsymbol{\Gamma}_0^* = \begin{pmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \dots & \boldsymbol{\Gamma}_{p-1} \\ \boldsymbol{\Gamma}_1' & \boldsymbol{\Gamma}_0 & \dots & \boldsymbol{\Gamma}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{p-1}' & \boldsymbol{\Gamma}_{p-2}' & \dots & \boldsymbol{\Gamma}_0 \end{pmatrix}.$$

- Solving for $\boldsymbol{\Gamma}_0^*$ gives the $\boldsymbol{\Gamma}_0, \dots, \boldsymbol{\Gamma}_{p-1}$.
- Higher-order $\boldsymbol{\Gamma}_p, \boldsymbol{\Gamma}_{p+1}, \dots$ can now be obtained from the Yule–Walker equations on the previous slide.

Autocovariances of a Stationary VAR(p)

Causal Representation

- Consider a VAR(p) in lag-operator notation $\phi(L)\boldsymbol{z}_t = \phi_0 + \boldsymbol{a}_t$.
- If $\phi(L)$ is **invertible** (i.e., $|\phi(z)| \neq 0$ for $|z| \leq 1$), then there exists $\boldsymbol{\phi}^{-1}(L) := \sum_{i=0}^{\infty} \boldsymbol{\theta}_i L^i$, such that

$$\boldsymbol{\phi}^{-1}(L)\phi(L) = \boldsymbol{I}_K. \tag{2.7}$$

- Pre-multiplying by $\boldsymbol{\phi}^{-1}(L)$ gives the causal representation

$$\boldsymbol{z}_t = \left(\sum_{i=0}^{\infty} \boldsymbol{\theta}_i\right) \phi_0 + \sum_{i=0}^{\infty} \boldsymbol{\theta}_i \boldsymbol{a}_{t-i}, \tag{2.8}$$

where $\boldsymbol{\mu} = \left(\sum_{i=0}^{\infty} \boldsymbol{\theta}_i\right) \phi_0$.

- How do we obtain the coefficients $\boldsymbol{\theta}_i$?
- Use (2.7):

$$
\begin{aligned}
\boldsymbol{I}_K &= (\boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 L + \boldsymbol{\theta}_2 L^2 + \ldots)(\boldsymbol{I}_K - \boldsymbol{\phi}_1 L - \ldots - \boldsymbol{\phi}_p L^p) \\
&= \boldsymbol{\theta}_0 + (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0 \boldsymbol{\phi}_1) L + (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 \boldsymbol{\phi}_1 - \boldsymbol{\theta}_0 \boldsymbol{\phi}_2) L^2 + \ldots
\end{aligned}
$$

- Thus, the $\boldsymbol{\theta}_i$ can be computed recursively from

$$
\begin{cases}
\boldsymbol{\theta}_0 = \boldsymbol{I}_K, \\
\boldsymbol{\theta}_i = \sum_{j=1}^{i} \boldsymbol{\theta}_{i-j} \boldsymbol{\phi}_j, \quad i = 1, 2, \ldots
\end{cases}
\tag{2.9}
$$

Autocovariances of a Stationary VAR(p)

Causal Representation

- As before, the causal representation can be used to compute the CCMs.
- They can also be used to compute the variances of the forecast errors below.

Implied Component Models

- Theorem 2.10 can be generalized:

Theorem 2.13:

If $\boldsymbol{z}_t$ is a K-dimensional VAR(p), then each component z_{it} follows a univariate ARMA($Kp, (K-1)p$) model.

- The ARMA orders are again the maximum orders. In real-world applications, the orders of the univariate ARMA models are typically quite low, e.g., (1,1).
- It is an R exercise to show that using the multivariate VAR model improves forecast precision vis-à-vis using an implied component model.

Outline

Estimation of VAR Models

- Assume that $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T$ are observations from the VAR(p) model

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \ldots + \boldsymbol{\phi}_p \boldsymbol{z}_{t-p} + \boldsymbol{a}_t, \qquad t = p+1, \ldots, T. \tag{3.1}$$

- The parameters of interest are $\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p$ and $\boldsymbol{\Sigma}_{\boldsymbol{a}} = \mathrm{Var}(\boldsymbol{a}_1)$.
- In the following, we discuss two methods for estimation:

  1. Least squares (LS) estimation,
  2. Maximum likelihood (ML) estimation.

- To derive asymptotic properties, we often require

Definition 3.1: Standard White Noise.

A white noise process $\boldsymbol{a}_t$ is **standard white noise**, if the $\boldsymbol{a}_t$ are continuous random vectors with components having finite fourth moments.

1. Least Squares Estimation

LS Estimation

- Rewrite (3.1) as
$$\boldsymbol{Z} = \boldsymbol{X}\beta + \boldsymbol{A}, \tag{3.2}$$
where
$$\underset{(T-p)\times K}{\boldsymbol{Z}} = \begin{pmatrix} \boldsymbol{z}_{p+1}' \\ \vdots \\ \boldsymbol{z}_T' \end{pmatrix}, \quad \underset{(T-p)\times(Kp+1)}{\boldsymbol{X}} = \begin{pmatrix} 1 & \boldsymbol{z}_p' & \dots & \boldsymbol{z}_1' \\ \vdots & \vdots & & \vdots \\ 1 & \boldsymbol{z}_{T-1}' & \dots & \boldsymbol{z}_{T-p}' \end{pmatrix},$$
$$\underset{(T-p)\times K}{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{a}_{p+1}' \\ \vdots \\ \boldsymbol{a}_T' \end{pmatrix}, \quad \underset{(Kp+1)\times K}{\beta} = \begin{pmatrix} \phi_0' \\ \phi_1' \\ \vdots \\ \phi_p' \end{pmatrix}.$$
- Our first goal is to estimate β.

LS Estimation

- From (3.2) and Proposition 1.2.7, we get

$$\operatorname{vec}(\boldsymbol{Z}) = (\boldsymbol{I}_K \otimes \boldsymbol{X}) \operatorname{vec}(\beta) + \operatorname{vec}(\boldsymbol{A}). \tag{3.3}$$

- Since $\operatorname{Var}(\operatorname{vec}(\boldsymbol{A})) = \boldsymbol{\Sigma}_{\boldsymbol{a}} \otimes \boldsymbol{I}_{T-p}$, the generalized least squares (GLS) estimate of β is obtained by minimizing

$$\begin{aligned} S(\beta) &= \operatorname{vec}(\boldsymbol{A})'(\boldsymbol{\Sigma}_{\boldsymbol{a}} \otimes \boldsymbol{I}_{T-p})^{-1} \operatorname{vec}(\boldsymbol{A}) && (3.4)\\ &\overset{(3.2)}{=} \operatorname{vec}(\boldsymbol{Z} - \boldsymbol{X}\beta)'(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{I}_{T-p}) \operatorname{vec}(\boldsymbol{Z} - \boldsymbol{X}\beta) && (3.5)\\ &= \operatorname{tr}[(\boldsymbol{Z} - \boldsymbol{X}\beta)\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\beta)'], && (3.6) \end{aligned}$$

where we have also used Proposition 1.2.4 to obtain (3.5). The last equality holds because $\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}$ is a symmetric matrix and Proposition 1.2.9.

UNIVERSITÄT DUISBURG ESSEN

- From (3.5) and rules of Proposition 1.2,

$$\begin{aligned} S(\beta) &= [\mathrm{vec}(\boldsymbol{Z}) - (\boldsymbol{I}_K \otimes \boldsymbol{X})\,\mathrm{vec}(\beta)]'(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{I}_{T-p})[\mathrm{vec}(\boldsymbol{Z}) - (\boldsymbol{I}_K \otimes \boldsymbol{X})\,\mathrm{vec}(\beta)] \\ &= [\mathrm{vec}(\boldsymbol{Z})' - \mathrm{vec}(\beta)'(\boldsymbol{I}_K \otimes \boldsymbol{X}')](\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{I}_{T-p})[\mathrm{vec}(\boldsymbol{Z}) - (\boldsymbol{I}_K \otimes \boldsymbol{X})\,\mathrm{vec}(\beta)] \\ &= \mathrm{vec}(\boldsymbol{Z})'(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{I}_{T-p})\,\mathrm{vec}(\boldsymbol{Z}) - 2\,\mathrm{vec}(\beta)'(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{X}')\,\mathrm{vec}(\boldsymbol{Z}) \\ &\qquad + \mathrm{vec}(\beta)'(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\,\mathrm{vec}(\beta). \end{aligned}$$

- Taking partial derivatives with respect to $\mathrm{vec}(\beta)$ and using Proposition 1.4, we obtain

$$\frac{\partial S(\beta)}{\partial\,\mathrm{vec}(\beta)} = -2(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{X}')\,\mathrm{vec}(\boldsymbol{Z}) + 2(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\,\mathrm{vec}(\beta).$$

- Equating to $\mathbf{0}$ gives the normal equations for $\widehat{\beta}$, i.e.,

$$(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\,\mathrm{vec}(\widehat{\beta}) = (\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{X}')\,\mathrm{vec}(\boldsymbol{Z}).$$

- Consequently, the GLS estimate of a VAR(p) model is

$$\begin{aligned}
\operatorname{vec}(\widehat{\beta}) &= (\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{X}')\operatorname{vec}(\boldsymbol{Z}) \\
&= (\boldsymbol{\Sigma}_{\boldsymbol{a}} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1})(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{X}')\operatorname{vec}(\boldsymbol{Z}) && \text{[by Proposition 1.2.4]} \\
&= (\boldsymbol{I}_K \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\operatorname{vec}(\boldsymbol{Z}) && \text{[by Proposition 1.2.3]} \\
&= \operatorname{vec}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}) && \text{[by Proposition 1.2.7]}
\end{aligned} \tag{3.7}$$

- Hence, we obtain the **GLS estimator**

$$\boxed{\widehat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z},} \tag{3.8}$$

which—interestingly—does not depend on $\boldsymbol{\Sigma}_{\boldsymbol{a}}$.

LS Estimation

Discussion

Remark 3.2:

- The result in (3.7) shows that the GLS estimate can also be obtained from equation-by-equation LS, i.e., from the K linear regressions of z_{it} on $(1, \boldsymbol{z}'_{t-1}, \ldots, \boldsymbol{z}'_{t-p})$. Showing this is left as an exercise; see Lütkepohl (2005, p. 72).
- Replacing $\boldsymbol{\Sigma}_{\boldsymbol{a}}$ (corresponding to GLS) with $\boldsymbol{I}_K$ (corresponding to OLS) in (3.4) leads to an identical $\widehat{\boldsymbol{\beta}}$ in (3.8). Thus, OLS $=$ GLS in VAR(p) models. This is a classic result due to Zellner (1962).

LS Estimation

- It remains to obtain an estimate of $\boldsymbol{\Sigma_a}$.
- The LS **residuals** are

$$\widehat{\boldsymbol{a}}_t = \boldsymbol{z}_t - \widehat{\boldsymbol{\phi}}_0 - \sum_{i=1}^{p} \widehat{\boldsymbol{\phi}}_i \boldsymbol{z}_{t-i}, \qquad t = p+1, \ldots, T. \tag{3.9}$$

- The **residual matrix** is $\widehat{\boldsymbol{A}} = \boldsymbol{Z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$.
- The LS estimate of $\boldsymbol{\Sigma_a}$ is

$$\boxed{\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{a}} = \frac{1}{T-p} \sum_{t=p+1}^{T} \widehat{\boldsymbol{a}}_t \widehat{\boldsymbol{a}}_t' = \frac{1}{T-p} \widehat{\boldsymbol{A}}' \widehat{\boldsymbol{A}}.}$$

Theorem 3.3: Lütkepohl (2005), Proposition 3.1.

Let $\boldsymbol{z}_t$ be a stationary, K-dimensional VAR(p) with standard white noise $\boldsymbol{a}_t$. Then,

$$\sqrt{T}\left[\mathrm{vec}(\widehat{\beta}) - \mathrm{vec}(\beta)\right] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}} \otimes \boldsymbol{G}^{-1}), \qquad \text{as } T \to \infty,$$

where $\boldsymbol{G} = \operatorname{plim} \boldsymbol{X}\boldsymbol{X}'/T$, i.e., $\boldsymbol{X}\boldsymbol{X}'/T \xrightarrow{p} \boldsymbol{G}$ as $T \to \infty$.

- To obtain an estimate of the asymptotic variance $\boldsymbol{\Sigma}_{\boldsymbol{a}} \otimes G^{-1}$, we only require a consistent estimate of $\boldsymbol{\Sigma}_{\boldsymbol{a}}$, since $\boldsymbol{G}$ can be consistently estimated by $\widehat{\boldsymbol{G}} = \boldsymbol{X}\boldsymbol{X}'/T$.

Proposition 3.4: Lütkepohl (2005), Corollary 3.2.1.

Under the conditions of Theorem 3.3,

$$\operatorname{plim} \widehat{\boldsymbol{\Sigma}}_a = \boldsymbol{\Sigma}_{\boldsymbol{a}}.$$

- Together, Theorem 3.3 and Proposition 3.4 imply that t-ratios behave as usual:

$$\frac{\widehat{\beta}_i - \beta_i}{\widehat{s}_i} \xrightarrow{d} N(0,1), \tag{3.10}$$

where β_i ($\widehat{\beta}_i$) is the i-th component of $\mathrm{vec}(\beta)$ ($\mathrm{vec}(\widehat{\beta})$) and $\widehat{s}_i$ is the square root of the i-th diagonal element of

$$\widehat{\boldsymbol{\Sigma}}_a \otimes (\boldsymbol{X}\boldsymbol{X}')^{-1}.$$

- In small samples, one typically finds that instead of the $N(0,1)$-distribution, the $t_{T-p-Kp-1}$-distribution provides a better fit in (3.10).
  - $T - p =$ number of effective observations
  - $Kp + 1 =$ number of parameters in each equation-by-equation regression of z_{it} on $(1, \boldsymbol{z}'_{t-1}, \ldots, \boldsymbol{z}'_{t-p})$

2. Maximum Likelihood Estimation

- Recall **maximum likelihood estimation**: Given a sequence of realizations $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T$ of i.i.d. draws from a distribution with density $f(\boldsymbol{z}; \boldsymbol{\theta})$, the ML estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \prod_{t=1}^{T} f(\boldsymbol{z}_t; \boldsymbol{\theta}). \tag{3.11}$$

- Hence, the idea is to choose the model (here: simply a parameter) with the highest likelihood of the realizations $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T$ occurring.

ML Estimation

However, two problems render estimation of VAR(p) processes via ML infeasible.

1. We do not have a closed form for the density $f(\boldsymbol{z}; \boldsymbol{\theta} = (\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p, \boldsymbol{\Sigma}_{\boldsymbol{a}})')$ in general.

2. Realizations of a VAR(p) process are not i.i.d. Thus, the factorization of the densities in (3.11) does not hold.

In the following we shall see that **conditional maximum likelihood estimation** is possible.

ML Estimation

- Recall that for random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ the conditional density is given by

$$f_{\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})}{f_{\boldsymbol{Y}}(\boldsymbol{y})}, \qquad \text{if } f_{\boldsymbol{Y}}(\boldsymbol{y}) > 0.$$

- Denote by $\boldsymbol{z}_{h:q}$ the observations from $t = h$ to $t = q$ (inclusive).
- Thus, for (not necessarily i.i.d.) realizations $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T$ the **conditional likelihood** is given by

$$f_{\boldsymbol{z}_{(p+1):T}|\boldsymbol{z}_{1:p}}(\boldsymbol{z}_{p+1}, \ldots, \boldsymbol{z}_T) = \prod_{t=p+1}^{T} f_{\boldsymbol{z}_t|\boldsymbol{z}_{1:(t-1)}}(\boldsymbol{z}_t) \tag{3.12}$$

- Hence, while a factorization of the joint density as in (3.11) is only obtained for independent data, a product factorization of the joint *conditional* density also holds for *non*-independent data.

ML Estimation

- Thus, in the following we seek to maximize the **conditional likelihood**

$$L(\beta, \boldsymbol{\Sigma}_{\boldsymbol{a}}) = \prod_{t=p+1}^{T} f_{\boldsymbol{z}_t|\boldsymbol{z}_{1:(t-1)}}(\boldsymbol{z}_t;\ \beta, \boldsymbol{\Sigma}_{\boldsymbol{a}}) \tag{3.13}$$

- The **maximum likelihood estimator** of $(\beta, \boldsymbol{\Sigma}_{\boldsymbol{a}})$ is then

$$\boxed{(\widehat{\beta}, \widehat{\boldsymbol{\Sigma}}_a) = \underset{(\beta, \boldsymbol{\Sigma}_{\boldsymbol{a}})}{\arg\max}\, L(\beta, \boldsymbol{\Sigma}_{\boldsymbol{a}}).}$$

- Thus, we seek an expression for $f_{\boldsymbol{z}_t|\boldsymbol{z}_{1:(t-1)}}(\boldsymbol{x};\ \beta, \boldsymbol{\Sigma}_{\boldsymbol{a}})$ in (3.13).
- To obtain an expression for $f_{\boldsymbol{z}_t|\boldsymbol{z}_{1:(t-1)}}(\boldsymbol{x};\ \beta, \boldsymbol{\Sigma}_{\boldsymbol{a}})$ we assume for the moment that

$$\boldsymbol{a}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}}).$$

ML Estimation

- We calculate $f_{\boldsymbol{z}_t|\boldsymbol{z}_{1:(t-1)}}(\boldsymbol{x};\ \boldsymbol{\beta}, \boldsymbol{\Sigma_a})$ by first calculating the distribution function (d.f.):

$$\begin{aligned}\mathrm{P}\{\boldsymbol{z}_t \leq \boldsymbol{x} \mid \boldsymbol{z}_{1:(t-1)}\} &= \mathrm{P}\{\boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \ldots + \boldsymbol{\phi}_p \boldsymbol{z}_{t-p} + \boldsymbol{a}_t \leq \boldsymbol{x} \mid \boldsymbol{z}_{1:(t-1)}\} \\ &= \mathrm{P}\{\boldsymbol{a}_t \leq \boldsymbol{x} - \boldsymbol{\phi}_0 - \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} - \ldots - \boldsymbol{\phi}_p \boldsymbol{z}_{t-p} \mid \boldsymbol{z}_{1:(t-1)}\} \\ &= \boldsymbol{\Phi}(\boldsymbol{x} - \boldsymbol{\phi}_0 - \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} - \ldots - \boldsymbol{\phi}_p \boldsymbol{z}_{t-p};\ \boldsymbol{\Sigma_a}),\end{aligned}$$

  where $\boldsymbol{\Phi}(\,\cdot\,;\ \boldsymbol{\Sigma_a})$ is the d.f. of a $N(\mathbf{0}, \boldsymbol{\Sigma_a})$-distribution.

- Taking the derivative of the d.f. with respect to $\boldsymbol{x}$ gives the desired density

$$f_{\boldsymbol{z}_t|\boldsymbol{z}_{1:(t-1)}}(\boldsymbol{x};\ \boldsymbol{\beta}, \boldsymbol{\Sigma_a}) = \phi(\boldsymbol{x} - \boldsymbol{\phi}_0 - \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} - \ldots - \boldsymbol{\phi}_p \boldsymbol{z}_{t-p};\ \boldsymbol{\Sigma_a}),$$

  where $\phi(\,\cdot\,;\ \boldsymbol{\Sigma_a})$ is the density function of a $N(\mathbf{0}, \boldsymbol{\Sigma_a})$-distribution.

ML Estimation

- Plugging the expression for $f_{\mathbf{z}_t|\mathbf{z}_{1:(t-1)}}(\boldsymbol{x};\ \beta, \boldsymbol{\Sigma}_{\boldsymbol{a}})$ in (3.13) gives

$$\begin{aligned} L(\beta, \boldsymbol{\Sigma}_{\boldsymbol{a}}) &= \prod_{t=p+1}^{T} \phi(\boldsymbol{a}_t;\ \boldsymbol{\Sigma}_{\boldsymbol{a}}) \\ &= \prod_{t=p+1}^{T} \frac{1}{(2\pi)^{-\frac{K}{2}(T-p)}|\boldsymbol{\Sigma}_{\boldsymbol{a}}|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{a}_t'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}\boldsymbol{a}_t\right\} \\ &= (2\pi)^{-\frac{K}{2}(T-p)}|\boldsymbol{\Sigma}_{\boldsymbol{a}}|^{-\frac{T-p}{2}} \exp\left\{-\frac{1}{2}\sum_{t=p+1}^{T} \operatorname{tr}(\boldsymbol{a}_t'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}\boldsymbol{a}_t)\right\}. \end{aligned}$$

- With $c = -\frac{K}{2}(T-p)\log(2\pi)$, the conditional log-likelihood is

$$l(\beta, \boldsymbol{\Sigma}_{\boldsymbol{a}}) = c - \frac{T-p}{2}\log(|\boldsymbol{\Sigma}_{\boldsymbol{a}}|) - \frac{1}{2}\sum_{t=p+1}^{T} \operatorname{tr}(\boldsymbol{a}_t'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}\boldsymbol{a}_t).$$

ML Estimation

- Using

$$\begin{aligned}\sum_{t=p+1}^{T}\operatorname{tr}(\boldsymbol{a}_t'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}\boldsymbol{a}_t) &= \operatorname{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}\sum_{t=p+1}^{T}\boldsymbol{a}_t'\boldsymbol{a}_t\right) && \text{[by Proposition 1.2.11]}\\ &= \operatorname{tr}(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}\boldsymbol{A}'\boldsymbol{A}) && \text{[by (3.2)]}\\ &= \operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}\boldsymbol{A}') && \text{[by Proposition 1.2.10]}\\ &= S(\boldsymbol{\beta}) && \text{[by (3.6)]}\end{aligned}$$

- Plugging this into the expression for the log-likelihood gives

$$l(\boldsymbol{\beta},\boldsymbol{\Sigma}_{\boldsymbol{a}}) = c - \frac{T-p}{2}\log(|\boldsymbol{\Sigma}_{\boldsymbol{a}}|) - \frac{1}{2}S(\boldsymbol{\beta}). \tag{3.14}$$

- Since maximizing $l(\boldsymbol{\beta},\boldsymbol{\Sigma}_{\boldsymbol{a}})$ in $\boldsymbol{\beta}$ is equivalent to minimizing $S(\boldsymbol{\beta})$ the ML estimate $\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} l(\boldsymbol{\beta},\boldsymbol{\Sigma}_{\boldsymbol{a}})$ is equal to the LS estimate.

- To obtain the ML estimate of $\boldsymbol{\Sigma_a}$, we take the partial derivative of $l(\widehat{\boldsymbol{\beta}}, \boldsymbol{\Sigma_a})$ with respect to $\boldsymbol{\Sigma_a}$:

$$\begin{aligned}
\frac{\partial l(\widehat{\boldsymbol{\beta}}, \boldsymbol{\Sigma_a})}{\partial \boldsymbol{\Sigma_a}} &= -\frac{T-p}{2}\frac{\partial \log(|\boldsymbol{\Sigma_a}|)}{\partial \boldsymbol{\Sigma_a}} - \frac{1}{2}\frac{\partial \operatorname{tr}(\widehat{\boldsymbol{A}}\boldsymbol{\Sigma_a}^{-1}\widehat{\boldsymbol{A}}')}{\partial \boldsymbol{\Sigma_a}} \\
&= -\frac{T-p}{2}(\boldsymbol{\Sigma_a}')^{-1} + \frac{1}{2}(\boldsymbol{\Sigma_a}^{-1}\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{A}}\boldsymbol{\Sigma_a}^{-1})' \\
&= -\frac{T-p}{2}\boldsymbol{\Sigma_a}^{-1} + \frac{1}{2}\boldsymbol{\Sigma_a}^{-1}\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{A}}\boldsymbol{\Sigma_a}^{-1},
\end{aligned}$$

  where the second equality follows from Proposition 1.4.4 and 1.4.5 and the third by recalling that $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$ and $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$.

- Equating to $\mathbf{0}$, the ML estimate is (as before)

$$\widehat{\boldsymbol{\Sigma}}_a = \frac{1}{T-p}\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{A}} = \frac{1}{T-p}\sum_{t=p+1}^{T}\widehat{\boldsymbol{a}}_t\widehat{\boldsymbol{a}}_t'. \tag{3.15}$$

- Since ML=LS, the same asymptotic properties as in Theorem 3.3 and Proposition 3.4 hold.
- Thus, the innovations do not even have to be normal for the ML estimator to work.
- This is a strong robustness result for ML!

ML Estimation I

GDP Growth in R

Example 3.5:

Consider the GDP growth rates $\boldsymbol{z}_t = (z_{1t}, z_{2t}, z_{3t})'$ of the UK, Canada and the US from 1980Q2 to 2011Q2. They have been calculated as

$$z_{it} = \log(GDP_t^i / GDP_{t-1}^i),$$

where GDP^i denotes the GDP in millions of local currency. We estimate a VAR(2) with $K = 3$ and $T = 125$. It is left as an exercise to compute the LS estimates by hand; see Tsay (2014, p. 51).

```
da  = read.table("Resources/Data/q-gdp-ukcaus.txt", header=T) # red in data
head(da)
```

ML Estimation II

GDP Growth in R

```
##   year mon     uk     ca      us
## 1 1980   1 172436 624794 5908500
## 2 1980   4 169359 623433 5787400
## 3 1980   7 169038 623215 5776600
## 4 1980  10 167180 630215 5883500
## 5 1981   1 166052 645957 6005700
## 6 1981   4 166393 651954 5957800
```

```r
gdp = log(da[, 3:5])
z   = diffM(gdp)                                  # difference log GDP
z   = z*100
dim(z)
```

```
## [1] 125   3
```

```r
m1  = VAR(z,2)                                    # Use VAR command in MTS
```

ML Estimation III

GDP Growth in R

```
## Constant term:
## Estimates:  0.1258 0.1232 0.2896
## Std.Error:  0.07266 0.07383 0.08169
## AR coefficient matrix
## AR( 1 )-matrix
##       [,1]  [,2]   [,3]
## [1,] 0.393 0.103 0.0521
## [2,] 0.351 0.338 0.4691
## [3,] 0.491 0.240 0.2356
## standard error
##        [,1]   [,2]   [,3]
## [1,] 0.0934 0.0984 0.0911
## [2,] 0.0949 0.1000 0.0926
## [3,] 0.1050 0.1106 0.1024
## AR( 2 )-matrix
##         [,1]   [,2]     [,3]
## [1,]  0.0566  0.106  0.01889
## [2,] -0.1914 -0.175 -0.00868
## [3,] -0.3120 -0.131  0.08531
## standard error
##        [,1]   [,2]   [,3]
## [1,] 0.0924 0.0876 0.0938
## [2,] 0.0939 0.0890 0.0953
```

ML Estimation IV

GDP Growth in R

```
## [3,] 0.1038 0.0984 0.1055
##
## Residuals cov-mtx:
##          [,1]    [,2]    [,3]
## [1,] 0.28244 0.02654 0.07435
## [2,] 0.02654 0.29158 0.13949
## [3,] 0.07435 0.13949 0.35697
##
## det(SSE) =  0.02259
## AIC =  -3.502
## BIC =  -3.095
## HQ  =  -3.337
```

■

Outline

Model Selection and Checking

In this section, we learn how to

1. choose the order p of a VAR model in practice,
2. check the adequacy of the fitted model (with order p from 1.), and
3. simplify the adequate fitted model from 2.

1. Order Selection

Order Selection

- Just like the model parameters in β, the order p of a VAR model is typically unknown.
- To choose p, we use **information criteria**.
- The idea is to trade off model fit and model complexity.
  - A popular measure for model fit is the maximized log-likelihood of a VAR(l), i.e.,

$$l(\widehat{\beta}, \widehat{\Sigma}_a) = c - \frac{T-p}{2} \log |\widehat{\Sigma}_{a,l}| - K\frac{T-p}{2}.$$

  - Model complexity is measured by the number of parameters.

Definition 4.1:

For a K-dimensional VAR(l) model the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the Hannan and Quinn (HQ) criterion are

$$\begin{aligned} \mathrm{AIC}(l) &= \log|\widehat{\boldsymbol{\Sigma}}_{a,l}| + \frac{2}{T} lK^2, \\ \mathrm{BIC}(l) &= \log|\widehat{\boldsymbol{\Sigma}}_{a,l}| + \frac{\log T}{T} lK^2, \\ \mathrm{HQ}(l) &= \log|\widehat{\boldsymbol{\Sigma}}_{a,l}| + 2\frac{\log(\log T)}{T} lK^2, \end{aligned}$$

where T is the sample size, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{a},l}$ is the ML estimate of $\boldsymbol{\Sigma}_{\boldsymbol{a}}$ in (3.15).

- AIC was introduced by Akaike (1973), BIC by Schwarz (1978) and HQ by Hannan and Quinn (1979).

Remark 4.2:

- If P is the maximum order, one should use the same number of observations $\boldsymbol{z}_{P+1}, \ldots, \boldsymbol{z}_T$ to evaluate the likelihood for all VAR(l) models, $l = 0, 1, \ldots, P$.

- The order $\widehat{p} \in \{0, 1, \ldots, P\}$ producing the lowest value of AIC, BIC or HQ is preferred.

- AIC penalizes each order by a factor of 2. BIC and HQ, on the other hand, employ penalties that depend on the sample size.

Definition 4.3:

An estimator $\widehat{p}$ of the VAR order p is **consistent** if

$$\lim_{T\to\infty} \mathrm{P}\{\widehat{p} = p\} = 1.$$

Theorem 4.4: Lütkepohl (2005), Proposition 4.2.

Let $\boldsymbol{z}_t$ be a stationary, K-dim VAR(p) with standard white noise $\boldsymbol{a}_t$. Suppose the maximum order $P \geq p$ and $\widehat{p}$ is chosen to minimize

$$C(l) = \log|\widehat{\boldsymbol{\Sigma}_{a,l}}| + \frac{l}{T} c_T, \qquad l = 0, 1, \ldots, P.$$

Then, $\widehat{p}$ is consistent if and only if

$$c_T \to \infty \quad \text{and} \quad \frac{c_T}{T} \to 0 \quad \text{as } T \to \infty.$$

Corollary 4.5:

Under the conditions of Theorem 4.4,

- $\widehat{p}(\text{AIC})$ is not consistent,
- $\widehat{p}(\text{BIC})$ and $\widehat{p}(\text{HQ})$ are consistent.

The proof is left as an exercise.

Order Selection I

GDP Growth in R

Example 4.6:

Reconsider Example 3.5. We select the order of a VAR for the growth rates. The maximum order entertained is $P = 14$. This example demonstrates that different information criteria may select different orders. Keep in mind that the selected orders are estimates; see the plot below!

```
m2 = VARorder(z)
```

```
## selected order: aic =  2
## selected order: bic =  1
## selected order: hq =  1
## Summary table:
##         p    AIC    BIC     HQ    M(p) p-value
##  [1,]  0 -3.325 -3.325 -3.325   0.000  0.0000
##  [2,]  1 -4.252 -4.048 -4.169 115.133  0.0000
##  [3,]  2 -4.333 -3.926 -4.168  23.539  0.0051
##  [4,]  3 -4.293 -3.682 -4.044  10.486  0.3126
##  [5,]  4 -4.266 -3.452 -3.935  11.577  0.2382
##  [6,]  5 -4.151 -3.133 -3.737   2.741  0.9737
##  [7,]  6 -4.080 -2.858 -3.584   6.782  0.6598
##  [8,]  7 -3.987 -2.562 -3.408   4.547  0.8719
```

Order Selection II
GDP Growth in R

```
## [9,]  8 -4.126 -2.497 -3.464  24.483  0.0036
## [10,]  9 -4.059 -2.226 -3.314   6.401  0.6992
## [11,] 10 -3.968 -1.932 -3.141   4.323  0.8889
## [12,] 11 -3.973 -1.733 -3.063  11.492  0.2435
## [13,] 12 -3.987 -1.544 -2.994  11.817  0.2238
## [14,] 13 -4.041 -1.393 -2.965  14.127  0.1179
```

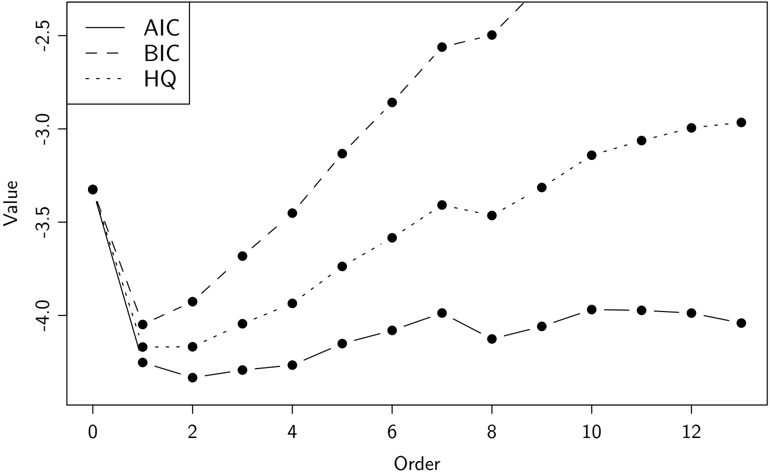
```r
names(m2)
```

```
## [1] "aic"   "aicor" "bic"   "bicor" "hq"    "hqor"  "Mstat" "Mpv"
```

```r
plot(0:13, m2$aic,  type="b", lty=1, ylim = c(-4.4, -2.4),
                                          xlab="Order", ylab="Value")
lines(0:13, m2$bic, type="b", lty=2)
lines(0:13, m2$hq,  type="b", lty=3)
legend("topleft", legend=c("AIC", "BIC", "HQ"), lty=1:3)
```

Order Selection III

GDP Growth in R



2. Model Checking

Model Checking

- Model checking plays an important role in model building.
- Its main objectives are
  1. to ensure that the fitted model is adequate and
  2. to suggest directions for further improvements if needed.
- A fitted model is **adequate** if
  1. all estimated parameters are statistically significant,
  2. the residuals have no significant serial cross-correlations,
  3. there exist no outlying observations, and
  4. the residuals do not violate distributional assumptions (e.g., finite 4-th moments).

Model Checking

- The residuals of an adequate model should be close to i.i.d.
- Thus, we check for residual cross-correlations
- The lag-l cross-covariance matrix of the residuals $\widehat{\boldsymbol{a}}_t$ from (3.9) is

$$\widehat{\boldsymbol{C}}_l = \frac{1}{T-p} \sum_{t=p+l+1}^{T} \widehat{\boldsymbol{a}}_t \widehat{\boldsymbol{a}}'_{t-l}.$$

- In particular, we have $\widehat{\boldsymbol{C}}_0 = \widehat{\boldsymbol{\Sigma}}_a$.
- The lag-l CCM is thus

$$\widehat{\boldsymbol{R}}_l = \widehat{\boldsymbol{D}}^{-1} \widehat{\boldsymbol{C}}_l \widehat{\boldsymbol{D}}^{-1},$$

where $\widehat{\boldsymbol{D}} = \sqrt{\mathrm{diag}(\widehat{\boldsymbol{C}}_0)}$ is the diagonal matrix of the residual standard errors.

Model Checking

- Let $\boldsymbol{R}_l$ be the theoretical lag-l CCM of the innovations $\boldsymbol{a}_t$.
- The hypothesis of interest in model checking is

$$\boxed{H_0: \ \boldsymbol{R}_1 = \ldots = \boldsymbol{R}_m = \boldsymbol{0} \quad \text{vs.} \quad H_1: \ \boldsymbol{R}_j \neq \boldsymbol{0} \text{ for some } 1 \leq j \leq m,}$$

  where m is a prespecified integer.
- The Ljung–Box test statistic is

$$\boxed{Q_K(m) = T^2 \sum_{l=1}^{m} \frac{1}{T-l} \widehat{\boldsymbol{B}}_l' \left(\widehat{\boldsymbol{R}}_0^{-1} \otimes \widehat{\boldsymbol{R}}_0^{-1} \right) \widehat{\boldsymbol{B}}_l,}$$

  where $\widehat{\boldsymbol{B}}_l = \operatorname{vec}(\widehat{\boldsymbol{R}}_l')$.

Theorem 4.7: Tsay (2014), Theorem 2.6.

Let $\boldsymbol{z}_t$ be a K-dimensional VAR(p) with standard white noise $\boldsymbol{a}_t$. Then,

$$Q_K(m) \xrightarrow{d} \chi^2_{(m-p)K^2}, \qquad \text{as } T \to \infty.$$

Remark 4.8:

- Compared with Corollary 1.20, the degrees of freedom of the χ^2-distribution is adjusted by

$$pK^2 = \text{number of AR parameters in a VAR}(p).$$

- If some parameters are set to zero, the adjustment is set to the number of estimated parameters.

Example 4.9:

Reconsider Example 3.5. We apply the Pormanteau test of Theorem 4.7 to check for model adequacy of the fitted VAR($p = 2$). Note that the degrees of freedom $(m - p)K^2$ are only defined if $m > p = 2$. Overall, the VAR(2) captures dynamic dependence of growth series, except for some minor violations at $m = 4$.

```
names(m1)
```

```
##  [1] "data"      "cnst"      "order"     "coef"      "aic"
##  [6] "bic"       "hq"        "residuals" "secoef"    "Sigma"
## [11] "Phi"       "Ph0"       "fixed"
```

```
resi = m1$residuals         # residuals of VAR(2) fit
mq(resi, lag=20, adj=18)    # adjustment of 18=2*3^2=pK^2 for degrees of freedom
```

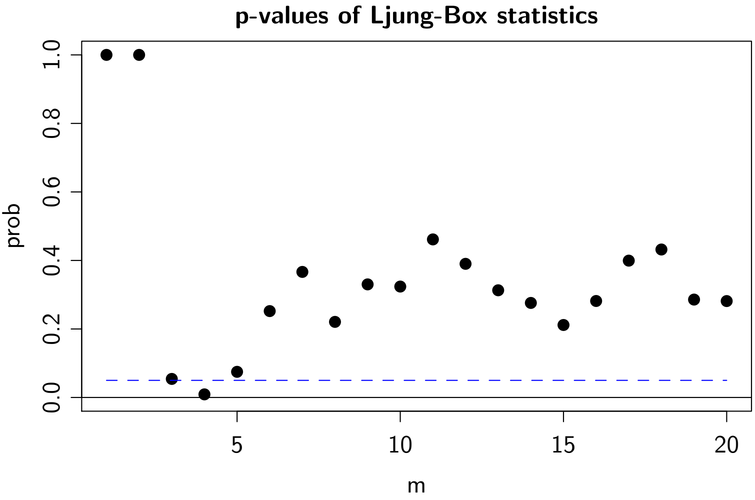
Model Checking II

GDP Growth in R

```
## Ljung-Box Statistics:
##             m       Q(m)     df    p-value
##  [1,]    1.000     0.816  -9.000      1.00
##  [2,]    2.000     3.978   0.000      1.00
##  [3,]    3.000    16.665   9.000      0.05
##  [4,]    4.000    35.122  18.000      0.01
##  [5,]    5.000    38.189  27.000      0.07
##  [6,]    6.000    41.239  36.000      0.25
##  [7,]    7.000    47.621  45.000      0.37
##  [8,]    8.000    61.677  54.000      0.22
##  [9,]    9.000    67.366  63.000      0.33
## [10,]   10.000    76.930  72.000      0.32
## [11,]   11.000    81.567  81.000      0.46
## [12,]   12.000    93.112  90.000      0.39
## [13,]   13.000   105.327  99.000      0.31
## [14,]   14.000   116.279 108.000      0.28
## [15,]   15.000   128.974 117.000      0.21
## [16,]   16.000   134.704 126.000      0.28
## [17,]   17.000   138.552 135.000      0.40
## [18,]   18.000   146.256 144.000      0.43
## [19,]   19.000   162.418 153.000      0.29
## [20,]   20.000   171.948 162.000      0.28
```

Model Checking III

GDP Growth in R



3. Model Simplification

Model Simplification

- Multivariate time series models may contain many parameters if the dimension K is moderate or large.
  - A K-dimensional VAR(p) has $K + K^2 p$ free parameters, which may quickly approach the total number of observations, KT, and thus exhaust degrees of freedom.
- In practice, some parameters may not be significant.
- Then, it is advantageous to remove these to simplify the model and to improve the estimation precision of the remaining parameter.
  - This is particularly so when no prior knowledge supports these parameters.
- Next, we discuss some methods for model simplification.

Model Simplification
Testing Zero Parameters

- First, we identify **target parameters** for removal as those parameters whose *individual* t-ratios (see (3.10)) are smaller in absolute value than the α-critical value (e.g., 1.96 for $\alpha = 0.05$).
- To test whether these parameters are insignificant *jointly*, we employ a test based on Theorem 3.3.
- Let

$$\widehat{\omega} = v\text{-dimensional vector containing target parameters,}$$
$$\omega = \text{counterpart of } \widehat{\omega} \text{ in the parameter matrix } \boldsymbol{\beta} \text{ from (3.2).}$$

- The hypothesis of interest is

$$H_0: \ \boldsymbol{\omega} = \mathbf{0} \qquad \text{vs.} \qquad H_a: \ \boldsymbol{\omega} \neq \mathbf{0}. \tag{4.1}$$

- Clearly, there exists a $v \times K(Kp+1)$-**locating matrix** $\boldsymbol{K}$ such that

$$\boldsymbol{K}\,\mathrm{vec}(\boldsymbol{\beta}) = \boldsymbol{\omega} \qquad \text{and} \qquad \boldsymbol{K}\,\mathrm{vec}(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{\omega}}.$$

- By Theorem 3.3 and porperties of the multivariate normal distribution,

$$\sqrt{T-p}(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{K}(\boldsymbol{\Sigma_a} \otimes \boldsymbol{G}^{-1})\boldsymbol{K}').$$

- Hence, we obtain

Corollary 4.10: Tsay (2014), Section 2.7.3.

Under the conditions of Theorem 3.3 and $H_0: \ \boldsymbol{\omega} = \mathbf{0}$, the **Wald statistic**

$$\lambda = (T-p)\widehat{\boldsymbol{\omega}}' \left[\boldsymbol{K}\left\{\widehat{\boldsymbol{\Sigma}}_a \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right\}\boldsymbol{K}'\right]^{-1}\widehat{\boldsymbol{\omega}} \xrightarrow{d} \chi^2_v, \qquad \text{as } T \to \infty.$$

Example 4.11:

Reconsider Example 3.5, which shows that several parameters in the fitted VAR(2) are insignificant. To simplify the model, we apply Corollary 4.10 with $\alpha = 0.05$ (and thus critical value of 1.96). The *p*-value of 0.056 suggests that the 8 targeted parameters with the smallest (in absolute value) *t*-ratios are 0. Consequently we can simplify the VAR(2) model by setting these 8 parameters to 0.

```
m2 = VARchi(z, p=2, thres=1.645)
```

```
## Number of targeted parameters:  8
## Chi-square test and p-value:  15.16 0.05604
```

```
m3 = VARchi(z, p=2, thres=1.96)
```

```
## Number of targeted parameters:  10
## Chi-square test and p-value:  31.69 0.0004514
```

■

Model Simplification

Information Criteria

- An alternative approach to the Wald-test of Corollary 4.10 is to use information criteria $\mathrm{IC} \in \{\mathrm{AIC}, \mathrm{BIC}, \mathrm{HQ}\}$.
- To do so, estimate the
  1. unconstrained VAR(p) model under H_a and the
  2. constrained VAR(p) model under H_0.
- If $\mathrm{IC}_{\mathrm{constr}} < \mathrm{IC}_{\mathrm{unconstr}}$, then H_0 is not rejected.

Example 4.11: Continued.

We use the `refVAR` command of the `MTS` package, which uses a critical value (`thres`) to select the target parameters and computes ICs of the simplified models. Since $\text{AIC}_{\text{constr}} = -3.53 < -3.50 = \text{AIC}_{\text{unconstr}}$ (see also Example 3.5), the simplified model is preferred.

```r
m.ref = refVAR(m1, thres=1.96)   # model refinement
```

```
## Constant term:
## Estimates:  0.1628 0 0.2828
## Std.Error:  0.06814 0 0.07973
## AR coefficient matrix
## AR( 1 )-matrix
##       [,1]  [,2]  [,3]
## [1,] 0.467 0.207 0.000
## [2,] 0.334 0.270 0.496
## [3,] 0.468 0.225 0.232
## standard error
##        [,1]   [,2]   [,3]
## [1,] 0.0790 0.0686 0.0000
```

Model Simplification II

GDP Growth in R

```
## [2,] 0.0921 0.0875 0.0913
## [3,] 0.1027 0.0963 0.1023
## AR( 2 )-matrix
##        [,1] [,2] [,3]
## [1,]  0.000    0    0
## [2,] -0.197    0    0
## [3,] -0.301    0    0
## standard error
##        [,1] [,2] [,3]
## [1,] 0.0000    0    0
## [2,] 0.0921    0    0
## [3,] 0.1008    0    0
##
## Residuals cov-mtx:
##         [,1]    [,2]    [,3]
## [1,] 0.29004 0.01803 0.07056
## [2,] 0.01803 0.30803 0.14598
## [3,] 0.07056 0.14598 0.36269
##
## det(SSE) =  0.02494
## AIC =  -3.531
## BIC =  -3.305
## HQ  =  -3.439
```

Example 4.11: Continued.

The refined VAR(2) must also be checked with `MTSdiag(m.ref, adj=12)`, since there are now only 12 parameters in the fitted model. The final refined VAR(2) is

$$\boldsymbol{z}_t = \begin{pmatrix} 0.16 \\ - \\ 0.28 \end{pmatrix} + \begin{pmatrix} 0.47 & 0.21 & - \\ 0.33 & 0.27 & 0.50 \\ 0.47 & 0.23 & 0.23 \end{pmatrix} \boldsymbol{z}_{t-1} + \begin{pmatrix} - & - & - \\ -0.20 & - & - \\ -0.30 & - & - \end{pmatrix} \boldsymbol{z}_{t-2} + \boldsymbol{a}_t,$$

where the residual correlation matrix is

$$\boldsymbol{R}_0 = \begin{pmatrix} 1.00 & 0.06 & 0.22 \\ 0.06 & 1.00 & 0.44 \\ 0.22 & 0.44 & 1.00 \end{pmatrix}.$$

Example 4.11: Continued.

The fitted simplified VAR(2) shows that

- GDP growth in the UK and Canada are not instantaneously correlated.
- GDP growth in the UK does not depend on past US growth in the presence of lagged Canadian growth.
- GDP growth in Canada (US) is related to growth in the UK and US (Canada).

Note that even though 10 parameters were targeted, only 9 were set to zero in the refined VAR(2). This is because the 10 parameters are significant jointly with p-value 0.0004514, but the 9 parameters set to zero are not jointly significant.

- Of course, we have to re-estimate the simplified model
- Any linear parameter constraint (in particular $\boldsymbol{\omega} = \mathbf{0}$ in (4.1)) can be expressed as
$$\operatorname{vec}(\boldsymbol{\beta}) = \boldsymbol{J}\boldsymbol{\gamma} + \boldsymbol{r}, \tag{4.2}$$
where $\boldsymbol{\beta}$ is defined in (3.2) and
  - $\boldsymbol{J}$ is a *known* $K(Kp+1) \times M$-matrix of rank M,
  - $\boldsymbol{r}$ a $K(Kp+1)$-vector of *known* parameters,
  - $\boldsymbol{\gamma}$ a M-vector of *unknown* parameters.
- If we write the linear parameter constraint $\boldsymbol{\omega} = \mathbf{0}$ in the form (4.2), then $\boldsymbol{\gamma}$ contains all non-zero elements of $\boldsymbol{\beta}$, $\boldsymbol{r} = \mathbf{0}$ and $\boldsymbol{J}$ contains only zeros and ones; see also Example 4.12 below.

Example 4.12:

Consider the 2-dimensional VAR(1) model $\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t$ with

$$\boldsymbol{\beta} = \begin{pmatrix} \phi_{0,1} & \phi_{0,2} \\ \phi_{1,11} & \phi_{1,21} \\ \phi_{1,12} & \phi_{1,22} \end{pmatrix}, \quad \text{where } \boldsymbol{\phi}_0 = \begin{pmatrix} \phi_{0,1} \\ \phi_{0,2} \end{pmatrix},\ \boldsymbol{\phi}_1 = [\phi_{1,ij}]_{i,j=1,2}.$$

Example 4.12: Continued.

Suppose the true model is

$$\boldsymbol{z}_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.6 & 0 \\ 0.2 & 0.8 \end{pmatrix} \boldsymbol{z}_{t-1} + \boldsymbol{a}_t.$$

Then, $\boldsymbol{r} = \boldsymbol{0}$ and

$$\operatorname{vec}(\beta) = \begin{pmatrix} \phi_{0,1} \\ \phi_{1,11} \\ \phi_{1,12} \\ \phi_{0,2} \\ \phi_{1,21} \\ \phi_{1,22} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1.0 \\ 0.6 \\ 0.2 \\ 0.8 \end{pmatrix} = \boldsymbol{J}\boldsymbol{\gamma}.$$

- Under the linear constraints (4.2), equation (3.3) becomes
$$\operatorname{vec}(\boldsymbol{Z}) = (\boldsymbol{I}_K \otimes \boldsymbol{X})(\boldsymbol{J}\gamma + \boldsymbol{r}) + \operatorname{vec}(\boldsymbol{A}).$$
- The GLS estimate $\widehat{\gamma}$ of the only unknown parameters γ is obtained by minimzing
$$S(\gamma) = \operatorname{vec}(\boldsymbol{A})'(\boldsymbol{\Sigma_a} \otimes \boldsymbol{I}_{T-p})^{-1} \operatorname{vec}(\boldsymbol{A}).$$

Theorem 4.13:

Under (4.2) and the conditions of Theorem 3.3, $\widehat{\gamma}$ is a consistent estimator of γ and asymptotically normal with

$$\sqrt{T}(\widehat{\gamma} - \gamma) \xrightarrow{d} N\left(\boldsymbol{0}, [\boldsymbol{J}'(\boldsymbol{\Sigma_a}^{-1} \otimes \boldsymbol{G})\boldsymbol{J}]^{-1}\right), \qquad \text{as } T \to \infty.$$

Corollary 4.14: Lütkepohl (2005), Proposition 5.3.

Under the conditions of Theorem 4.13, the estimator $\mathrm{vec}(\widetilde{\beta}) = \boldsymbol{J}\widehat{\gamma} + \boldsymbol{r}$ satisfies

$$\sqrt{T}(\mathrm{vec}(\widetilde{\beta}) - \mathrm{vec}(\beta)) \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{J}[\boldsymbol{J}'(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{G})\boldsymbol{J}]^{-1}\boldsymbol{J}'\right), \qquad \text{as } T \to \infty.$$

- The asymptotic variance of the unrestrained estimator of Theorem 3.3 is $\boldsymbol{\Sigma}_{\boldsymbol{a}} \otimes \boldsymbol{G}^{-1}$.
- One can show (see Lütkepohl, 2005, p. 199) that

$$\boldsymbol{\Sigma}_{\boldsymbol{a}} \otimes \boldsymbol{G}^{-1} - \boldsymbol{J}[\boldsymbol{J}'(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \otimes \boldsymbol{G})\boldsymbol{J}]^{-1}\boldsymbol{J}' \geq 0.$$

- This shows that imposing restrictions increases asymptotic efficiency.

Outline

Forecasting

- Prediction is one of the main objectives of time series analysis.
- In this section, we are concerned with
  1. point forecasts,
  2. forecast intervals.
- Forecast intervals are important to quantify the uncertainty in the point forecasts.
  - Just ask any economist why his predictions are so wrong...

1. Point Forecasts

Point Forecasts

- Suppose $\boldsymbol{z}_t$ is a K-dimensional VAR(p) process
- For now, we assume—unrealistically—that the model parameters are known.
- We want to predict $\boldsymbol{z}_{T+h}$ based on information at T
  - T is the **forecast origin** and h the **forecast horizon**
- The information Ω_T at T is typcially given by $\boldsymbol{z}_T, \boldsymbol{z}_{T-1}, \ldots$.
- A prediction $\widehat{\boldsymbol{z}}_{T,T+h}$ is called an ***h*-step ahead forecast at time T**.
- The **forecast error** is $\boldsymbol{z}_{T+h} - \widehat{\boldsymbol{z}}_{T,T+h}$.

Point Forecasts
Minimum MSE Predictor

- Now, what is a good forecast? One with a 'small' forecast error!
- What is 'small'? Small mean squared error (MSE):

$$\mathrm{MSE}(\widehat{\boldsymbol{z}}_{T,T+h}) = \mathrm{E}\left[(\boldsymbol{z}_{T+h} - \widehat{\boldsymbol{z}}_{T,T+h})(\boldsymbol{z}_{T+h} - \widehat{\boldsymbol{z}}_{T,T+h})'\right].$$

- The **minimum MSE predictor** for forecast horizon h and forecast origin T is

$$\boxed{\boldsymbol{z}_T(h) := \mathrm{E}[\boldsymbol{z}_{T+h} \mid \Omega_T] = \mathrm{E}[\boldsymbol{z}_{T+h} \mid \boldsymbol{z}_T, \boldsymbol{z}_{T-1}, \ldots].}$$

- This predictor minimizes the MSE of each component of $\boldsymbol{z}_t$.
- In other words, for any other predictor $\widehat{\boldsymbol{z}}_{T,T+h}$ based on Ω_T

$$\mathrm{MSE}(\widehat{\boldsymbol{z}}_{T,T+h}) \geq \mathrm{MSE}(\boldsymbol{z}_T(h)), \tag{5.1}$$

where $\geq$ means that the difference of the left-hand and right-hand matrix is p.s.d.

- Why does (5.1) hold?
- Rewrite

$$
\begin{aligned}
\mathrm{MSE}(\widehat{\boldsymbol{z}}_{T,T+h}) &= \mathrm{E}[\{\boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h) + \boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}\times \\
&\qquad\qquad \{\boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h) + \boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}'] \\
&= \mathrm{E}[\{\boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h)\}\{\boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h)\}'] \\
&\qquad\qquad + \mathrm{E}[\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}'] \\
&= \mathrm{MSE}[\boldsymbol{z}_T(h)] \\
&\qquad\qquad + \mathrm{E}[\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}'],
\end{aligned}
$$

where for the second equality we have used the property

$$
\begin{aligned}
&\mathrm{E}[\{\boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h)\}\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}'] \\
&\quad = \mathrm{E}\left(\mathrm{E}[\{\boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h)\}\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}' \mid \Omega_T]\right) \\
&\quad = \mathrm{E}\left(\mathrm{E}[\{\boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h)\} \mid \Omega_T]\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}'\right) = \mathbf{0}. \qquad (5.2)
\end{aligned}
$$

UNIVERSITÄT DUISBURG ESSEN

- Since $\mathrm{E}[\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_{T,T+h}\}']$ is a p.s.d. matrix, we conclude that
$$\mathrm{MSE}(\widehat{\boldsymbol{z}}_{T,T+h}) \geq \mathrm{MSE}[\boldsymbol{z}_T(h)],$$
i.e., (5.1) holds.

- Two implications to remember are:

  1. $\boldsymbol{z}_T(h)$ is an unbiased predictor, i.e., (using the law of iterated expectations)
  $$\mathrm{E}[\boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h)] = \mathrm{E}(\mathrm{E}[\boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h) \mid \Omega_T]) = \mathbf{0}.$$

  2. the optimal forecast error $\boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h)$ is uncorrelated with the predictors $\boldsymbol{z}_T, \boldsymbol{z}_{T-1}, \ldots$—otherwise we could improve the forecast!

Point Forecasts

Calculating the MSE Predictor

- How do we calculate $\boldsymbol{z}_T(h)$ for VAR(p) processes?
- Applying the conditional expectation operator, $\mathrm{E}[\cdot \mid \Omega_T]$, to the VAR model equation, we get from $\mathrm{E}[\boldsymbol{a}_{T+h} \mid \Omega_T] = \boldsymbol{0}$ (by independence of $\boldsymbol{a}_{T+h}$ from $\boldsymbol{z}_T, \boldsymbol{z}_{T-1}, \ldots$) that

$$\boldsymbol{z}_T(h) = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_T(h-1) + \ldots + \boldsymbol{\phi}_p \boldsymbol{z}_T(h-p), \tag{5.3}$$

  where $\boldsymbol{z}_T(j) = \boldsymbol{z}_{T+j}$ for $j \leq 0$.

- The optimal h-step ahead forecast can be computed recursively from (5.3) using

$$\begin{aligned}
\boldsymbol{z}_T(1) &= \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_T + \ldots + \boldsymbol{\phi}_p \boldsymbol{z}_{T-p+1}, \\
\boldsymbol{z}_T(2) &= \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_T(1) + \boldsymbol{\phi}_2 \boldsymbol{z}_T + \ldots + \boldsymbol{\phi}_p \boldsymbol{z}_{T-p+2}, \\
&\vdots
\end{aligned}$$

Example 5.1:

Applying the above recursion to a VAR(1), we obtain

$$\boldsymbol{z}_T(h) = (\boldsymbol{I}_K + \phi_1 + \ldots + \phi_1^{h-1})\phi_0 + \phi_1^h \boldsymbol{z}_T.$$

- Using $\phi_0 = (\boldsymbol{I}_K - \sum_{i=1}^{p} \phi_i)\boldsymbol{\mu}$, we rewrite (5.3) as

$$\boldsymbol{z}_T(h) - \boldsymbol{\mu} = \sum_{i=1}^{p} \phi_i \left[\boldsymbol{z}_T(h-i) - \boldsymbol{\mu}\right].$$

- This implies

$$\boldsymbol{Z}_T(h) = \boldsymbol{\Phi}_1 \boldsymbol{Z}_T(h-1), \qquad h > 1, \tag{5.4}$$

for $\boldsymbol{\Phi}_1$ as in (2.6) and

$$\boldsymbol{Z}_T(h) = \left(\left[\boldsymbol{z}_T(h) - \boldsymbol{\mu}\right]', \ldots, \left[\boldsymbol{z}_T(h-p+1) - \boldsymbol{\mu}\right]'\right)'.$$

- Repeated iterations of (5.4) gives

$$\boldsymbol{Z}_T(h) = \boldsymbol{\Phi}_1^{h-1}\boldsymbol{Z}_T(1), \qquad h > 1. \tag{5.5}$$

- For a stationary VAR(p), all eigenvalues of $\boldsymbol{\Phi}_1$ are less than 1 in absolute value.
- Therefore, $\boldsymbol{\Phi}_1^h \longrightarrow \mathbf{0}$, as $h \to \infty$.
- Consequently, from (5.5)

$$\boldsymbol{z}_T(h) - \boldsymbol{\mu} \underset{(h\to\infty)}{\longrightarrow} \mathbf{0}.$$

- As the forecast horizon increases, the optimal forecast converges to the mean—the optimal forecast of a white noise process.
- In other words, the past of the process contains no information on the development in the distant future!

2. Forecast Intervals

Forecast Intervals

- To set up forecast intervals, we need to study optimal forecast errors.
- To study the h-step ahead forecast error $\boldsymbol{e}_T(h) = \boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h)$, the causal representation in (2.8) is helpful:

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\theta}_i \boldsymbol{a}_{t-i}, \quad \text{where } \boldsymbol{\mu} = [\phi(1)]^{-1}\phi_0. \tag{5.6}$$

- Using this representation, it can be shown that

$$\boldsymbol{e}_T(h) = \sum_{i=0}^{h-1} \boldsymbol{\theta}_i \boldsymbol{a}_{T+h-i}, \tag{5.7}$$

where

$$\text{cov}(\boldsymbol{e}_T(h)) = \sum_{i=0}^{h-1} \boldsymbol{\theta}_i \boldsymbol{\Sigma}_{\boldsymbol{a}} \boldsymbol{\theta}_i'. \tag{5.8}$$

Forecast Intervals

- To simplify things, we assume that $\boldsymbol{a}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}})$.
- Then, by well-known properties of the normal distribution,

$$\boldsymbol{e}_T(h) = \sum_{i=0}^{h-1} \boldsymbol{\theta}_i \boldsymbol{a}_{T+h-i} \sim N(\mathbf{0}, \operatorname{cov}(\boldsymbol{e}_T(h))). \tag{5.9}$$

- Thus, for the forecast errors of the individual components

$$\frac{z_{T+h}^{(i)} - z_T^{(i)}(h)}{\sigma_{ii}(h)} \sim N(0,1), \tag{5.10}$$

  where $z_{T+h}^{(i)}$ ($z_T^{(i)}(h)$) is the i-th component of $\boldsymbol{z}_{T+h}$ ($\boldsymbol{z}_T(h)$), and $\sigma_{ii}(h)$ is the square root of the i-th diagonal element of $\operatorname{cov}(\boldsymbol{e}_T(h))$.

- Hence, a $(1-\alpha)$-**forecast interval** for the i-th component of $\boldsymbol{z}_{T+h}$ is

$$\boxed{z_T^{(i)}(h) \pm \Phi^{-1}(1-\alpha/2)\sigma_{ii}(h).} \tag{5.11}$$

- The result in (5.9) can also be used to construct confidence regions for the whole vector $\boldsymbol{z}_{T+h}$.
- A $(1-\alpha)$-**confidence region** for $\boldsymbol{z}_{T+h}$ is the ellipsoid determined by

$$\boxed{\left\{\boldsymbol{z} \in \mathbb{R}^K : \ (\boldsymbol{z}_T(h) - \boldsymbol{z})' \operatorname{cov}(\boldsymbol{e}_T(h))^{-1} (\boldsymbol{z}_T(h) - \boldsymbol{z}) \leq \chi^2_{K,1-\alpha}\right\},}$$

  where $\chi^2_{K,1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2_K-distribution.
- If forecast regions of this type are computed repeatedly from a large number of realizations of the considered process, then about $(1-\alpha)100\%$ of the confidence regions will contain the actual value of $\boldsymbol{z}_{T+h}$.

- When parameters are estimated, the optimal forecast in (5.3) is now determined from

$$\widehat{\boldsymbol{z}}_T(h) = \widehat{\phi}_0 + \sum_{i=1}^{p} \widehat{\phi}_i \widehat{\boldsymbol{z}}_T(h-i), \qquad \text{where} \quad \widehat{\boldsymbol{z}}_T(j) = \boldsymbol{z}_{T+j} \text{ for } j \leq 0.$$

- Thus, the point forecast using estimated parameters remains the same.
- However, the associated forecast error is

$$\begin{aligned} \widehat{\boldsymbol{e}}_T(h) &= \boldsymbol{z}_{T+h} - \widehat{\boldsymbol{z}}_T(h) = \boldsymbol{z}_{T+h} - \boldsymbol{z}_T(h) + \boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_T(h) \\ &= \boldsymbol{e}_T(h) + [\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_T(h)]. \end{aligned}$$

- Note that $\boldsymbol{e}_T(h)$ and $[\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_T(h)]$ are uncorrelated, due to (5.2).

Forecast Intervals

The Case of Estimated Parameters

- Hence, we have

$$\text{cov}(\widehat{\boldsymbol{e}}_T(h)) = \text{cov}(\boldsymbol{e}_T(h)) + E\left[\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_T(h)\}\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_T(h)\}'\right]. \quad (5.12)$$

- An estimate of $\text{cov}(\boldsymbol{e}_T(h))$ can be obtained from (5.8).
- It remains to derive an estimate of $E\left[\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_T(h)\}\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_T(h)\}'\right]\ldots$

- To that end, assume that the parameter estimates satisfy

$$\sqrt{T_p}[\text{vec}(\widehat{\beta}) - \text{vec}(\beta)] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_\beta), \qquad \text{as } T \to \infty,$$

where $T_p = T - p$.

- (See Chapter 3, for examples of such estimators.)
- Since $\boldsymbol{z}_T(h)$ is a differentiable function of vec(β), the Delta method implies

$$\sqrt{T_p}[\widehat{\boldsymbol{z}}_T(h) - \boldsymbol{z}_T(h) \mid \Omega_T] \xrightarrow{d} N\left(\mathbf{0},\ \boldsymbol{\Omega}_h = \frac{\partial \boldsymbol{z}_T(h)}{\partial\, \text{vec}(\beta)'} \boldsymbol{\Sigma}_\beta \frac{\partial \boldsymbol{z}_T(h)}{\partial\, \text{vec}(\beta)}\right).$$

- This suggests we can approximate $E\left[\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_T(h)\}\{\boldsymbol{z}_T(h) - \widehat{\boldsymbol{z}}_T(h)\}'\right]$ in (5.12) by (an estimate of) $\boldsymbol{\Omega}_h / T_p$.

Forecast Intervals
The Case of Estimated Parameters

- The general formula for $\boldsymbol{\Omega}_h$ is complicated; see Tsay (2014, p. 87).
- For $h = 1$, it is simply
$$\boldsymbol{\Omega}_1 = (Kp + 1)\boldsymbol{\Sigma}_{\boldsymbol{a}},$$
such that
$$\operatorname{cov}(\widehat{\boldsymbol{e}}_T(1)) = \boldsymbol{\Sigma}_{\boldsymbol{a}} + \frac{\boldsymbol{\Omega}_1}{T_p} = \frac{T_p + Kp + 1}{T_p}\boldsymbol{\Sigma}_{\boldsymbol{a}}.$$
- Since $Kp + 1$ is the number of parameters in the model equation for z_{it}, this can be interpreted as follows:
  - Each parameter used increases the MSE of one-step ahead forecasts by a factor of $1/T_p$.
- This provides support for removing insignificant parameters, i.e., using parsimonious models.

Forecast Intervals
The Case of Estimated Parameters

- Denote by $\widehat{\sigma}_{ii}(h)$ the square root of the i-th diagonal element of $\text{cov}(\widehat{\boldsymbol{e}}_T(h))$.
- Assume again that $\boldsymbol{a}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}})$.
- Then, a $(1-\alpha)$-forecast interval for the i-th component of $\boldsymbol{z}_{T+h}$, that also factors in the parameter estimation uncertainty, is given by

$$\boxed{\widehat{\boldsymbol{z}}_T^{(i)}(h) \pm \Phi^{-1}(1-\alpha/2)\widehat{\sigma}_{ii}(h).}$$

- Just like $\sigma_{ii}(h)$ in (5.11), $\widehat{\sigma}_{ii}(h)$ has to be estimated for feasible forecast interval construction.

Example 5.2:

Reconsider Example 3.5. We produce h-step ahead forecasts ($h = 1, \dots, 4$) of GDP growth at forecast origin 2011Q2. We also provide estimates of the

- standard deviations $\sigma_{ii}(h)$ (from (5.10)) and
- root mean square errors (RMSEs)—including parameter estimation uncertainty—of the predictions, i.e., $\widehat{\sigma}_{ii}(h)$.

We make the following observations from the output:

1. The point forecasts move closer to the sample means as h increases.
2. The standard error and the RMSE increase with the forecast horizon.
3. The effect of using estimated parameters is large (small) when the forecast horizon is small (large).

```
VARpred(m1, 4)   # forecasts using VAR(2)
```

```
## orig  125
## Forecasts at origin:  125
##           uk      ca     us
## [1,] 0.3129 0.05166 0.1660
## [2,] 0.2647 0.31687 0.4889
## [3,] 0.3143 0.48231 0.5205
## [4,] 0.3839 0.53053 0.5998
## Standard Errors of predictions:
##         [,1]   [,2]   [,3]
## [1,] 0.5315 0.5400 0.5975
## [2,] 0.5804 0.7165 0.7077
## [3,] 0.6202 0.7672 0.7345
## [4,] 0.6484 0.7785 0.7442
## Root mean square errors of predictions:
##         [,1]   [,2]   [,3]
## [1,] 0.5461 0.5549 0.6140
## [2,] 0.6001 0.7799 0.7499
## [3,] 0.6365 0.7879 0.7456
## [4,] 0.6601 0.7832 0.7484
```

```
colMeans(z)      # compute sample means
```

```
##     uk     ca     us
## 0.5223 0.6154 0.6474
```

Outline

UNIVERSITÄT DUISBURG ESSEN

Structural Analysis

- VAR models are used to model relationships among a set of variables.
- In the following, we discuss three closely related ways to interpret the relationships implied by VAR models:
  1. Granger causality and instantaneous causality,
  2. impulse response functions (IRFs),
  3. forecast error variance decomposition.

1. Granger Causality and Instantaneous Causality

Granger Causality and Instantaneous Causality

- Granger (1969) introduced a concept of causality that is easy to deal with for VAR models
- The idea is that cause cannot come after the effect.
- Thus, if a variable $\boldsymbol{x}$ affects a variable $\boldsymbol{y}$, $\boldsymbol{x}$ should help in improving predictions of $\boldsymbol{y}$.
- We formalize this idea next...

- Let Ω_t contain past $\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \ldots$
- Let $\boldsymbol{y}_t(h \mid \Omega_t)$ be the optimal (minimum MSE) h-step predictor of $\boldsymbol{y}_{t+h}$ based on information in Ω_t.
- Let $\text{cov}(\boldsymbol{e}_t(h) \mid \Omega_t)$ denote the corresponding forecast MSE

Definition 6.1: Granger Causality.

The process $\boldsymbol{x}_t$ **Granger-causes** (or briefly: **causes**) $\boldsymbol{y}_t$ if for some t and h,

$$\text{cov}(\boldsymbol{e}_t(h) \mid \Omega_t) \neq \text{cov}(\boldsymbol{e}_t(h) \mid \Omega_t \setminus \{\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \ldots\}).$$

- In other words, if $\boldsymbol{y}_t$ can be predicted more efficiently using additionally $\boldsymbol{x}_t$, then $\boldsymbol{x}_t$ is **Granger causal** for $\boldsymbol{y}_t$.

Definition 6.2: Instantaneous Causality.

There is **instantaneous causality** between $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$, if

$$\operatorname{cov}(\boldsymbol{e}_t(1) \mid \Omega_t \cup \{\boldsymbol{x}_{t+1}\}) \neq \operatorname{cov}(\boldsymbol{e}_t(1) \mid \Omega_t).$$

- In other words, in period t, adding $\boldsymbol{x}_{t+1}$ to the information set helps to improve the forecast of $\boldsymbol{y}_{t+1}$.

Remark 6.3:

The concept of causality is symmetric, i.e., instantaneous causality between $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ is the same as instantaneous causality between $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$; see Lütkepohl (2005, Proposition 2.3).

Granger Causality and Instantaneous Causality

- Granger causality can easily be identified for a VAR(p) model:

$$\boldsymbol{z}_t = \begin{pmatrix} \boldsymbol{x}_t \\ \boldsymbol{y}_t \end{pmatrix} = \begin{pmatrix} \phi_{0,1} \\ \phi_{0,2} \end{pmatrix} + \begin{pmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_{t-1} \\ \boldsymbol{y}_{t-1} \end{pmatrix} + \ldots + \begin{pmatrix} \phi_{p,11} & \phi_{p,12} \\ \phi_{p,21} & \phi_{p,22} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_{t-p} \\ \boldsymbol{y}_{t-p} \end{pmatrix} + \begin{pmatrix} \boldsymbol{a}_{1t} \\ \boldsymbol{a}_{2t} \end{pmatrix}. \quad (6.1)$$

Theorem 6.4: Characterization of Granger-Noncausality.

If $\boldsymbol{z}_t$ is a stationary VAR(p) as in (6.1), then

$$\boldsymbol{x}_t(h \mid \{\boldsymbol{z}_t, \boldsymbol{z}_{t-1}, \ldots\}) = \boldsymbol{x}_t(h \mid \{\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \ldots\}), \qquad h = 1, 2, \ldots$$
$$\iff \qquad \phi_{i,12} = 0 \qquad \text{for } i = 1, \ldots, p.$$

In other words, $\boldsymbol{y}_t$ does not Granger-cause $\boldsymbol{x}_t$ if and only if $\phi_{i,12} = 0$ for $i = 1, \ldots, p$.

Example 6.5:

Consider a system with variables

$$z_1 = \text{rate of change of investment},$$
$$z_2 = \text{income},$$
$$z_3 = \text{consumption}.$$

Suppose the model is

$$\begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \end{pmatrix} = \phi_0 + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} \begin{pmatrix} z_{1,t-1} \\ z_{2,t-1} \\ z_{3,t-1} \end{pmatrix} + \boldsymbol{a}_t.$$

Then,

- $\boldsymbol{y}_t = (z_{2,t}, z_{3,t})'$ (i.e., (consumption, income)) does not Granger-cause $\boldsymbol{x}_t = z_{1,t}$ (i.e., investment).
- Conversely however, investment Granger-causes (income, consumption).

Example 6.6: GDP Growth.

Reconsider the refined VAR(2) from Example 4.11:

$$\boldsymbol{z}_t = \begin{pmatrix} 0.16 \\ - \\ 0.28 \end{pmatrix} + \begin{pmatrix} 0.47 & 0.21 & - \\ 0.33 & 0.27 & 0.50 \\ 0.47 & 0.23 & 0.23 \end{pmatrix} \boldsymbol{z}_{t-1} + \begin{pmatrix} - & - & - \\ -0.20 & - & - \\ -0.30 & - & - \end{pmatrix} \boldsymbol{z}_{t-2} + \boldsymbol{a}_t,$$

It is tempting to conclude that GDP growth in the US, $\boldsymbol{y}_t = z_{3,t}$, does not Granger-cause growth in the UK, $\boldsymbol{x}_t = z_{1,t}$. However, Theorem 6.4 only considers the case of two subvectors $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ of $\boldsymbol{z}_t = (\boldsymbol{x}_t', \boldsymbol{y}_t')'$.

- To study instantaneous causality, we use the causal representation of the VAR(p) in (2.8).

- For the symmetric, p.d. matrix $\boldsymbol{\Sigma_a}$, the **Cholesky decomposition** of Lemma 1.7 states

$$\boldsymbol{\Sigma_a} = \boldsymbol{L}\boldsymbol{L}',$$

  where $\boldsymbol{L}$ is a lower triangular matrix with positive diagonal elements.

- Write the causal representation with $\boldsymbol{\theta}_0 = \boldsymbol{I}_K$ as

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\theta}_i \boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{a}_{t-i} = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Theta}_i \boldsymbol{\eta}_{t-i}, \tag{6.2}$$

  where $\boldsymbol{\Theta}_i = \boldsymbol{\theta}_i \boldsymbol{L}$ and the **orthogonal innovations** $\boldsymbol{\eta}_t = \boldsymbol{L}^{-1}\boldsymbol{a}_t$ are white noise with

$$\operatorname{cov}(\boldsymbol{\eta}_t) = \boldsymbol{L}^{-1}\boldsymbol{\Sigma_a}(\boldsymbol{L}^{-1})' = \boldsymbol{L}^{-1}(\boldsymbol{L}\boldsymbol{L}')(\boldsymbol{L}^{-1})' = \boldsymbol{I}_K.$$

- Using (6.2) it is not difficult to show:

Theorem 6.7: Characterization of Instantaneous Causality.

Let $\boldsymbol{z}_t$ be as in (6.1). Then, there is no instantaneous causality between $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ if and only if

$$\mathrm{E}[\boldsymbol{a}_{1t}\boldsymbol{a}_{2t}'] = 0.$$

- This theorem provides an easy-to-check condition for instantaneous causality.
- It shows that instantaneous causality merely describes a nonzero correlation between two sets of variables, but nothing about a cause and effect relation
  - The direction of instantaneous causality cannot be derived, but must come from (e.g.) economic theory.

Example 6.8:

Suppose the white noise covariance matrix in Example 6.5 is

$$\boldsymbol{\Sigma_a} = \begin{pmatrix} 2.25 & 0 & 0 \\ 0 & 1.0 & 0.5 \\ 0 & 0.5 & 0.74 \end{pmatrix}.$$

Then, there is no instantaneous causality between (income, consumption) and investment.

2. Impulse Response Functions

Impulse Response Functions

- In applied work, it is often valuable to know the *response* of one variable to an *impulse* in another variable.
- Here, we study this by tracing out the effect of an exogenous shock to one variable on all of the other variables.

Example 6.9:

In a system consisting of inflation rate and interest rate, the effect of an increase in inflation may be of interest. In the real world, such an increase may be induced exogenously (i.e., outside the simple two variable system) by an increase in oil prices due to the joint action of the OPEC in 1973 to raise prices.

Impulse Response Functions

- In impulse response analysis we can assume $\mathrm{E}[\boldsymbol{z}_t] = \mathbf{0}$, because the mean does not affect the response of $\boldsymbol{z}_t$ to any shock.
- Hence, the causal representation of $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{Kt})'$ becomes

$$\boldsymbol{z}_t = \sum_{i=0}^{\infty} \boldsymbol{\theta}_i \boldsymbol{a}_{t-i} = \boldsymbol{a}_t + \sum_{i=1}^{\infty} \boldsymbol{\theta}_i \boldsymbol{a}_{t-i} \tag{6.3}$$

- To study the effects of changes in z_{1t} on $\boldsymbol{z}_{t+j}$ for $j > 0$, while holding other quantities unchanged, we assume $\boldsymbol{z}_t = 0$ for $t < 0$, i.e., $\boldsymbol{a}_t = 0$ for $t < 0$ in (6.3).
- We now trace out what happens to the system in $t \geq 0$ when $\boldsymbol{a}_0 = (1, 0, \ldots, 0)'$, i.e., a unit shock in z_{1t} occurs, if no further shocks occur, i.e., $\boldsymbol{a}_t = 0$ for $t > 0$.

Impulse Response Functions

- Using (6.3) we obtain

$$\boldsymbol{z}_0 = \boldsymbol{a}_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{z}_1 = \boldsymbol{\theta}_1 \boldsymbol{a}_0 = \begin{pmatrix} \theta_{1,11} \\ \theta_{1,21} \\ \vdots \\ \theta_{1,K1} \end{pmatrix}, \quad \boldsymbol{z}_2 = \boldsymbol{\theta}_2 \boldsymbol{a}_0 = \begin{pmatrix} \theta_{2,11} \\ \theta_{2,21} \\ \vdots \\ \theta_{2,K1} \end{pmatrix}, \ldots$$

- In short, $\boldsymbol{z}_t = \boldsymbol{\theta}_{t,\cdot 1}$, where $\boldsymbol{\theta}_{t,\cdot 1}$ denotes the first column of $\boldsymbol{\theta}_t$.
- Similarly, if the effect of a unit shock in z_{it} on $\boldsymbol{z}_{t+j}$ is traced out, we get

$$\boxed{\boldsymbol{z}_t = \boldsymbol{\theta}_{t,\cdot i}, \qquad t \geq 0.}$$

- Therefore, the coefficient matrices $\boldsymbol{\theta}_i$ contain the coefficients of the *impulse responses*.

Impulse Response Functions

- Sometimes interest centers on the accumulated effect over several periods.
- The kth column of

$$\underline{\boldsymbol{\theta}_n} = \sum_{i=0}^{n} \boldsymbol{\theta}_i$$

contains the **accumulated responses** over n periods to a unit shock in the kth variable.

- The infinite sum

$$\underline{\boldsymbol{\theta}_\infty} = \sum_{i=0}^{\infty} \boldsymbol{\theta}_i$$

is called the matrix of **long-run effects**.

Remark 6.10:

One can show that a shock to variable k has no effect (i.e., zero impulse responses) on the other variables, if variable k does not Granger-cause the remaining variables.

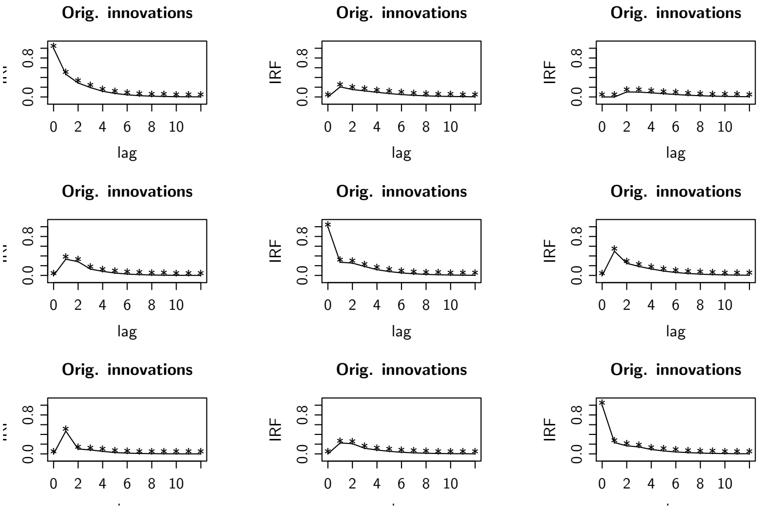
Example 6.11: GDP Growth.

We calculate the impulse response functions for the refined VAR(2) from Example 4.11. We make the following observations:

1. The impulse responses decay to 0 quickly.
2. The upper-right plot shows that there is a delayed effect on the UK GDP growth if one changes the US growth rate by 1. This is because the US rate at t affects the Canadian rate at $t+1$, which affects the UK rate at $t+2$.
3. The accumulated responses converge quickly to the long-run effect.

```
Phi = m.ref$Phi                # m.ref is the simplified VAR(2) model
Sig = m.ref$Sigma
VARirf(Phi, Sig, orth=FALSE)  # IRFs for the original innovations
```

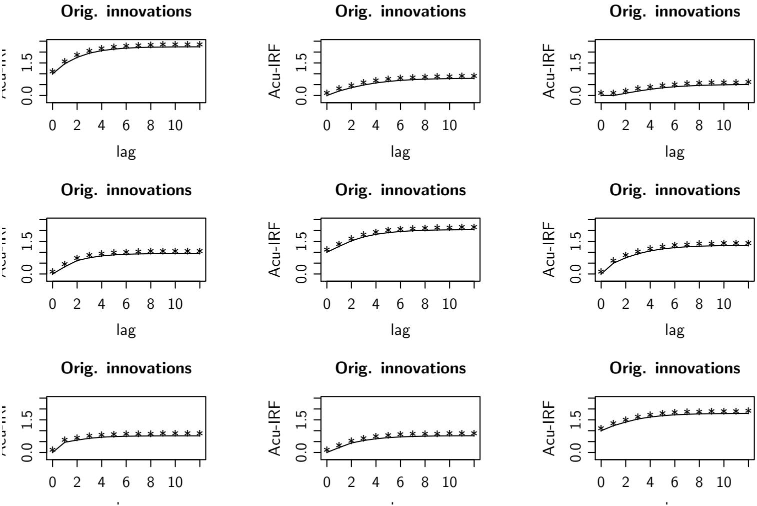
Impulse Response Functions II
GDP Growth in R




```
## Press return to continue
```

Impulse Response Functions III

GDP Growth in R



- Previously, we have studied the effects of a unit shock in $\boldsymbol{a}_t = (a_{1t}, a_{2t}, \ldots, a_{Kt})' = (1, 0, \ldots, 0)'$ on $\boldsymbol{z}_{t+j}$.
- However, the components a_{kt} $(k = 1, \ldots, K)$ consist of all influences not included in the $\boldsymbol{z}$ variables.
- Thus, in addition to forces that affect all variables, there may be forces that affect only variable 1.
  - If a_{1t} does not contain forces that also affect the other variables (i.e., if a_{1t} is uncorrelated with a_{kt}, $k = 2, \ldots, K$), the $\boldsymbol{\theta}_i$ can be interpreted—just as desired—as dynamic responses to a unit shock only in the first variable.
  - If this is not the case, a shock to a_{1t} also affects some other a_{kt} $(k = 2, \ldots, K)$, so that setting $a_{kt} = 0$ provides a misleading picture of the actual dynamics and, hence, the responses cannot be interpreted as above.

Example 6.12:

The white noise covariance matrix in Example 6.8 is

$$\boldsymbol{\Sigma_a} = \begin{pmatrix} 2.25 & 0 & 0 \\ 0 & 1.0 & 0.5 \\ 0 & 0.5 & 0.74 \end{pmatrix}.$$

- Obviously, there is strong positive correlation between a_{2t} and a_{3t}, the residuals of the income and consumption equations, respectively.
- Consequently, at time t a shock in income may be accompanied by a shock in consumption.
- Therefore, forcing the consumption innovations to zero when the effect of an income shock is traced, may in fact obscure the actual relation between the variables.

- To overcome the above difficulty, we **orthogonalize** the $\boldsymbol{a}_t$.
- One way of doing so is via the Cholesky decomposition, i.e., by considering

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Theta}_i \boldsymbol{\eta}_{t-i} \tag{6.4}$$

  from (6.2), where the components of $\boldsymbol{\eta}_t$ are uncorrelated with unit variance.
- Thus, we can interpret $\{\Theta_{l,ij}\}_{l=0,1,2,\dots}$ of the matrix sequence

$$\boldsymbol{\Theta}_l = \frac{\partial \boldsymbol{z}_{t+l}}{\partial \boldsymbol{\eta}_t'}$$

  as properly orthogonalized **impulse response functions**.
- Specifically, $\Theta_{l,ij}$ denotes the impact of a shock of 'one standard deviation' of $\eta_{j,0}$ on $z_{i,l}$.

- In $\boldsymbol{z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Theta}_i \boldsymbol{\eta}_{t-i}$, we have

$$\boldsymbol{a}_t = \boldsymbol{L}\boldsymbol{\eta}_t = \begin{pmatrix} L_{11} & 0 & \cdots & 0 \\ L_{21} & L_{22} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ L_{K1} & \cdots & \cdots & L_{KK} \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \vdots \\ \eta_{Kt} \end{pmatrix}$$

- That is, by orthogonalizing we have implicitly imposed the assumption that—contemporaneously—z_{1t} is only affected by its own shock η_{1t}, whereas z_{2t} is affected by its own shock η_{2t} and η_{1t}; z_{Kt} is affected by all η_{jt}.
- The **results of the impulse response analysis may be affected** by which variable is called z_{1t} etc.

- Unfortunately, economic theory rarely provides much guidance regarding the above identifying assumption.
- One would have to answer the question ‘Which variable does not react instantaneously to the others?’
- It is then good practice to try a few orderings and to hope not much changes.

Impulse Response Functions I

GDP Growth in R

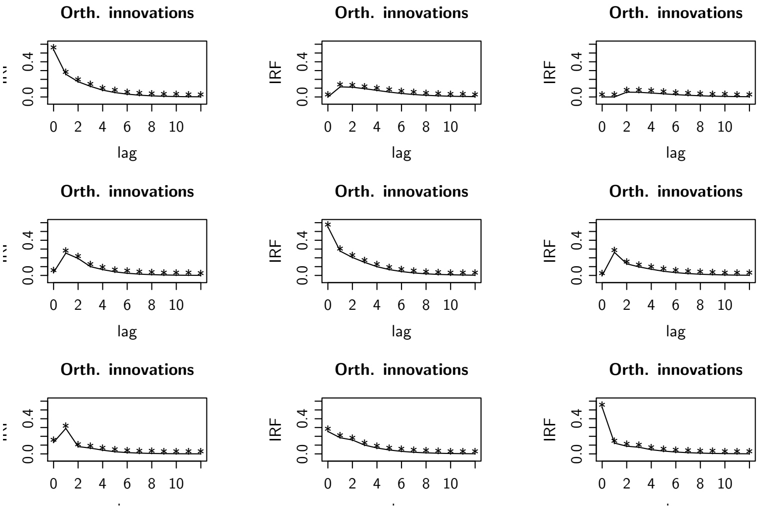
Example 6.13: GDP Growth.

We calculate the orthogonalized impulse response functions for the refined VAR(2) from Example 4.11. The results are quite similar to those in Example 6.11, because the correlations between the components of $\boldsymbol{a}_t$ are not strong.

```
VARirf(Phi, Sig)       # IRFs for orthogonal innovations
```

Impulse Response Functions II

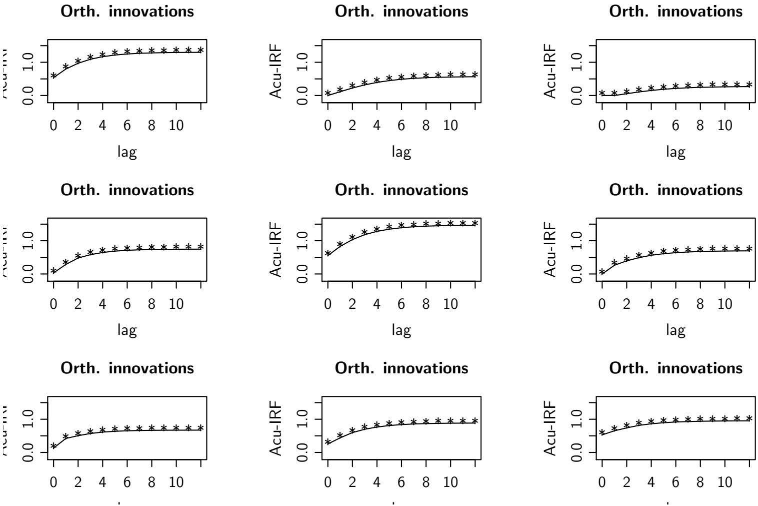
GDP Growth in R




```
## Press return to continue
```

Impulse Response Functions III

GDP Growth in R



3. Forecast Error Variance Decomposition

- From (6.4) we see that—similarly as in (5.7)—the forecast error is $\boldsymbol{e}_T(h) = \sum_{i=0}^{h-1} \boldsymbol{\Theta}_i \boldsymbol{\eta}_{T+h-i}$, where

$$\operatorname{cov}(\boldsymbol{e}_T(h)) = \sum_{l=0}^{h-1} \boldsymbol{\Theta}_l \boldsymbol{\Theta}_l'. \tag{6.5}$$

- From (6.5), the variance of the ith component, $e_T^{(i)}(h)$, of the forecast error is

$$\operatorname{Var}(e_T^{(i)}(h)) = \sum_{l=0}^{h-1} \sum_{j=1}^{K} \Theta_{l,ij}^2 = \sum_{j=1}^{K} \sum_{l=0}^{h-1} \Theta_{l,ij}^2 \equiv \sum_{j=1}^{K} w_{ij}(h), \tag{6.6}$$

  where

$$w_{ij}(h) = \sum_{l=0}^{h-1} \Theta_{l,ij}^2.$$

- $w_{ij}(h)$ can be interpreted as the contribution of the kth shock η_{jt} to the variance of the h-step ahead forecast error of z_{it}.

Forecast Error Variance Decomposition

- Equation (6.6), $\boxed{\operatorname{Var}(e_T^{(i)}(h)) = \sum_{j=1}^{K} w_{ij}(h)}$, is known as the **forecast error decomposition**.
  - The forecast error variance is decomposed into components accounted for by innovations in the different variables of the system.
- In particular,
$$\frac{w_{ij}(h)}{\operatorname{Var}(e_T^{(i)}(h))}$$
is the percentage contribution of the shock η_{jt} to the h-step ahead forecast error variance of z_{it}.
- If $\frac{w_{ii}(h)}{\operatorname{Var}(e_T^{(i)}(h))} = 1$, only z_{it}'s own shock η_{it} explains its movements.
- Note that since we again work with the Cholesky ordering induced by $\boldsymbol{L}$, the same comments as for impulse responses apply.

Forecast Error Variance Decomposition I

GDP Growth in R

Example 6.14: GDP Growth.

Consider again the refined VAR(2) from Example 4.11. We calculate the forecast error variance decompositions for the one-step to four-step ahead forecasts at the forecast origin 2011Q2.

```
Theta = NULL
FEVdec(Phi, Theta, Sig, lag=4)
```

```
## Order of the ARMA mdoel:
## [1] 2 0
## Standard deviation of forecast error:
##        [,1]   [,2]   [,3]   [,4]   [,5]
## [1,] 0.5386 0.6083 0.6444 0.6645 0.6746
## [2,] 0.5550 0.7198 0.7839 0.8100 0.8218
## [3,] 0.6022 0.7041 0.7317 0.7453 0.7510
## Forecast-Error-Variance Decomposition
## Forecast horizon:  1
##          [,1]   [,2]   [,3]
## [1,] 1.000000 0.0000 0.0000
## [2,] 0.003641 0.9964 0.0000
## [3,] 0.047328 0.1801 0.7726
```

Forecast Error Variance Decomposition II

GDP Growth in R

```
## Forecast horizon:  2
##         [,1]    [,2]   [,3]
## [1,] 0.9645 0.03548 0.0000
## [2,] 0.1267 0.74004 0.1333
## [3,] 0.2044 0.19992 0.5956
## Forecast horizon:  3
##         [,1]    [,2]     [,3]
## [1,] 0.9317 0.06115 0.007115
## [2,] 0.1674 0.69182 0.140735
## [3,] 0.2022 0.23200 0.565786
## Forecast horizon:  4
##         [,1]    [,2]    [,3]
## [1,] 0.9095 0.07753 0.01298
## [2,] 0.1722 0.68152 0.14625
## [3,] 0.2028 0.24163 0.55561
## Forecast horizon:  5
##         [,1]    [,2]   [,3]
## [1,] 0.8956 0.08745 0.0170
## [2,] 0.1738 0.67671 0.1495
## [3,] 0.2028 0.24596 0.5512
```

Outline

Vector Error Correction Models

- So far, we have dealt with stationary models, i.e., models with time invariant first and second moments.
- Stationary models cannot capture some main features of economic time series
  - stochastic trends
  - deterministic trends
  - seasonalities
- Hence, we investigate nonstationary models in the following.

- The outline of this chapter is as follows:

  1. Univariate Integrated Processes
  2. Cointegration and Error Correction
  3. Inclusion of Deterministic Trends
  4. Forecasting Cointegrated VARs
  5. Causality Analysis
  6. Impulse Response Analysis

1. Univariate Integrated Processes

Univariate Integrated Processes

Random Walks

- Recall that a VAR(p) process $\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \ldots + \boldsymbol{\phi}_p \boldsymbol{z}_{t-p} + \boldsymbol{a}_t$ is stationary if the stability condition holds, i.e.,

$$|\boldsymbol{I}_K - \boldsymbol{\phi}_1 z - \ldots - \boldsymbol{\phi}_p z^p| \neq 0 \qquad \text{for } |z| \leq 1.$$

- For a univariate AR(1), $z_t = \phi_0 + \phi_1 z_{t-1} + a_t$, this is satisfied if

$$1 - \phi_1 z \neq 0 \quad \text{for } |z| \leq 1 \qquad \Longleftrightarrow \qquad |\phi_1| < 1.$$

- In the borderline case $\phi_1 = 1$, the resulting process

$$z_t = \phi_0 + \phi_1 z_{t-1} + a_t = \ldots = z_0 + t\phi_0 + \sum_{i=1}^{t} a_i, \qquad z_0 \in \mathbb{R},$$

is called a **random walk with drift** if $\phi_0 \neq 0$, or simply a **random walk** if $\phi_0 = 0$.

- A random walk with drift is not stationary, as

$$\mathrm{E}[z_t] = t\phi_0 \quad \text{and} \quad \mathrm{Var}(z_t) = t\sigma_a^2 (\underset{(t\to\infty)}{\longrightarrow} \infty).$$

- For any $h > 0$, the correlation

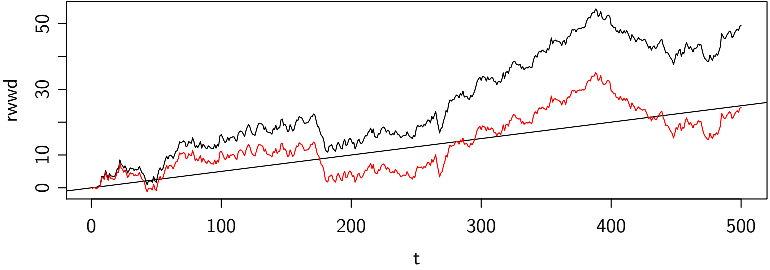
$$\begin{aligned} \operatorname{corr}(z_t, z_{t+h}) &= \frac{\mathrm{E}\left[\left(\sum_{i=1}^{t} a_i\right)\left(\sum_{i=1}^{t+h} a_i\right)\right]}{\sqrt{t\sigma_a^2 (t+h)\sigma_a^2}} \\ &= \frac{t}{\sqrt{t^2 + th}} \underset{(t\to\infty)}{\longrightarrow} 1. \end{aligned}$$

- This means that z_t and z_{t+h} are strongly correlated even if they are far apart in time.

Univariate Integrated Processes

Random Walk Trajectories in R

```
wn   <- rnorm(500, 0, 1)         # standard normal white noise
rw   <- cumsum(wn)               # random walk
phi0 <- 0.05                     # drift term
rwwd <- cumsum(wn + 0.05)        # random walk with drift
plot(rwwd, type="l", xlab="t",
      ylim=c(min(c(rw, rwwd)), max(c(rw,rwwd))))  # black: random walk with drift
abline(a=0, b=0.05)
lines(rw, col="red")             # red:   random walk
```



Univariate Integrated Processes

- More general AR(p) processes $z_t = \phi_0 + \phi_1 z_{t-1} + \ldots + \phi_p z_{t-p} + a_t$ can also display random walk-type behavior, if

$$1 - \phi_1 z - \ldots - \phi_p z^p = 0 \qquad \text{for } z = 1.$$

- Recall that by the **fundamental theorem of algebra**

$$1 - \phi_1 z - \ldots - \phi_p z^p = (1 - \lambda_1 z) \cdot \ldots \cdot (1 - \lambda_p z), \tag{7.1}$$

  where $\lambda_1, \ldots, \lambda_p$ are the reciprocals of the roots of the polynomial.

- If there is just one **unit root** (a root equal to 1) and all other roots are outside the complex unit circle, the AR(p) behaves similarly as a random walk
  - first differences are stationary,
  - linear trend in mean if $\phi_0 \neq 0$, and variances increase linearly in t,
  - correlations tend to 1.

Univariate Integrated Processes

- AR processes with $d \in \{0, 1, \ldots, p\}$ unit roots in (7.1) and all other roots outside the complex unit circle are called **integrated of order** d .
- An AR process that is integrated of order d can be made stationary by differencing d times, that is

$$\Delta^d z_t = (1 - L)^d z_t.$$

- In light of this, we generalize the above definition to more general processes:

Definition 7.1:

A multivariate process $\boldsymbol{z}_t$ is called **integrated of order** d ($I(d)$) if $\Delta^d \boldsymbol{z}_t$ is stationary and $\Delta^{d-1} \boldsymbol{z}_t$ is not. The order d is referred to as the **order of integration** or the **multiplicity of a unit root**.

2. Cointegration and Error Correction

Spurious Regression

- More often than not, we work with multiple nonstationary time series in economics.
- To illustrate take two variables z_{1t} and z_{2t}.
- We might ask whether there are any **equilibrium relationships** among the two, although they are individually nonstationary.
  - Such a relation might be useful for forecasting!

Spurious Regression

- To investigate the existence of equilibrium relationships, we can run a regression of z_{1t} on z_{2t}:

$$z_{1t} = \alpha + \beta z_{2t} + e_t, \qquad t = 1, \dots, T. \tag{7.2}$$

- Suppose there is no relationship between z_{1t} and z_{2t}. For simplicity, assume both are independent random walks, i.e.,

$$z_{1t} = z_{1,t-1} + a_{1t},$$
$$z_{2t} = z_{2,t-1} + a_{2t},$$

  where $(a_{1t}, a_{2t})' \overset{\text{i.i.d.}}{\sim} (\mathbf{0}, \boldsymbol{I}_2)$.

- We would expect that the OLS regression coefficient $\widehat{\beta}$ satisfies $\widehat{\beta} \xrightarrow{p} 0$.
- It turns out that this is not the case—we have a **spurious regression**.
- This was discovered in a simulation study by Granger and Newbold (1974).

Spurious Regression

- At an intuitive level, the problem with (7.2) is that

$$z_{1t} = \alpha + \beta z_{2t} + e_t \quad \stackrel{\alpha=\beta=0}{\Longrightarrow} \quad e_t = z_{1t},$$

that is, the error is $I(1)$—in violation of classical regression assumptions.

- More formally, Phillips (1986) showed the following:

Theorem 7.2:

The OLS estimators $(\widehat{\alpha}, \widehat{\beta})$ in (7.2) satisfy

$$\begin{pmatrix} T^{-1/2}\widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}, \qquad \text{as } T \to \infty,$$

where

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 1 & \int_0^1 W_2(r)\,\mathrm{d}r \\ \int_0^1 W_2(r)\,\mathrm{d}r & \int_0^1 W_2(r)^2\,\mathrm{d}r \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 W_1(r)\,\mathrm{d}r \\ \int_0^1 W_1(r)W_2(r)\,\mathrm{d}r \end{pmatrix}$$

and $W_1(r)$ and $W_2(r)$ are independent Brownian motions.

Spurious Regression

- Hence, $\widehat{\beta} \xrightarrow{d} g_2$ and not $\widehat{\beta} \xrightarrow{p} 0$, as it 'should'!
- Furthermore, $\widehat{\alpha}$ diverges, as it needs to be divided by $T^{1/2}$ to become a random variable with non-degenerate distribution.
- Above and beyond Theorem 7.2 one can show that the t-statistic for $H_0: \ \beta = 0$ diverges with rate $T^{1/2}$.
  - That is, although there is no relationship, a t-test will falsely indicate one.
- Therefore, such regressions are also appropriately called **nonsense regressions**.

Spurious Regression

- If we are sure that there is no relationship between the two variables, we can construct first differences Δz_{1t} and Δz_{2t} and regress these on each other.
- This is again a standard regression of stationary variables, and $\widehat{\beta} \xrightarrow{p} 0$ will result.
- However, it would be premature to recommend routine differencing of $I(1)$ variables—you would miss out on what comes next!

Example 7.3:

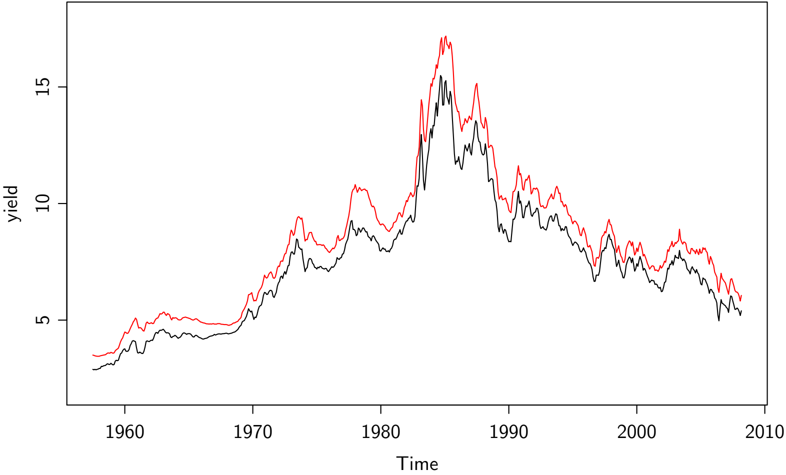
Consider the 609 monthly yields of Moody's seasoned corporate Aaa and Baa bonds from 1954M7 to 2005M2. The following figure shows the time plots of the bond yields.

```
da = read.table("Resources/Data/m-bnd.txt")
head(da)
```

```
##      V1 V2 V3   V4   V5
## 1 1954  7  1 2.89 3.50
## 2 1954  8  1 2.87 3.49
## 3 1954  9  1 2.89 3.47
## 4 1954 10  1 2.87 3.46
## 5 1954 11  1 2.89 3.45
## 6 1954 12  1 2.90 3.45
```

```
Aaa = ts(da[, 4], start = c(1957, 7), frequency = 12) # from 1954M7 to 2005M2
Baa = ts(da[, 5], start = c(1957, 7), frequency = 12) # from 1954M7 to 2005M2
plot(Aaa, ylab="yield", ylim=c(2, 18))
lines(Baa, col="red")
```

Cointegration II



- The previous plot suggests that two variables can be individually $I(1)$, yet there is some linear combination of the two that is $I(0)$.
- This phenomenon is called **cointegration**.

Example 7.4:

There is cointegration in the model

$$z_{1t} = \beta z_{2t} + e_{1t}, \tag{7.3}$$

$$z_{2t} = z_{2,t-1} + e_{2t}, \tag{7.4}$$

where $(e_{1t}, e_{2t})' \overset{\text{i.i.d.}}{\sim} (\mathbf{0}, \boldsymbol{\Sigma})$. Obviously, z_{2t} is $I(1)$ and hence also z_{1t}. Since the linear combination $z_{1t} - \beta z_{2t}$ is stationary, z_{1t} and z_{2t} are cointegrated.

Example 7.4: Continued.

Why is this interesting? Let $\Delta z_{1t} = z_{1t} - z_{1,t-1}$. Subtract $z_{1,t-1}$ from both sides of (7.3) and insert (7.4) to get

$$\begin{aligned}\Delta z_{1t} &= -1(z_{1,t-1} - \beta z_{2,t-1}) + \beta e_{2t} + e_{1t} \\ &=: -1(z_{1,t-1} - \beta z_{2,t-1}) + a_{1t}. \end{aligned} \tag{7.5}$$

This is an **error correction representation**—a statistical way of thinking about long-run equilibria. That is levels of variables contain useful information for forecasting the changes of variables.

Where might this be interesting? Suppose (somewhat unrealistically) that $z_{1,t}$ and $z_{2,t}$ denote the stock prices of Coca-Cola and Pepsi. Then, if Coca-Cola is cheap relative to Pepsi (i.e., if $z_{1,t-1} < \beta z_{2,t-1}$), we predict the change in Coca-Cola's share price $\Delta z_{1,t}$ to be positive. This strategy, called *pairs trading*, is useful indeed.

Definition 7.5:

A multivariate $I(d)$ process $\boldsymbol{z}_t$ is **cointegrated of order** (d, h), if there exists a linear combination $\beta' \boldsymbol{z}_t$ with $\beta \neq \mathbf{0}$ that is $I(d-h)$. Here, β is called a **cointegrating vector** and $\boldsymbol{z}_t$ a **cointegrated process**.

Remark 7.6:

a) In most real applications, only the case $(d, h) = (1, 1)$ is of interest. In this case, we merely say that $\boldsymbol{z}_t$ is cointegrated without specifying the order.

b) Cointegrating vectors are not unique. E.g., in the above definition, $c\beta'$ $(c \neq 0)$ is also a cointegrating vector. We shall impose some normalization subsequently.

Cointegration

- There is a plethora of economic theories that imply possible cointegrating relationships.
- Examples include:
  - Purchasing Power Parity (PPP): Apart from transportation costs, goods should be sold for the same effective price in two countries.
  - Fisher hypothesis: Real interest rate equals nominal interest rate minus expected inflation
  - Unbiased forward rate hypothesis.
  - Certain economic theories predict that consumption will be a constant proportion of income. Both are nonstationary, so cointegration implies stationarity of the log-difference.

Example 7.7: Term structure theory.

Long-and short term interest rates are (probably) nonstationary variables.

Yet, the theory of the term structure of interest rates asserts a relationship between interest rates of different maturities.

Denote by r_{Lt} and r_{St} long- and short-term interest rates at time t. Consider the following error-correction model.

$$\begin{aligned}\Delta r_{St} &= \alpha_S(r_{Lt-1} - \beta r_{St-1}) + \epsilon_{St} \\ \Delta r_{Lt} &= \alpha_L(r_{Lt-1} - \beta r_{St-1}) + \epsilon_{Lt}\end{aligned}$$

The cointegrating vector in this system then is $(1 \ -\beta)$.

We expect $\alpha_S > 0$ and $\alpha_L < 0$.

- In general, there can be more than one cointegrating vector.

- In our discussion of multiple cointegrating vectors, we focus on a cointegrated K-dimensional VAR(p) processes

$$\boldsymbol{z}_t = \phi_1 \boldsymbol{z}_{t-1} + \ldots + \phi_p \boldsymbol{z}_{t-p} + \boldsymbol{a}_t. \tag{7.6}$$

  that consists of variables z_{kt} $(k = 1, \ldots, K)$ that are $I(0)$ or $I(1)$ individually.

  - Representation (7.6) is mainly useful for forecasting; see Chapter 7.4
  - We will add a constant and a trend in (7.6) later on.

Question 7.8:

How many linearly independent cointegrating vectors are there in model (7.6)?

- Of course, any linear combination of two cointegrating vectors is a cointegrating vector again. Hence, we are only interested in linearly independent cointegrating vectors.

- The question is easy to answer for a stationary VAR(p), i.e., a VAR model with
$$|\boldsymbol{I}_K - \phi_1 z - \ldots - \phi_p z^p| \neq 0 \qquad \text{for } |z| \leq 1.$$
- For instance, the basis vectors $\boldsymbol{e}_i$ $(i = 1, \ldots, K)$ in $\mathbb{R}^K$ are K linearly independent cointegrating vectors, because $\boldsymbol{e}_i \boldsymbol{z}_t$ is stationary.
- Can there be more than K linearly independent cointegrating vectors?
- No, since $\boldsymbol{e}_i$ $(i = 1, \ldots, K)$ already span the whole $\mathbb{R}^K$!
- Hence:

$$\boxed{\text{There are } K \text{ linearly independent cointegrating vectors in a stationary VAR.}}$$

- But what if there is a unit root, i.e.,

$$|\boldsymbol{I}_K - \phi_1 z - \ldots - \phi_p z^p| = 0 \qquad \text{for } z = 1,$$

  with all other roots lie outside the unit circle?

- (As for univariate processes, we do not consider roots inside the unit circle, as these lead to—empirically irrelevant—explosive processes.)

- Here, it will turn out that the number of linearly independent cointegrating vectors equals the cointegration rank (introduced on the next slide).

Definition 7.9:

The VAR in (7.6) is called **cointegrated of rank r** if $\mathrm{rk}(\boldsymbol{\Pi}) = r$, where

$$\boldsymbol{\Pi} := \boldsymbol{\phi}_1 + \ldots + \boldsymbol{\phi}_p - \boldsymbol{I}_K.$$

- Thus, $\boldsymbol{\Pi}$ can be factorized as $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(K \times r)$-matrices with $\mathrm{rk}(\boldsymbol{\alpha}) = \mathrm{rk}(\boldsymbol{\beta}) = r$.

Definition 7.10:

The matrix $\boldsymbol{\beta}$ is called **cointegrating matrix** and $\boldsymbol{\alpha}$ the **loading matrix**. The $(K - r) \times K$-matrix $\boldsymbol{\alpha}_\perp$ is the **orthogonal complement matrix** of $\boldsymbol{\alpha}$ that satisfies

$$\underset{K\times(K-r)}{\boldsymbol{\alpha}_\perp'} \cdot \underset{K\times r}{\boldsymbol{\alpha}} = \underset{(K-r)\times r}{\mathbf{0}}.$$

The components of $\boldsymbol{\eta}_t = \boldsymbol{\alpha}_\perp' \boldsymbol{z}_t$ are called the **common trends**.

- In a cointegrated VAR of rank r, it will turn out that

  $$\boxed{\text{the number of linearly independent cointegrating vectors is precisely } r}$$

  and the cointegrating vectors are given as the columns of β.
- To see this, we introduce error correction models next...

- The VAR $\boldsymbol{z}_t$ from (7.6) can be written as an **vector error correction model (VECM)**

$$\Delta \boldsymbol{z}_t = \boldsymbol{\Pi} \boldsymbol{z}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{z}_{t-i} + \boldsymbol{a}_t \tag{7.7}$$
$$= \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{z}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{z}_{t-i} + \boldsymbol{a}_t,$$

where

$$\boldsymbol{\Gamma}_j = -(\boldsymbol{\phi}_{j+1} + \ldots + \boldsymbol{\phi}_p), \qquad j = 1, \ldots, p-1. \tag{7.8}$$

- Note how this generalizes (7.5).
- The term $\boldsymbol{\beta}' \boldsymbol{z}_{t-1}$ can be regarded as a ‘compensation’ term for the overdifferenced system $\Delta \boldsymbol{z}_t$.

- The proof of (7.7) for general VAR(p) models is left as an exercise.
- We show (7.7) here for a VAR(2)

$$\boldsymbol{z}_t = \phi_1 \boldsymbol{z}_{t-1} + \phi_2 \boldsymbol{z}_{t-2} + \boldsymbol{a}_t, \tag{7.9}$$

- Subtract $\boldsymbol{z}_{t-1}$ from both sides of (7.9) to obtain

$$\boldsymbol{z}_t - \boldsymbol{z}_{t-1} = (\phi_1 + \phi_2 - \boldsymbol{I}_K)\boldsymbol{z}_{t-1} - \phi_2 \boldsymbol{z}_{t-1} + \phi_2 \boldsymbol{z}_{t-2} + \boldsymbol{a}_t$$

or

$$\Delta \boldsymbol{z}_t = \boldsymbol{\Pi} \boldsymbol{z}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{z}_{t-1} + \boldsymbol{a}_t, \qquad \text{where } \boldsymbol{\Gamma}_1 = -\phi_2.$$

- Suppose that $\boldsymbol{z}_t$ is an $I(1)$ process (which is the most relevant case empirically), such that $\Delta \boldsymbol{z}_t$ is stationary ($I(0)$).

- Then, $\Delta \boldsymbol{z}_{t-i}$ $(i = 0, \ldots, p-1)$ and $\boldsymbol{a}_t$ in the VECM

$$\Delta \boldsymbol{z}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{z}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{z}_{t-i} + \boldsymbol{a}_t$$

  are $I(0)$.

- Hence, $\boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{z}_{t-1}$ must also be $I(0)$.

- Premultiplying $\boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{z}_{t-1}$ with $(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'$, we obtain the $I(0)$ variable

$$(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{z}_{t-1} = \boldsymbol{\beta}' \boldsymbol{z}_{t-1}.$$

- Hence, there are r cointegrating vectors, given by the rows of $\boldsymbol{\beta}'$, which provides the answer to Question 7.8.

Error Correction Models

- It will turn out that VECMs are also useful for estimation; see Chapter 8
  - Why? Because they only contain stationary random variables.

Error Correction Models

Example 7.11:

Suppose that $(z_{1t}, z_{2t})'$ denote prices of some commodity in two different markets. Assume that in equilibrium $z_{1t} = \beta z_{2t}$ and that changes in z_{1t} and z_{2t} depend on the deviations from this equilibrium in $t-1$.

This can be modelled by the simple error correction model

$$\begin{aligned}
\Delta z_{1t} &= \alpha_1(z_{1,t-1} - \beta_1 z_{2,t-1}) + a_{1t}, \\
\Delta z_{2t} &= \alpha_2(z_{1,t-1} - \beta_1 z_{2,t-1}) + a_{2t}.
\end{aligned}$$

If there is an 'error' in the sense that there is a deviation $z_{1,t-1} > \beta_1 z_{2,t-1}$ ($z_{1,t-1} < \beta_1 z_{2,t-1}$) from equilibrium, then this is 'corrected' through the changes Δz_{1t} and Δz_{2t}.

In this example, we have $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ and $\boldsymbol{\beta} = (1, -\beta_1)'$.

Consider again the general VECM from (7.7):

$$\Delta \boldsymbol{z}_t = \boldsymbol{\Pi} \boldsymbol{z}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{z}_{t-i} + \boldsymbol{a}_t.$$

We distinguish 3 cases for $\mathrm{rk}(\boldsymbol{\Pi})$:

(I) $\mathrm{rk}(\boldsymbol{\Pi}) = 0$: $\boldsymbol{\Pi} = 0$;

(II) $0 < \mathrm{rk}(\boldsymbol{\Pi}) < K$: $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$;

(III) $\mathrm{rk}(\boldsymbol{\Pi}) = K$: $|\boldsymbol{\Pi}| = |-\boldsymbol{\phi}(1)| \neq 0$.

Error Correction Models

I. $\operatorname{rk}(\boldsymbol{\Pi}) = 0$

In case $\operatorname{rk}(\boldsymbol{\Pi}) = 0$, it follows that:

- $\boldsymbol{\Pi} = 0$
- Because $\boldsymbol{z}_t$ is not cointegrated, there does not exist a linear combination of the $I(1)$ variables that is stationary.
- The VECM reduces to a stationary VAR$(p-1)$ in differences:

$$\Delta \boldsymbol{z}_t = \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{z}_{t-i} + \boldsymbol{a}_t.$$

In case $0 < \text{rk}(\boldsymbol{\Pi}) < K$, it follows that:

- We can factorize $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' \neq \mathbf{0}$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(K \times r)$-matrices with $\text{rk}(\boldsymbol{\alpha}) = \text{rk}(\boldsymbol{\beta}) = r$.
- The VECM becomes

$$\Delta \boldsymbol{z}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{z}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{z}_{t-i} + \boldsymbol{a}_t.$$

- This means that $\boldsymbol{z}_t$ is cointegrated with r linearly independent cointegrating vectors, giving $\boldsymbol{w}_t = \boldsymbol{\beta}' \boldsymbol{z}_t$.
- $\boldsymbol{z}_t$ has $(K - r)$ unit roots that give $(K - r)$ **common stochastic trends** of $\boldsymbol{z}_t$.

Overall there are

- r cointegrating relations (eigenvalues of $\boldsymbol{\Pi}$ different from 0) giving the stationary r-dimensional $\boldsymbol{w}_t = \beta' \boldsymbol{z}_t$ and
- $(K - r)$ stochastic trends $\boldsymbol{\alpha}'_{\perp} \boldsymbol{z}_t$ (see below)

In case $\boldsymbol{\Pi}$ has full rank ($\text{rk}(\boldsymbol{\Pi}) = r = K$), it follows that:

- $|\boldsymbol{\Pi}| = |-\phi(1)| \neq 0$.
- $\boldsymbol{z}_t$ has no unit root, that is $\boldsymbol{z}_t$ is $I(0)$.
- There are $(K - r) = 0$ stochastic trends.
- As a consequence, we model the relationship of the $\boldsymbol{z}_t$'s in levels, not in differences.
- In particular, there is no need to refer to the error correction representation.

Error Correction Models

II. $0 < \text{rk}(\boldsymbol{\Pi}) < K$ (continued)

- A general way to obtain the $(K-r)$ common trends is to use the orthogonal complement $(K-r) \times K$-matrix $\boldsymbol{\alpha}_\perp$ of $\boldsymbol{\alpha}$.
- If the VECM is pre-multiplied by $\boldsymbol{\alpha}'_\perp$, the error correction term vanishes:

$$\boldsymbol{\alpha}'_\perp \boldsymbol{\Pi} = (\boldsymbol{\alpha}'_\perp \boldsymbol{\alpha})\boldsymbol{\beta}' = \mathbf{0}.$$

- Thus, since $\boldsymbol{\alpha}'_\perp \Delta \boldsymbol{z}_t = \Delta(\boldsymbol{\alpha}'_\perp \boldsymbol{z}_t)$,

$$\Delta(\boldsymbol{\alpha}'_\perp \boldsymbol{z}_t) = \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta(\boldsymbol{\alpha}'_\perp \boldsymbol{z}_{t-i}) + (\boldsymbol{\alpha}'_\perp \boldsymbol{a}_t).$$

- The resulting system is a $(K-r)$-dimensional system of first differences, corresponding to $(K-r)$ independent random walks

$$\boldsymbol{\alpha}'_\perp \boldsymbol{z}_t,$$

which are the common trends.

Example 7.11: Continued.

Since $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$, we have $\boldsymbol{\alpha}_\perp = (-1, \alpha_1/\alpha_2)$. Hence, we easily obtain the common trends in Example 7.11.

Error Correction Models

Non-Uniqueness of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$

- The choice of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is not unique:
- For any orthogonal $r \times r$-matrix $\boldsymbol{\Omega}$ with $\boldsymbol{\Omega}\boldsymbol{\Omega}' = \boldsymbol{I}_r$,

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}(\boldsymbol{\Omega}\boldsymbol{\Omega}')\boldsymbol{\beta}' = (\boldsymbol{\alpha}\boldsymbol{\Omega})(\boldsymbol{\beta}\boldsymbol{\Omega})' =: \boldsymbol{\alpha}^*(\boldsymbol{\beta}^*)',$$

  where both $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ are of rank r.
- To obtain a unique representation, usually the normalization

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{I}_r \\ \boldsymbol{\beta}_{(K-r)} \end{pmatrix} \tag{7.10}$$

  is imposed, where $\boldsymbol{\beta}_{(K-r)}$ is an arbitrary $(K-r) \times r$-matrix.
- The normalization in (7.10) is always possible if the variables are arranged in a suitable way; see Lütkepohl (2005, p. 250) for possible exercises on this.

3. Inclusion of Deterministic Trends

Inclusion of Deterministic Trends

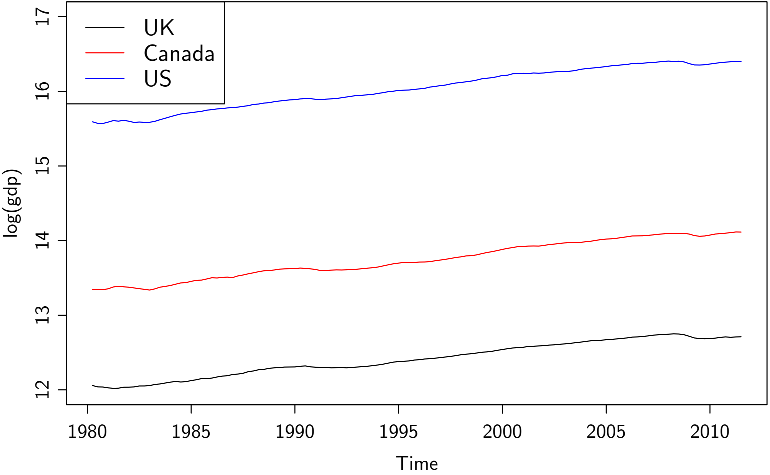
- In the discussion of cointegrated VAR models we have ignored deterministic terms so far.
- However, these terms are empirically very relevant; see the following example.
- From the discussion of the random walk with drift, it should be clear that deterministic terms have a different impact in a VAR with unit roots than in a stationary VAR.
- For example,
  - an intercept term in a random walk generates a linear trend in the mean, whereas
  - an intercept term in a stationary AR process just implies a constant mean value.

Example 7.12:

We plot the logarithm of GDP (in millions of local currency) in the UK / Canada / US over time. The data are from Example 3.5.

```
UK.gdp = ts(gdp[, 1], start = c(1980, 2), frequency = 4)
CA.gdp = ts(gdp[, 2], start = c(1980, 2), frequency = 4)
US.gdp = ts(gdp[, 3], start = c(1980, 2), frequency = 4)
plot(UK.gdp, ylab="log(gdp)", ylim=c(12, 17))
lines(CA.gdp, col="red")
lines(US.gdp, col="blue")
legend("topleft", legend=c("UK", "Canada", "US"),
       lty=c(1, 1, 1), col=c("black", "red", "blue"))
```

Inclusion of Deterministic Trends II



Inclusion of Deterministic Trends

- To explore the implications of the deterministic term, we consider

$$\boxed{\boldsymbol{z}_t = \boldsymbol{\mu}_t + \boldsymbol{x}_t,} \tag{7.11}$$

  where $\boldsymbol{x}_t$ is a (possibly cointegrated) zero-mean VAR(p) process and $\boldsymbol{\mu}_t$ the **deterministic term**.

- Advantage of (7.11):
  - Mean of $\boldsymbol{z}_t$ is clearly specified and does not have to be derived from model parameters.

- Disadvantage of (7.11):
  - Stochastic part $\boldsymbol{x}_t$ is not directly observable.

- Therefore, we rewrite the model in terms of the observable $\boldsymbol{z}_t$ for (e.g.) estimation purposes.

Inclusion of Deterministic Trends

- Suppose that we have the following VECM representation for $\boldsymbol{x}_t$:

$$\begin{aligned}\Delta \boldsymbol{x}_t &= \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{x}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{x}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t \\ &= \boldsymbol{\Pi} \boldsymbol{x}_{t-1} + \ldots \end{aligned} \tag{7.12}$$

- For the trend we consider the empirically most relevant cases:

  1. $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$ (Intercept-only)
  2. $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$ (Intercept & Trend)

where $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ are elements of $\mathbb{R}^K$.

Inclusion of Deterministic Trends

Case 1: Intercept-only (No Constant)

- For $\boldsymbol{\mu}_t = 0$, we have $\boldsymbol{x}_t = \boldsymbol{z}_t$ such that $\Delta \boldsymbol{z}_t = \Delta \boldsymbol{x}_t$ and, hence,

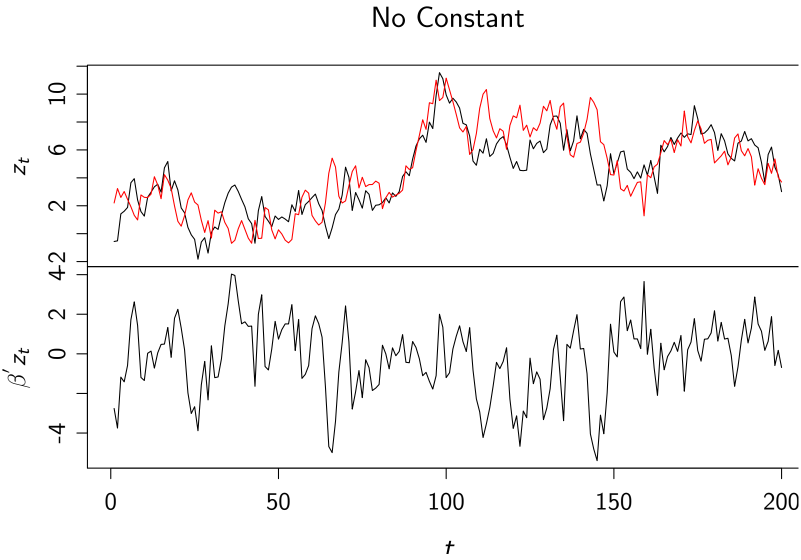
$$\begin{aligned}\Delta \boldsymbol{z}_t &= \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{z}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{z}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{z}_{t-p+1} + \boldsymbol{a}_t \\ &= \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{z}_{t-1} + \ldots .\end{aligned}$$

- (In this section, the $\ldots$ stand for $\boldsymbol{\Gamma}_1 \Delta \boldsymbol{z}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{z}_{t-p+1} + \boldsymbol{a}_t$.)
- Thus, all components of $\boldsymbol{z}_t$ are $I(1)$ (or $I(0)$) without a drift and the cointegrating relations $\boldsymbol{\beta}'\boldsymbol{z}_t$ have zero mean.

(no component has a drift)

Inclusion of Deterministic Trends

Case 1: Intercept-only (No Constant)



Inclusion of Deterministic Trends

Case 2: Intercept-only (Restricted Constant)

- For $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$, we have $\boldsymbol{x}_t = \boldsymbol{z}_t - \boldsymbol{\mu}_0$ such that $\Delta\boldsymbol{z}_t = \Delta\boldsymbol{x}_t$ and, hence,

$$\Delta\boldsymbol{z}_t = \boldsymbol{\nu}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{z}_{t-1} + \ldots,$$

  where $\boldsymbol{\nu}_0 = -\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\mu}_0$.

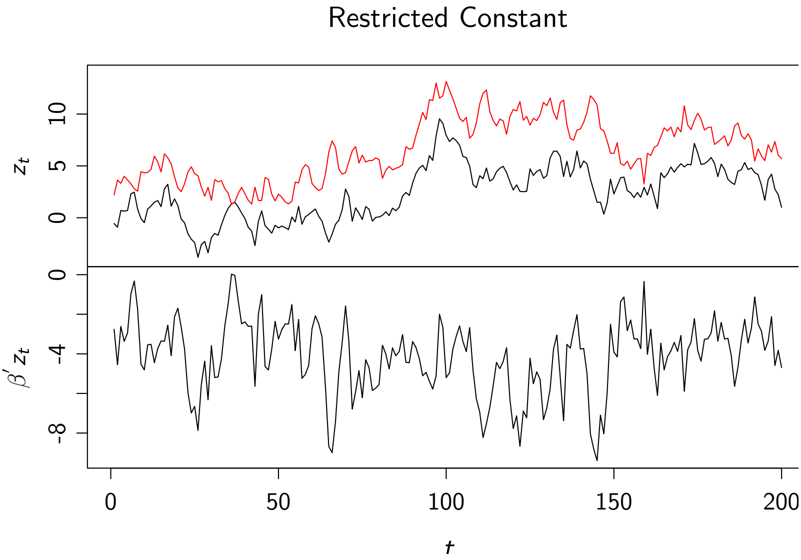
- Thus, a constant mean in the additive representation (7.11) becomes an intercept in the cointegrating relationship, since

$$\boldsymbol{\nu}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{z}_{t-1} = \boldsymbol{\alpha}\boldsymbol{\beta}'(\boldsymbol{z}_{t-1} - \boldsymbol{\mu}_0).$$

- Thus, all components of $\boldsymbol{z}_t$ are $I(1)$ (or $I(0)$) without a drift and the cointegrating relations have a non-zero mean.

(no component has a drift)

Case 2: Intercept-only (Restricted Constant)



Inclusion of Deterministic Trends

Case 4: Intercept & Trend (Restricted Trend)

- If $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$, we have $\boldsymbol{x}_t = \boldsymbol{z}_t - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 t$ and $\Delta \boldsymbol{x}_t = \Delta \boldsymbol{z}_t - \boldsymbol{\mu}_1$ and, hence, from (7.12)

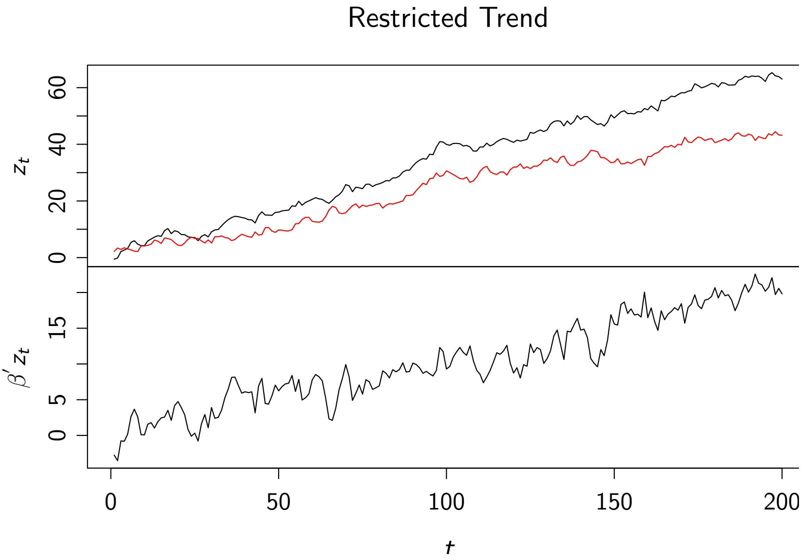
$$\Delta \boldsymbol{z}_t = \boldsymbol{\nu} + \boldsymbol{\alpha} \begin{pmatrix} \boldsymbol{\beta}' & \boldsymbol{\eta}' \end{pmatrix} \begin{pmatrix} \boldsymbol{z}_{t-1} \\ t-1 \end{pmatrix} + \ldots, \tag{7.13}$$

where

$$\begin{aligned} \boldsymbol{\nu} &= -\boldsymbol{\Pi} \boldsymbol{\mu}_0 + (\boldsymbol{I}_K - \boldsymbol{\Gamma}_1 - \ldots - \boldsymbol{\Gamma}_{p-1}) \boldsymbol{\mu}_1, \\ \boldsymbol{\eta}' &= -\boldsymbol{\beta}' \boldsymbol{\mu}_1. \end{aligned}$$

- Now, $\boldsymbol{\nu}$ is unrestricted and can take on any value in $\mathbb{R}^K$ (depending of course on $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_1$, and the other parameters)
- In contrast, the trend term can be absorbed into the cointegrating relationship.
- Thus, the components in $\boldsymbol{z}_t$ are $I(1)$ (or $I(0)$) with a drift vector $\boldsymbol{\nu}$ and the cointegrating relations have a linear trend.

(at least one component drifts)



Restricted Trend
z_t
0
20
40
60
$\beta' z_t$
0
5
15
0
50
100
150
200
t

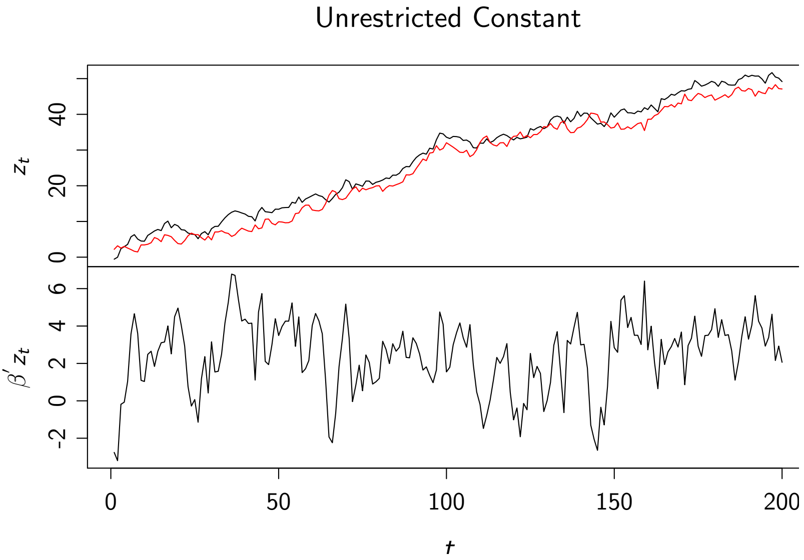
Inclusion of Deterministic Trends

Case 3: Intercept-only (Unrestricted Constant)

- It may happen that $\beta' \boldsymbol{\mu}_1 = \mathbf{0}$ and, hence, $\boldsymbol{\eta} = \mathbf{0}$ in (7.13).
- The VECM is then
$$\Delta \boldsymbol{z}_t = \boldsymbol{\nu} + \boldsymbol{\alpha} \beta' \boldsymbol{z}_{t-1} + \dots .$$
- Hence,
  1. there is a linear trend in the variables (generated by the intercept $\boldsymbol{\nu}$),
  2. but not in the cointegrating relations.
- This situation can arise if the trend slope is the same for all variables, which have a linear trend.
- The components in $\boldsymbol{z}_t$ are $I(1)$ (or $I(0)$) with drift vector $\boldsymbol{\nu}_0$ and the cointegrating relations $\beta' \boldsymbol{z}_t$ have a non-zero mean.
  - Intercept $\boldsymbol{\nu}$ may be split in a drift component and a constant vector in the cointegrating relations

(at least one component drifts)

- The VECM with unrestricted $\boldsymbol{\nu}$ is popular in applied work, because a trend in the cointegrating relationships is sometimes regarded as implausible.
  - Since cointegrating relationships are equilibrium relationships, it is implausible that the variables are driven apart by a deterministic trend.


Unrestricted Constant
z_t
$\beta' z_t$
0
20
40
6
4
2
0
-2
0
50
100
150
200
t

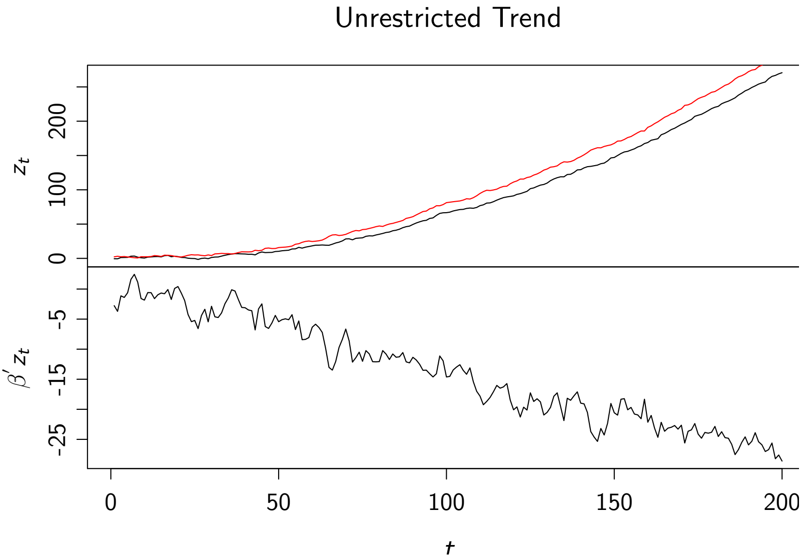
- Model (7.13) with unrestricted intercept and trend, i.e.,

$$\Delta \boldsymbol{z}_t = \boldsymbol{\nu}_0 + \boldsymbol{\nu}_1 t + \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{z}_{t-1} + \ldots,$$

  generates $I(1)$ components in $\Delta \boldsymbol{z}_t$ with a linear trend (quadratic trend in levels) and the cointegrating relations $\boldsymbol{\beta}' \boldsymbol{z}_t$ have a linear trend.

- (Strictly speaking, the model is not covered under the framework of (7.11).)

(at least one component has a quadratic trend)



Unrestricted Trend
z_t
0
100
200
$\beta' z_t$
-5
-15
-25
0
50
100
150
200
t

Inclusion of Deterministic Trends

Summary

- If no component of $\Delta \boldsymbol{z}_t$ drifts, we use Case 1 or 2
  - Case 1 is not really relevant for empirical work.
  - The restricted constant Case 2 is appropriate for non-trending $I(1)$ data like interest rates and exchange rates.
- If at least one component of $\Delta \boldsymbol{z}_t$ drifts, we use Case 3 or 4
  - The unrestricted constant Case 3 is appropriate for trending $I(1)$ data like asset prices, macroeconomic aggregates (real GDP, consumption, employment, etc.)
  - The restricted trend Case 4 is also appropriate for trending $I(1)$ as in Case 3. However, notice the deterministic trend in the cointegrating residual in Case 4 as opposed to the stationary residuals in Case 3.
- If at least one component of $\Delta \boldsymbol{z}_t$ has a quadratic trend, we use Case 5.
  - The unrestricted trend Case 5 is appropriate for $I(1)$ data with a quadratic trend. An example might be nominal price data during times of extreme inflation.

Remark 7.13:

Why is the specification of the deterministic trend important? In particular, why not always consider the most general Case 5?
It is important for testing (e.g., cointegration tests): Choosing the 'correct' specification of the trend leads to more powerful tests! Intuitively, suppose Case 1 is the true model, yet one operates under Case 3. Yet, Case 3 allows for common linear trends in the variables, which will make it much harder to establish cointegration between the variables than if the trend is not present (as in Case 1).

Remark 7.14:

Then, how can we choose between the different trend specifications? Use AIC, BIC or some other information criterion; see Chapter 9.1.

4. Forecasting Cointegrated VARs

- As in Chapter 5, we assume model parameters to be known.
- Forecasting the deterministic term $\boldsymbol{\mu}_t$ is trivial, such that we focus on the stochastic VAR process

$$\boldsymbol{x}_t = \phi_1 \boldsymbol{x}_{t-1} + \ldots + \phi_p \boldsymbol{x}_{t-p} + \boldsymbol{a}_t. \tag{7.14}$$

- Again, the h-step minimum MSE predictor is given by the conditional expectation, which leads to the recursion

$$\boldsymbol{x}_T(h) = \phi_1 \boldsymbol{x}_T(h-1) + \ldots \phi_p \boldsymbol{x}_T(h-p),$$

  where $x_T(j) = x_{T+j}$ for $j \leq 0$.

- Similarly as in (5.7) and (5.8) it can be shown that the forecast error $\boldsymbol{e}_T(h) = \boldsymbol{x}_{T+h} - \boldsymbol{x}_T(h)$ is

$$\boldsymbol{e}_T(h) = \sum_{i=0}^{h-1} \boldsymbol{\theta}_i \boldsymbol{a}_{T+h-i}$$

with forecast MSE matrix (recall (2.9))

$$\operatorname{cov}(\boldsymbol{e}_T(h)) = \sum_{i=0}^{h-1} \boldsymbol{\theta}_i \boldsymbol{\Sigma}_{\boldsymbol{a}} \boldsymbol{\theta}_i'.$$

- The important difference is that for a stationary VAR

$$\operatorname{cov}(\boldsymbol{e}_T(h)) \underset{(h\to\infty)}{\longrightarrow} \operatorname{cov}(\boldsymbol{z}_T),$$

whereas for a cointegrated VAR some elements of $\operatorname{cov}(\boldsymbol{e}_T(h))$ will approach infinity as $h \to \infty$; see Exercises (Lütkepohl, 2005, pp. 260-261)

Remark 7.15: Deterministic Trends.

If $\boldsymbol{z}_t = \boldsymbol{\mu}_t + \boldsymbol{x}_t$, the h-step minimum MSE predictor of $\boldsymbol{z}_{T+h}$ is

$$\boldsymbol{z}_T(h) = \boldsymbol{\mu}_{T+h} + \boldsymbol{x}_T(h).$$

Remark 7.16: Unknown Parameters.

The parameters $\phi_1, \ldots, \phi_p, \boldsymbol{\Sigma}_{\boldsymbol{a}}$ and $\boldsymbol{\mu}_t$ are usually unknown. We will not explore this issue further and instead refer to Lütkepohl (2005, Chapter 7) for some discussion.

5. Causality Analysis

Causality Analysis

- From the discussion above, it follows easily that Theorem 6.4 carries over to cointegrated VARs

- More precisely, write $\boldsymbol{z}_t$ in (7.14) as in (6.1), viz.

$$\boldsymbol{z}_t = \begin{pmatrix} \boldsymbol{x}_t \\ \boldsymbol{y}_t \end{pmatrix} = \begin{pmatrix} \phi_{0,1} \\ \phi_{0,2} \end{pmatrix} + \sum_{i=1}^{p} \begin{pmatrix} \phi_{i,11} & \phi_{i,12} \\ \phi_{i,21} & \phi_{i,22} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_{t-i} \\ \boldsymbol{y}_{t-i} \end{pmatrix} + \begin{pmatrix} \boldsymbol{a}_{1t} \\ \boldsymbol{a}_{2t} \end{pmatrix}$$

- Then, $\boldsymbol{y}_t$ ($\boldsymbol{x}_t$) does not Granger-cause $\boldsymbol{x}_t$ ($\boldsymbol{y}_t$) if and only if $\phi_{i,12} = 0$ ($\phi_{i,21}$) for $i = 1, \ldots, p$.

- It is also easy to derive the corresponding restriction for the VECM

$$\begin{pmatrix} \boldsymbol{x}_t \\ \boldsymbol{y}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_{t-1} \\ \boldsymbol{y}_{t-1} \end{pmatrix} + \sum_{i=1}^{p-1} \begin{pmatrix} \boldsymbol{\Gamma}_{i,11} & \boldsymbol{\Gamma}_{i,12} \\ \boldsymbol{\Gamma}_{i,21} & \boldsymbol{\Gamma}_{i,22} \end{pmatrix} \begin{pmatrix} \Delta \boldsymbol{x}_{t-i} \\ \Delta \boldsymbol{y}_{t-i} \end{pmatrix} + \begin{pmatrix} \boldsymbol{a}_{1t} \\ \boldsymbol{a}_{2t} \end{pmatrix}.$$

- It follows immediately from (7.8) that $\phi_{i,12} = 0$ for $i = 1, \ldots, p$ is equivalent to

$$\boldsymbol{\Pi}_{12} = \mathbf{0} \quad \text{and} \quad \boldsymbol{\Gamma}_{i,12} = 0 \quad \text{for } i = 1, \ldots, p-1.$$

- In other words, to check Granger-causality, we just have to test linear hypotheses.
  - For cointegrated VARs this is more difficult than for stationary VARs.

6. Impulse Response Analysis

Impulse Response Analysis

- From Chapter 7.4, we see that the elements of $\boldsymbol{\theta}_i = (\theta_{i,jk})$ matrices may represent impulse responses—just as in the stationary case.
- Again, $\theta_{i,jk}$ represents the response of variable j to a unit shock in variable k, i periods ago.
- For stationary VARs, $\theta_{i,jk} \underset{(i\to\infty)}{\longrightarrow} 0$, which may not be the case for cointegrated VARs.

- For stationary VARs, we have also considered orthogonalized impulse responses and forecast error variance decompositions.
- The tools are available for cointegrated VARs as well, using precisely the same formulas as in Chapter 6.
- The only quantity that cannot be computed in general is the matrix of total long-run effects $\underline{\boldsymbol{\theta}_{\infty}}$.

Outline

Estimation of VECMs

- In this chapter, we discuss estimation of VECMs with known order p and cointegration rank r.
- Specifically, we cover
  1. ML estimation without deterministic trends, and
  2. ML estimation including deterministics trends
- Of course, there are a number of other estimation techniques available, including LS estimation; see Lütkepohl (2005, Chapter 7).

1. ML estimation without deterministic trends

ML Estimation

- We consider the VECM

$$\Delta \boldsymbol{z}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{z}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{z}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{z}_{t-p+1} + \boldsymbol{a}_t \qquad (8.1)$$

  with $0 < \text{rk}(\boldsymbol{\Pi}) = r < K$, $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$.

- The error $\boldsymbol{a}_t \sim (\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}})$ is a standard white noise
- Also, $\boldsymbol{z}_t$ is assumed to be an $I(1)$ process.
- These conditions are assumed to hold without further notice in this chapter.
- Let the sample be given by $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T$.

- For conciseness, we rewrite (8.1), i.e.,

$$\Delta \boldsymbol{z}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{z}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{z}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{z}_{t-p+1} + \boldsymbol{a}_t$$

in matrix notation as

$$\Delta \boldsymbol{Z} = \boldsymbol{\Pi} \boldsymbol{Z}_{-1} + \boldsymbol{\Gamma} \Delta \boldsymbol{X} + \boldsymbol{A}, \tag{8.2}$$

where

$$\begin{cases} \Delta \boldsymbol{Z} & = (\Delta \boldsymbol{z}_{p+1}, \ldots, \Delta \boldsymbol{z}_T), \\ \boldsymbol{Z}_{-1} & = (\boldsymbol{z}_p, \ldots, \boldsymbol{z}_{T-1}), \\ \boldsymbol{\Gamma} & = (\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}), \\ \boldsymbol{A} & = (\boldsymbol{a}_{p+1}, \ldots, \boldsymbol{a}_T), \end{cases} \qquad \Delta \boldsymbol{X} = \begin{pmatrix} \Delta \boldsymbol{z}_p & \Delta \boldsymbol{z}_{p+1} & \ldots & \Delta \boldsymbol{z}_{T-1} \\ \vdots & \vdots & & \vdots \\ \Delta \boldsymbol{z}_1 & \Delta \boldsymbol{z}_2 & \ldots & \Delta \boldsymbol{z}_{T-p+1} \end{pmatrix}.$$

ML Estimation

- As for ML Estimation of VAR models, we assume $\boldsymbol{a}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}})$.
- As in (3.14) (see also (3.6)), the log-likelihood function is

$$
\begin{aligned}
l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}_{\boldsymbol{a}}) = &-\frac{K}{2}(T-p)\log(2\pi) - \frac{T-p}{2}\log(|\boldsymbol{\Sigma}_{\boldsymbol{a}}|) \\
&- \frac{1}{2}\operatorname{tr}\left[(\Delta\boldsymbol{Z} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{Z}_{-1} - \boldsymbol{\Gamma}\Delta\boldsymbol{X})'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\Delta\boldsymbol{Z} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{Z}_{-1} - \boldsymbol{\Gamma}\Delta\boldsymbol{X})\right].
\end{aligned}
\quad (8.3)
$$

- The ML estimator is given by

$$
\boxed{\left(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{a}}\right) = \underset{(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}_{\boldsymbol{a}})}{\arg\max}\ l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}_{\boldsymbol{a}}).}
$$

ML Estimation

Remark 8.1:

- The ML estimates can be given in closed form (Lütkepohl, 2005, Prop. 7.3). For instance,

$$\widehat{\beta}' = (\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_r)' \boldsymbol{S}_{11}^{-1/2},$$
$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{S}_{01} \widehat{\beta} (\widehat{\beta}' \boldsymbol{S}_{11} \widehat{\beta})^{-1},$$

  where

$$\boldsymbol{M} = \boldsymbol{I}_T - \Delta \boldsymbol{X}' (\Delta \boldsymbol{X} \Delta \boldsymbol{X}')^{-1} \Delta \boldsymbol{X},$$
$$\boldsymbol{R}_0 = \Delta \boldsymbol{Z} \boldsymbol{M}, \quad \boldsymbol{R}_1 = \boldsymbol{Z}_{-1} \boldsymbol{M},$$
$$\boldsymbol{S}_{ij} = \boldsymbol{R}_i \boldsymbol{R}_j' / (T - p), \quad i = 0, 1,$$

  and $\lambda_1 \geq \ldots \geq \lambda_K$ are the eigenvalues of $\boldsymbol{S}_{11}^{-1/2} \boldsymbol{S}_{10} \boldsymbol{S}_{00}^{-1} \boldsymbol{S}_{01} \boldsymbol{S}_{11}^{-1/2}$ with corresponding eigenvectors $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_K$.
- Yet, $\widehat{\alpha}$ and $\widehat{\beta}$ are not unique, because for any invertible matrix $\boldsymbol{Q}$, $\widehat{\alpha} \boldsymbol{Q}^{-1}$ and $\widehat{\beta} \boldsymbol{Q}'$ are also valid ML estimators.

Theorem 8.2: Lütkepohl (2005), Proposition 7.4.

We have that for some singular covariance matrix $\boldsymbol{\Sigma}_{co}$,

$$\sqrt{T}\,\mathrm{vec}\left[(\widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}' \;:\; \widehat{\boldsymbol{\Gamma}}) - (\boldsymbol{\Pi} \;:\; \boldsymbol{\Gamma})\right] \to_{\mathrm{d}} N(\mathbf{0}, \boldsymbol{\Sigma}_{co}), \qquad \text{as } T \to \infty.$$

Remark 8.3:

The matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are not identified (not unique) and, hence, cannot be estimated consistently. To do so, one may impose the normalization

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{I}_r \\ \boldsymbol{\beta}_{(K-r)} \end{pmatrix}$$

to identify $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. In this case, a (super-)consistent estimator of $\boldsymbol{\beta}_{(K-r)}$ is

$$\widehat{\boldsymbol{\beta}}_{(K-r)} = \text{ last } (K-r) \text{ rows of } \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}_r^{-1},$$

where $\widehat{\boldsymbol{\beta}}_r$ consists of the first r rows of $\widehat{\boldsymbol{\beta}}$. More precisely, $T(\widehat{\boldsymbol{\beta}}'_{(K-r)} - \boldsymbol{\beta}'_{(K-r)})$ converges in distribution to a non-standard limit.

Remark 8.4:

1. The reason for the singularity of $\boldsymbol{\Sigma}_{co}$ is the superconsistency of $\widehat{\boldsymbol{\beta}}_{(K-r)}$. The singularity implies that t-tests can be carried out as usual. However, F-tests of linear restrictions on the parameters may not have the usual asymptotic χ^2-distributions that are obtained for stationary processes.
2. The assumption $\boldsymbol{a}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}})$ is not essential for Theorem 8.2. The theorem also holds under weaker conditions when quasi ML estimators based on the Gaussian likelihood-function are considered.

ML Estimation I

Example in R

```
coint1 <- ca.jo(cbind(Aaa, Baa), ecdet="none", K=2, spec="transitory")
beta   <- coint1@V
alpha  <- coint1@W
Pi     <- coint1@PI
Gam    <- coint1@GAMMA
summary(coint1)
```

```
##
## ######################
## # Johansen-Procedure #
## ######################
##
## Test type: maximal eigenvalue statistic (lambda max) , with linear trend
##
## Eigenvalues (lambda):
## [1] 0.054773 0.004665
##
## Values of teststatistic and critical values of test:
##
##           test 10pct  5pct  1pct
## r <= 1 |  2.84  6.50  8.18 11.65
## r = 0  | 34.19 12.91 14.90 19.19
##
```

ML Estimation II

Example in R

```
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##          Aaa.l1 Baa.l1
## Aaa.l1  1.0000  1.000
## Baa.l1 -0.8857 -2.724
##
## Weights W:
## (This is the loading matrix)
##
##          Aaa.l1   Baa.l1
## Aaa.d -0.04697 0.002477
## Baa.d  0.04047 0.002140
```

```
coint2 = ca.jo(cbind(Aaa, Baa), ecdet="const", K=2)
summary(coint2)
```

ML Estimation III

Example in R

```
##
## ######################
## # Johansen-Procedure #
## ######################
##
## Test type: maximal eigenvalue statistic (lambda max) , without linear trend and c
##
## Eigenvalues (lambda):
## [1] 5.477e-02 4.878e-03 2.375e-19
##
## Values of teststatistic and critical values of test:
##
##           test 10pct  5pct  1pct
## r <= 1 |  2.97  7.52  9.24 12.97
## r = 0  | 34.19 13.75 15.67 20.20
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##              Aaa.l2 Baa.l2 constant
## Aaa.l2    1.000000  1.000    1.000
## Baa.l2   -0.885675 -2.702   -3.315
## constant -0.003492 16.400  -13.777
```

ML Estimation IV

Example in R

```
##
## Weights W:
## (This is the loading matrix)
##
##         Aaa.l2   Baa.l2   constant
## Aaa.d -0.04700 0.002507 -4.587e-18
## Baa.d  0.04044 0.002166  4.529e-18
```

2. ML estimation with deterministic trends

Including Deterministics Trends

- Assume the observed process $\boldsymbol{z}_t$ can be represented as

$$\boldsymbol{z}_t = \boldsymbol{\mu}_t + \boldsymbol{x}_t,$$

where $\boldsymbol{x}_t$ has a VECM representation as in (8.1) and $\boldsymbol{\mu}_t$ is the deterministic trend.

- As in Chapter 7, we can set up the VECM for the observed $\boldsymbol{z}_t$ as

$$\begin{aligned}\boldsymbol{z}_t &= \boldsymbol{\alpha}(\boldsymbol{\beta}' \ \boldsymbol{\eta}')\begin{pmatrix}\boldsymbol{z}_{t-1}\\ \boldsymbol{D}_{t-1}^{co}\end{pmatrix} + \boldsymbol{C}\boldsymbol{D}_t + \dots \\ &=: \boldsymbol{\Pi}^+ \boldsymbol{z}_{t-1}^+ + \boldsymbol{C}\boldsymbol{D}_t + \dots,\end{aligned}$$

where $\boldsymbol{D}_t^{co}$ contains all deterministic terms in the cointegrating relation, $\boldsymbol{D}_t$ contains all remaining deterministics, and $\boldsymbol{\eta}'$ and $\boldsymbol{C}$ are the corresponding parameter matrices.

UNIVERSITÄT DUISBURG ESSEN

- As in (8.2), we write

$$\Delta \boldsymbol{Z} = \boldsymbol{\Pi}^{+} \boldsymbol{Z}_{-1}^{+} + \boldsymbol{\Gamma}^{+} \Delta \boldsymbol{X}^{+} + \boldsymbol{A},$$

where

$$\begin{cases} \boldsymbol{Z}_{-1}^{+} & = (\boldsymbol{z}_p^{+}, \ldots, \boldsymbol{z}_{T-1}^{+}), \\ \boldsymbol{\Gamma}^{+} & = (\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}, \boldsymbol{C}), \end{cases} \qquad \Delta \boldsymbol{X}^{+} = \begin{pmatrix} \Delta \boldsymbol{z}_p & \Delta \boldsymbol{z}_{p+1} & \ldots & \Delta \boldsymbol{z}_{T-1} \\ \vdots & \vdots & & \vdots \\ \Delta \boldsymbol{z}_1 & \Delta \boldsymbol{z}_2 & \ldots & \Delta \boldsymbol{z}_{T-p+1} \\ \boldsymbol{D}_{p+1} & \boldsymbol{D}_{p+2} & \ldots & \boldsymbol{D}_T \end{pmatrix}.$$

- The ML estimator can be obtained analogously as before.
- Hence, the computation is equally easy as in the case without deterministic terms.

Including Deterministics Trends

- The log-likelihood function is

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Gamma}^{+}, \boldsymbol{\Sigma_a}) = -\frac{K}{2}(T-p)\log(2\pi) - \frac{T-p}{2}\log(|\boldsymbol{\Sigma_a}|)$$
$$-\frac{1}{2}\operatorname{tr}\left[(\Delta \boldsymbol{Z} - \boldsymbol{\alpha}\beta' \boldsymbol{Z}_{-1}^{+} - \boldsymbol{\Gamma}^{+}\Delta \boldsymbol{X}^{+})' \boldsymbol{\Sigma_a}^{-1}(\Delta \boldsymbol{Z} - \boldsymbol{\alpha}\beta' \boldsymbol{Z}_{-1}^{+} - \boldsymbol{\Gamma}^{+}\Delta \boldsymbol{X}^{+})\right]. \tag{8.4}$$

- The ML estimator is given by

$$\boxed{\left(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\Gamma}}^{+}, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{a}}\right) = \underset{(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Gamma}^{+}, \boldsymbol{\Sigma_a})}{\arg\max}\, l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Gamma}^{+}, \boldsymbol{\Sigma_a}).}$$

Including Deterministics Trends

- The asymptotic properties of the parameter estimators are as before.
- However, asymptotic properties of the deterministic terms require some care, because the convergence rates depend on the specific terms included.
  - For example, if linear trends are included, the convergence rate of the slope parameter are different from $\sqrt{T}$.

Outline

Specification of VECMs

- In specifying VECMs, the lag order and the cointegration rank have to be determined.
- After that, the adequacy of the fitted model has to be checked.
- Hence, we will discuss
  1. lag order selection,
  2. tests for the cointegration rank,
  3. model checking.
- (Since many tests for the cointegration rank require knowledge of the lag order, we first discuss lag order choice.)

1. Lag Order Selection

Lag Order Selection

- It turns out that the consistency property of the information criteria from Chapter 4 (AIC, BIC, HQ) is maintained for integrated processes.

- Formally, we have the following analogue of Theorem 4.4:

Theorem 9.1: Lütkepohl (2005), Proposition 8.1.

Let $\boldsymbol{z}_t$ be a K-dim VAR(p) with standard white noise $\boldsymbol{a}_t$ and suppose that $|\boldsymbol{I}_K - \phi_1 z - \ldots - \phi_p z^p|$ has d unit roots, yet all other roots are outside the complex unit circle. Suppose the maximum order $P \geq p$ and $\widehat{p}$ is chosen to minimize

$$C(l) = \log|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{a},l}}| + \frac{l}{T} c_T, \qquad l = 0, 1, \ldots, P,$$

where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{a},l}$ is the Gaussian ML estimate of $\boldsymbol{\Sigma}_{\boldsymbol{a}}$. Then, $\widehat{p}$ is consistent if and only if

$$c_T \to \infty \quad \text{and} \quad \frac{c_T}{T} \to 0 \quad \text{as } T \to \infty.$$

Lag Order Selection

- One can show that

$$\widehat{p}(\mathrm{BIC}) \leq \widehat{p}(\mathrm{HQ}) \leq \widehat{p}(\mathrm{AIC}) \quad \text{for } T \geq 16.$$

- Recall that consistent order selection may not be a relevant objective, because the true process may not admit a finite order VAR representation.
- If some order $\widehat{p}$ is chosen for the VAR, $\widehat{p} - 1$ is the lag order for the corresponding VECM.
- If some variables are known to be integrated, the VAR order must be at least 1 and, hence, model selection can be performed over $1, \ldots, P$ instead of $0, 1, \ldots, P$.

Lag Order Selection

Example in R

Example 9.2:

Consider the trivariate system of log-GDP in the UK, CA and the US from Example 3.5. It is evident from Example 7.12 that a trend needs to be allowed for in choosing the lag order. The two consistent information criteria HQ and BIC choose a lag order of 2.

```
UKCAUS.gdp <- ts(cbind(gdp[, 1], gdp[, 2], gdp[, 3]),
               start = c(1980, 2), frequency = 4, names=c("UK", "CA", "US"))
VARselect(UKCAUS.gdp, lag.max = 5, type="both")
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      4      2      2      4
##
## $criteria
##                  1          2          3          4          5
## AIC(n) -3.073e+01 -3.151e+01 -3.153e+01 -3.158e+01 -3.155e+01
## HQ(n)  -3.059e+01 -3.128e+01 -3.122e+01 -3.118e+01 -3.108e+01
## SC(n)  -3.039e+01 -3.095e+01 -3.076e+01 -3.061e+01 -3.038e+01
## FPE(n)  4.489e-14  2.074e-14  2.037e-14  1.939e-14  1.989e-14
```

2. Tests for the Cointegration Rank

Tests for the Cointegration Rank

- Many different tests have been proposed and the properties of most of them depend on the deterministic terms included in the model.
- Therefore, we discuss models with different deterministic terms separately.
- The general model is again
$$\boldsymbol{z}_t = \boldsymbol{\mu}_t + \boldsymbol{x}_t, \tag{9.1}$$
where $\boldsymbol{\mu}_t$ is the deterministic part and $\boldsymbol{x}_t$ has VECM representation
$$\Delta \boldsymbol{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{x}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{x}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t,$$
where $\boldsymbol{a}_t$ is Gaussian white noise.
- We shall first discuss the case $\boldsymbol{\mu}_t = \mathbf{0}$.

Tests for the Cointegration Rank
$\boldsymbol{\mu}_t = \mathbf{0}$

- In case $\boldsymbol{\mu}_t = \mathbf{0}$, the VECM representation for $\boldsymbol{z}_t$ is

$$\Delta \boldsymbol{z}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{z}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{z}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{z}_{t-p+1} + \boldsymbol{a}_t,$$

  where $\text{rk}(\boldsymbol{\Pi}) = r$ with $0 \leq r \leq K$ and $\boldsymbol{a}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}})$.

- Suppose we wish to test

$$H_0 \;:\; \text{rk}(\boldsymbol{\Pi}) = r_0 \qquad \text{vs.} \qquad H_1 \;:\; r_0 < \text{rk}(\boldsymbol{\Pi}) \leq r_1. \tag{9.2}$$

- The likelihood ratio (LR) statistic for (9.2) can be shown to be

$$\lambda_{LR}(r_0, r_1) = 2\left[\log l(r_1) - \log l(r_0)\right] = -(T-p) \sum_{i=r_0+1}^{r_1} \log(1-\lambda_i),$$

  where $l(r_i)$ denotes the maxmimized Gaussian likelihood function for cointegrating rank r_i (from (8.3)) and λ_i from Remark 8.1.

- Two different pairs of hypothese have received prime attention:

$$H_0 \ : \ \mathrm{rk}(\boldsymbol{\Pi}) = r_0 \quad \text{vs.} \quad H_1 \ : \ r_0 < \mathrm{rk}(\boldsymbol{\Pi}) \leq K, \tag{9.3}$$

$$H_0 \ : \ \mathrm{rk}(\boldsymbol{\Pi}) = r_0 \quad \text{vs.} \quad H_1 \ : \ \mathrm{rk}(\boldsymbol{\Pi}) = r_0 + 1. \tag{9.4}$$

- The LR statistic $\lambda_{LR}(r_0, K)$ for testing (9.3) is the **trace statistic** and the LR statistic $\lambda_{LR}(r_0, r_0 + 1)$ for testing (9.4) is the **maxmimum eigenvalue statistics**.
- Unfortunately, the LR statistics do not have the usual χ^2-limits; instead the limiting distributions depend on the number of common trends $(K - r_0)$ under H_0 and on the alternative hypotheses...

Tests for the Cointegration Rank

$\boldsymbol{\mu}_t = \mathbf{0}$

- Johansen (1995) shows that under $H_0 \ : \ \mathrm{rk}(\boldsymbol{\Pi}) = r_0$, as $T \to \infty$,

$$\lambda_{LR}(r_0, K) \to_{\mathrm{d}} \mathrm{tr}(\mathcal{D}), \tag{9.5}$$

$$\lambda_{LR}(r_0, r_0+1) \to_{\mathrm{d}} \lambda_{\max}(\mathcal{D}), \tag{9.6}$$

where $\lambda_{\max}(\mathcal{D})$ denotes the maxmimum eigenvalue of the matrix

$$\mathcal{D} = \left(\int_0^1 \boldsymbol{W} d\boldsymbol{W}'\right)' \left(\int_0^1 \boldsymbol{W}\boldsymbol{W}' ds\right)^{-1} (\boldsymbol{W} d\boldsymbol{W}).$$

- Here, $\boldsymbol{W} = \boldsymbol{W}_{K-r_0}(s)$ stands for a $(K - r_0)$-dimensional standard Wiener process.
- Fortunately, critical values can easily be simulated from $\mathrm{tr}(\mathcal{D})$ and $\lambda_{\max}(\mathcal{D})$.
- The convergences in (9.5) and (9.6) hold even for (not necessarily Gaussian) standard white noise.

UNIVERSITÄT DUISBURG ESSEN

Strategy for Determining the Cointegration Rank

1. Test the sequence of null hypotheses

$$\begin{aligned} H_0 &: \operatorname{rk}(\boldsymbol{\Pi}) = 0, \\ H_0 &: \operatorname{rk}(\boldsymbol{\Pi}) = 1, \\ &\vdots \\ H_0 &: \operatorname{rk}(\boldsymbol{\Pi}) = K - 1 \end{aligned}$$

   using either the trace statistic or the maximal eigenvalue statistic.

2. Terminate the tests at the first rejection and choose the cointegration rank accordingly.
3. If none of the null hypotheses is rejected, this is evidence for $\operatorname{rk}(\boldsymbol{\Pi}) = K$ and, hence, one chooses a stationary VAR model.

Tests for the Cointegration Rank

- We shall now treat the cases
  - $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$
  - $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$
- The LR statistics in the above cases are now obtained from the corresponding maximized likelihoods in (8.4).
- For instance, if $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$, then

$$\Delta \boldsymbol{z}_t = \boldsymbol{\nu} + \boldsymbol{\Pi}^+ \boldsymbol{z}_{t-1}^+ + \dots,$$

where

$$\begin{aligned}
\boldsymbol{\nu} &= \boldsymbol{\Pi}\boldsymbol{\mu}_0 + (\boldsymbol{I}_K - \boldsymbol{\Gamma}_1 - \dots - \boldsymbol{\Gamma}_{p-1})\boldsymbol{\mu}_1, \\
\boldsymbol{\Pi}^+ &= (\boldsymbol{\Pi} \ : \ \boldsymbol{\nu}_1), \quad \text{with } \boldsymbol{\nu}_1 = -\boldsymbol{\Pi}\boldsymbol{\mu}_1, \\
\boldsymbol{z}_{t-1}^+ &= \begin{pmatrix} \boldsymbol{z}_{t-1} \\ t-1 \end{pmatrix}.
\end{aligned}$$

Tests for the Cointegration Rank

Theorem 9.3: Lütkepohl (2005), Proposition 8.2.

Suppose $\boldsymbol{z}_t = \boldsymbol{\mu}_t + \boldsymbol{x}_t$ is as in (9.1). Then, under $H_0: \ \mathrm{rk}(\boldsymbol{\Pi}) = r_0$, as $T \to \infty$,

$$\lambda_{LR}(r_0, K) \to_{\mathrm{d}} \mathrm{tr}(\mathcal{D}),$$
$$\lambda_{LR}(r_0, r_0+1) \to_{\mathrm{d}} \lambda_{\max}(\mathcal{D}),$$

where

$$\mathcal{D} := \left(\int_0^1 \boldsymbol{F} d\boldsymbol{W}_{K-r_0}' \right)' \left(\int_0^1 \boldsymbol{F}\boldsymbol{F}' ds \right)^{-1} \left(\int_0^1 \boldsymbol{F} d\boldsymbol{W}_{K-r_0}' \right)$$

with

(1) $\boldsymbol{F}(s) = \boldsymbol{W}_{K-r_0}(s)$, if $\boldsymbol{\mu}_t = 0$,

(2) $\boldsymbol{F}(s) = (\boldsymbol{W}_{K-r_0}'(s) \ : \ 1)'$, if $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$,

(3) $\boldsymbol{F}(s) = (\overline{\boldsymbol{W}}(s) \ : \ s - 1/2)'$, if $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$,

(4) $\boldsymbol{F}(s) = \breve{\boldsymbol{W}}(s)$, if $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t, \ \boldsymbol{\mu}_1 \neq 0$ and $\beta' \boldsymbol{\mu}_1 = 0$.

Remark 9.4:

- The precise form of $\overline{\boldsymbol{W}}(s)$ and $\breve{\boldsymbol{W}}(s)$ is of no relevance here. The main point is that different limits arise for different deterministic trends.
- Trace and maximum eigenvalue tests have similar power in many situations; see also the exercises.
- It is also possible to derive LR tests for other deterministics terms, e.g.,
  - higher-order polynomial trends,
  - seasonal dummies,
  - ...

Remark 9.5:

- The correct specification of the deterministic trend is extremely important to obtain powerful tests.
  - If you allow for (non-zero) deterministic trends, cointegration will be much harder to establish.
- The choice of $\boldsymbol{\mu}_t$ in (1)–(4) involves a robustness-efficiency trade-off.
  - Allowing for more general trends increases robustness, ...
  - ... while efficiency is reduced.

Example 9.5: Continued.

Since the log-GDP series are drifting, we should try both Case 3 and Case 4 and choose the better one. For Case 3, the below output allows us to determine the rank r of $\boldsymbol{\Pi}$. The estimated eigenvalues λ_i are sorted from largest to smallest. The best model in terms of AIC and BIC is the restricted trend model (Case 4). According to the maximal eigenvalue test, a model with cointegrating rank $r = 1$ is chosen.

```
C3 <- VECM(UKCAUS.gdp, lag=1, estim="ML", include = "const")    # For Case 3
C4 <- VECM(UKCAUS.gdp, lag=1, estim="ML", LRinclude = "trend")  # For Case 4
data.frame(AIC = AIC(C3), BIC = BIC(C3))
```

```
##     AIC   BIC
## 1 -3853 -3805
```

```
data.frame(AIC = AIC(C4), BIC = BIC(C4))
```

```
##     AIC   BIC
## 1 -3865 -3817
```

```
summary(rank.test(C4))
```

Tests for the Cointegration Rank II

Example in R

```
##   r  trace trace_pval trace_pval_T  eigen eigen_pval
## 1 0 37.380        0.2          0.2 26.062       0.04
## 2 1 11.318        0.9          0.9  8.645       0.76
## 3 2  2.673        0.9          0.9  2.673       0.90
```

```r
C4$coefficients
```

```
##                  ECT Intercept  UK -1  CA -1   US -1
## Equation UK -0.03260   -0.2019 0.4401 0.1555 0.05062
## Equation CA  0.07902    0.4948 0.1711 0.3257 0.35931
## Equation US  0.16030    1.0042 0.2933 0.3462 0.10177
```

```r
t(C4$model.specific$beta)
```

```
##    UK      CA      US    trend
## r1  1 -0.7515 -0.5299 0.002612
```

3. Model Checking

Model Checking

- Diagnostic checking is also an important stage in modeling VECMs.
- The tests discussed for stationary VAR processes can be extended to VECMs

$$\Delta \boldsymbol{z}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{z}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{z}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{z}_{t-p+1} + \boldsymbol{a}_t$$

  with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ $(K \times r)$–matrices of rank r and $\boldsymbol{a}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}})$.
- We assume the model has been estimated by ML as outlined in Chapter 8.
- Denote the residuals by $\widehat{\boldsymbol{a}}_t$.

Model Checking

- Everything is as in Chapter 4:
- Let $\boldsymbol{R}_l$ be the theoretical lag-l CCM of the innovations $\boldsymbol{a}_t$.
- The hypothesis of interest in model checking is

$$\boxed{H_0: \ \boldsymbol{R}_1 = \ldots = \boldsymbol{R}_m = \mathbf{0} \quad \text{vs.} \quad H_1: \ \boldsymbol{R}_j \neq \mathbf{0} \text{ for some } 1 \leq j \leq m,}$$

where m is a prespecified integer.

- The Ljung–Box test statistic is

$$\boxed{Q_K(m) = T^2 \sum_{l=1}^{m} \frac{1}{T-l} \widehat{\boldsymbol{B}}_l \left(\widehat{\boldsymbol{R}}_0^{-1} \otimes \widehat{\boldsymbol{R}}_0^{-1} \right) \widehat{\boldsymbol{B}}_l',}$$

where $\widehat{\boldsymbol{B}}_l = \mathrm{vec}(\widehat{\boldsymbol{R}}_l')$.

Model Checking

Asymptotic Properties

Theorem 9.6: Lütkepohl (2005), p. 346.

Under suitable regularity conditions, as $T \to \infty$,

$$Q_K(m) \to_{\mathrm{d}} \chi^2_{mK^2-(p-1)K^2-Kr}.$$

Remark 9.7:

Note that the degrees of freedom are adjusted relative to the stationary VAR case in Theorem 4.7, where

$$Q_K(m) \to_{\mathrm{d}} \chi^2_{(m-p)K^2}.$$

Now we substract from the number of autocovariances included in the statistic (mK^2) the number of estimated parameters—not counting the elements of the $(K \times r)$-cointegration matrix β (since this matrix can be estimated T-consistently).

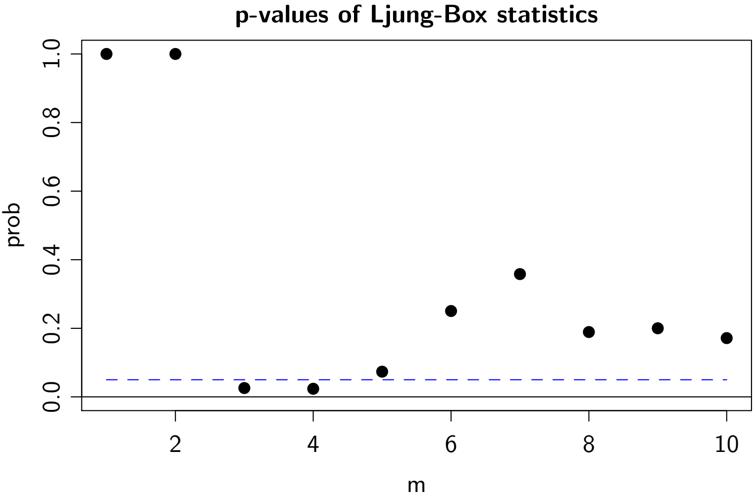
Again, it can be shown that deterministic terms do not affect the asymptotic distribution of the LB-test for residual autocorrelation.

Example 9.7: Continued.

How well does the restricted trend VECM with $r = 1$ cointegrating relationship fit the data? We apply the Portmanteau test from Theorem 9.6 to find out. Judging from the Ljung-Box statistics, the VECM provides a reasonable fit.

```
mq(C4$residuals, lag=10, adj=12)  # adjustment by (p-1)K^2 + Kr = 1*3^2 + 3*1 = 12
```

```
## Ljung-Box Statistics:
##          m       Q(m)     df    p-value
##  [1,]  1.00      7.36   -3.00      1.00
##  [2,]  2.00     14.33    6.00      1.00
##  [3,]  3.00     27.38   15.00      0.03
##  [4,]  4.00     39.58   24.00      0.02
##  [5,]  5.00     45.42   33.00      0.07
##  [6,]  6.00     47.75   42.00      0.25
##  [7,]  7.00     54.07   51.00      0.36
##  [8,]  8.00     69.45   60.00      0.19
##  [9,]  9.00     78.64   69.00      0.20
## [10,] 10.00     89.73   78.00      0.17
```

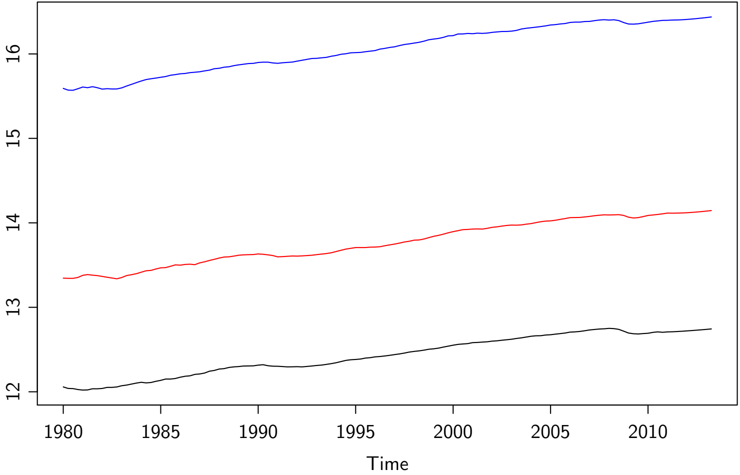


p-values of Ljung-Box statistics
prob
m
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10

Example 9.7: Continued.

Now that the VECM C4 has been found suitable, we can use it to forecast log-GDP (and hence GDP).

```
pred_C4 <- predict(C4, n.ahead = 8)
ee4     <- ts(rbind(UKCAUS.gdp, pred_C4), start = 1980, frequency = 4)
ts.plot(ee4, type = "l", col=c("black", "red", "blue"))
```



16
15
14
13
12
1980
1985
1990
1995
2000
2005
2010
Time

Outline

 References I

Akaike, H. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle." In *2nd International Symposium of Information Theory*, edited by B. N. Patrov and F. Csaki, 257:267–81. Budapest: Akademia Kiado.

Granger, C. W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37: 424–38.

Granger, C. W. J., and P. Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2 (2): 111–20.

Hannan, E. J., and B. G. Quinn. 1979. "The Determination of the Order of an Autoregression." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 41: 190–95.

Johansen, S. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Phillips, P. C. B. 1986. "Understanding Spurious Regressions in Econometrics." *Journal of Econometrics* 33 (3): 311–40.

References II

Schwarz, G. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6: 461–64.

Zellner, A. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57: 348–68.