Week 10: Multivariate Time Series analysis VAR models & IRFs, VECM

Advanced Econometrics 4EK608

Vysoká škola ekonomická v Praze

Content

VAR: Vector autoregression models

VAR model: introduction

Univariate autoregressive models (AR models) describe specific time-varying processes in nature, economy, etc.

- AR processes/models may be either stationary or non-stationary.
- AR(p) model (autoregressive model of order p): the modelled variable depends linearly on its own previous values and a stochastic term

AR(p): $y_t = c + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_p y_{t-p} + u_t = c + \sum_{i=1}^{p} \rho_i y_{t-i} + u_t$

VAR models generalize the univariate autoregressive model (AR model) by allowing for more than one evolving variable:
y_t becomes $\boldsymbol{y}_t$, where $\boldsymbol{y}_t' = (y_{1t}, y_{2t}, \ldots, y_{mt})$

- VAR models capture linear interdependencies among multiple time series
- Atheoretical: The only prior knowledge required for VAR modeling is a list of variables which can be hypothesized to affect each other inter-temporally

VAR: Vector autoregression models

VAR models: origins

C. Sims (Nobel price) reacted in the 1980ies against SEMs. His arguments:

- Large scale macroeconomic models failed in providing governments with adequate economic forecasts.
- For identification, some SEM variables must be omitted (so called zero restrictions on parameters). Such omissions are often arbitrary and lack justification (hence undermine model credibility).
- Endogenous/exogenous division of variables tends to be arbitrary.

VAR models vs. SEMs:

- Even very simple VAR models usually provide more realistic predictions of variables involved (compared to SEMs).
- In VAR models, all variables are treated as endogenous.
- No zero-restrictions placed on parameters (although such restrictions may be easily applied to VAR models if necessary).
- VAR models make for a versatile toolbox, many extensions and generalizations are possible.

VAR: Vector autoregression models

VAR model: notation

- We work with m-variable (m-dimensional) VAR(p) models
- A two-variable VAR(3) model may be denoted as follows:

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{A}_2\boldsymbol{y}_{t-2} + \boldsymbol{A}_3\boldsymbol{y}_{t-3} + \boldsymbol{u}_t = \boldsymbol{c} + \sum_{i=1}^{3} \boldsymbol{A}_i\boldsymbol{y}_{t-i} + \boldsymbol{u}_t$$

or:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{11,2} & a_{12,2} \\ a_{21,2} & a_{22,2} \end{bmatrix} \begin{bmatrix} y_{1t-2} \\ y_{2t-2} \end{bmatrix} + \\ + \begin{bmatrix} a_{11,3} & a_{12,3} \\ a_{21,3} & a_{22,3} \end{bmatrix} \begin{bmatrix} y_{1t-3} \\ y_{2t-3} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

or:

$$y_{1t} = c_1 + a_{11,1}y_{1t-1} + a_{12,1}y_{2t-1} + a_{11,2}y_{1t-2} + a_{12,2}y_{2t-2} + a_{11,3}y_{1t-3} + \\ + a_{12,3}y_{2t-3} + u_{1t}$$

$$y_{2t} = c_2 + a_{21,1}y_{1t-1} + a_{22,1}y_{2t-1} + a_{21,2}y_{1t-2} + a_{22,2}y_{2t-2} + a_{21,3}y_{1t-3} + \\ + a_{22,3}y_{2t-3} + u_{2t}$$

VAR: Vector autoregression models

VAR model: notation

- Any VAR(p) specification can be equivalently rewritten as a VAR(1) by stacking the lags of the VAR(p) variables and by appending identities to complete the number of equations;
- For example, a VAR(2) model $\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{A}_2\boldsymbol{y}_{t-2} + \boldsymbol{u}_t$ can be be written as a VAR(1):

$$\begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_{t-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{0} \end{bmatrix}$$

- In general, any m-dimensional VAR(p) model may be re-written as: $\boldsymbol{Y}_t = \boldsymbol{v} + \boldsymbol{A}\boldsymbol{Y}_{t-1} + \boldsymbol{U}_t$, where:

$$\boldsymbol{Y}_t := \begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_{t-1} \\ \vdots \\ \boldsymbol{y}_{t-p+1} \end{bmatrix}; \ \boldsymbol{v} := \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}; \ \boldsymbol{A} := \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \dots & \boldsymbol{A}_{p-1} & \boldsymbol{A}_p \\ \boldsymbol{I}_k & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_k & & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{I}_k & \boldsymbol{0} \end{bmatrix}; \ \boldsymbol{U}_t := \begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}$$

dim: $(mp \times 1)$ $(mp \times 1)$ $(mp \times mp)$ $(mp \times 1)$

VAR: Vector autoregression models

$$y_{1t} = c_1 + a_{11,1}y_{1t-1} + a_{12,1}y_{2t-1} + a_{11,2}y_{1t-2} + a_{12,2}y_{2t-2} + a_{11,3}y_{1t-3} + \\ + a_{12,3}y_{2t-3} + u_{1t}$$

$$y_{2t} = c_2 + a_{21,1}y_{1t-1} + a_{22,1}y_{2t-1} + a_{21,2}y_{1t-2} + a_{22,2}y_{2t-2} + a_{21,3}y_{1t-3} + \\ + a_{22,3}y_{2t-3} + u_{2t}$$

- All regressors are lagged variables - they can be assumed to be contemporaneously uncorrelated with the disturbances u_{1t}, u_{2t}. Hence, each equation can be consistently estimated by OLS on an individual basis.
- We have identical regressors in all individual VAR model equations ...
  Very often, we experience contemporaneous correlation between endogenous variables ...
  Hence, in practical applications, the elements of $\boldsymbol{u}_t$ tend to be contemporaneously correlated.
- For forecasting into the $t+1$ period, only current and past values of $\boldsymbol{y}$ variables are required (generally, these are readily available).

VAR: Model setup and testing

What do we need to set up a VAR?

A (small) set of endogenous variables

- Six is about the upper limit

A decision on a lag length

- VAR approach assumes the same length for each variable/equation.
- Longer is preferable with this method. Sufficient lags have to be included to ensure non-autocorrelated residuals in all equations.
- On the other hand, because of the degrees of freedom problem, variables frequently have to be excluded from the model and limit has to be placed on the length of lags.

Decide on inclusion of deterministic regressors

- Trends, time & seas. dummies, exogenous regressors (oil price). VAR models augmented by deterministic regressors are often called VARX(p) models.

VAR: Model setup and testing

How do we choose endogenous variables for a VAR model?

Prior information (economic theory) is applied to select model variables

Granger Causality tests are applied for setup verification

- Definition of Granger Causality (GC): X is said to be a Granger cause of Y if present Y can be predicted with greater accuracy by using past values of X rather than not using such past values, all other relevant information being identical.

  Definition easily extends to the situation where X and Y are multidimensional processes.

- GC is a statistical concept {its not an "actual" causality}.
- Different types of GC tests exist. (Sims test, modified Sims test)
- **GC tests apply to stationary series!**

VAR: Model setup and testing

Granger Causality Tests: *F* test for a 2-variable VAR(p) model (Direct GC Test: X $\underset{GC}{\rightarrow}$ Y)

In a VAR(p) model with two variables (Y, X), we can use a simple F test for multiple linear restrictions to test for the H_0 of no Granger causality:
We start with a full (unrestricted) VAR(p) model

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i x_{t-i} + u_t$$

Under H_0 of no GC, all $\beta_i = 0$:

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + u_t$$

Test statistic:

$$F = \frac{(SSR_R - SSR_{UR})}{q(SSR_{UR})}(T-k) \underset{H_0}{\sim} F(q \cdot T-k)$$

where

k - number of estimated parameters in the UR model

q - number of restrictions imposed on the R model

T - number of observations

Wald test for GC

In an m-dimensional VAR(p) model, we can partition $\boldsymbol{y}_t$ into two processes: $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$.

Then, the H_0 of non-causality between $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ (GC-type) may be characterized – and tested – using specific zero constraints on the coefficients of the estimated VAR(p) system.

Wald test may be used as an asymptotical test for such constraints.

(see Lütkepohl: "New introduction to multiple time series analysis")

Limitations and drawbacks in Granger Causality testing

- In practical applications, the "all other relevant information being identical" clause in the GC definition may cause problems, as the results of GC testing between X and Y are sensitive to information (variables, lags) included in the system.
- Data frequency can have important impact. For example, if GC is found in monthly data, this does not necessarily imply GC in daily/weekly/quarterly/annual series of the same variables. The same applies to seasonally adjusted/unadjusted series.
- GC tests are performed on estimated rather than known systems.

VAR: Model setup and testing

How do we decide on the lag-length of a VAR(p) model?

- When estimating VARs or conducting GC tests, results can be sensitive to the lag length of the VAR
- Sometimes the VAR model lag length corresponds to the data, such that quarterly data have 4 lags, monthly data have 12 lags, etc.
- A more rigorous way to determine the optimal lag length is to use the Akaike or Schwarz-Bayesian information criteria (IC).
- However, VAR model estimations tend to be sensitive to the presence of autocorrelation. In such case, after using IC, if there is any evidence of autocorrelation, further lags are added, above the number indicated by the IC, until the autocorrelation is removed.

VAR: Model setup and testing

How do we decide on the lag-length of a VAR(p) model?

- The main information criteria are the Schwarz-Bayesian criteria (SIC, SBIC) and the Akaike criteria (AIC).
- They operate on the basis that there are two competing factors related to adding more lags to a model. More lags will reduce the RSS, but also generate some loss of degrees of freedom (penalty for complexity).
- The aim is to minimize the IC value - adding an extra lag will only benefit the model if the reduction in the RSS outweighs the loss of degrees of freedom.
- In general, the SBIC has a harsher complexity penalty term than the AIC (sometimes leading to smaller p values).

VAR: Model setup and testing

How do we decide on the lag-length of a VAR(p) model?

Single-equation statistics

$$AIC = \log(\hat{\sigma}^2) + \frac{2k}{T}$$

$$SBIC = \log(\hat{\sigma}^2) + \frac{k}{T} \ln T$$

where:
$\hat{\sigma}^2$ - residual variance
T - sample size
k - number of parameters

Multivariate statistics

$$MAIC = \log|\hat{\Sigma}| + \frac{2k'}{T}$$

$$MSBIC = \log|\hat{\Sigma}| + \frac{k'}{T} \log(T)$$

where:
$\hat{\Sigma}$ - covariance matrix of the residuals
T - number of observations
k' - total number of regressors in all equations

VAR: Model setup and testing

Lag-length selection - empirical example:

varsoc oilp igae_s ex_rate impi ppi cpi i_rate if time>= tm(2001m7), maxlag(6)

Selection-order criteria

Sample: 2001m7 - 2013m2 Number of obs = 140

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	1951.92				2.0e-21	-27.7845	-27.7247	-27.6374
1	3360.47	2817.1	49	0.000	7.4e-30	-47.2067	-46.7285	-46.03*
2	3458.16	195.39	49	0.000	3.7e-30*	-47.9023*	-47.0058*	-45.6961
3	3494.82	73.311	49	0.014	4.5e-30	-47.7259	-46.411	-44.4901
4	3528.78	67.922	49	0.038	5.7e-30	-47.5111	-45.7778	-43.2457
5	3562.2	66.846	49	0.046	7.4e-30	-47.2886	-45.1369	-41.9936
6	3603.7	82.998	49	0.002	8.8e-30	-47.1814	-44.6113	-40.8569

Endogenous: oilp igae_s ex_rate impi ppi cpi i_rate

Exogenous: _cons

VAR: Model setup and testing

Stability testing of a VAR model:

A VAR(1) proces
$\boldsymbol{y}_t = \boldsymbol{v} + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{u}_t$ is stable if the following condition is met:

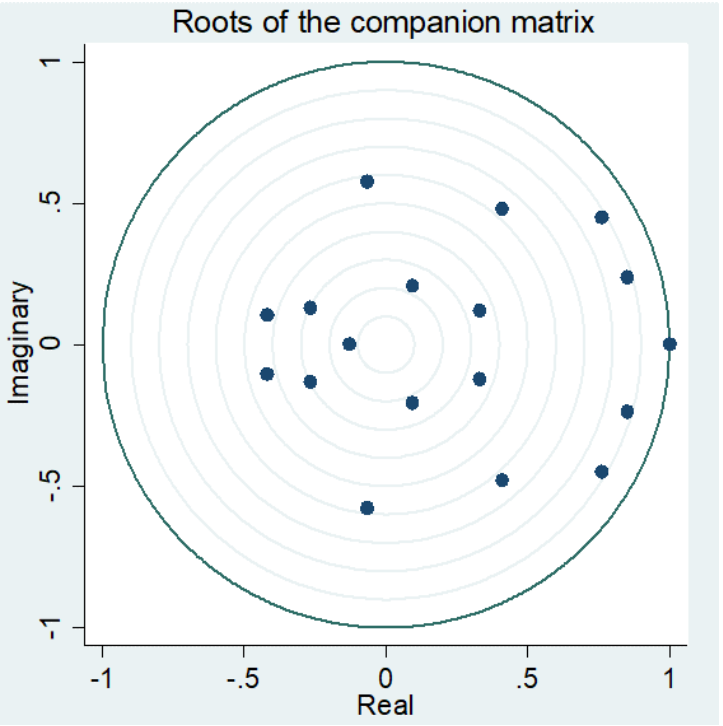
$$det(\boldsymbol{I}_m - \boldsymbol{A}_1 z) \neq 0 \text{ for } |z| \leq 1,$$

alternatively:

$$\lim_{n \to \infty} \boldsymbol{A}_1^n = \boldsymbol{0}_m$$

where:
z is a scalar (number)
$\boldsymbol{I}_m$ – identity matrix, where m is the number of variables in a VAR,
$\boldsymbol{0}_m$ – $(m \times m)$ zero matrix.

Note: Any VAR(p) model may be re-written as VAR(1) ...



Graphical representation of the stability condition (example): If the moduli of the eigenvalues of $\boldsymbol{A}_1$ are less than one, then the VAR(p)-process is stable.

VAR: Model setup and testing

VAR model testing (residuals):

- Serial correlation tests: Portmanteau, Breusch & Godfrey
- Heteroskedasticity: ARCH
- Normality tests: Jarque & Bera, etc.
- Structural stability: CUSUM

Functions **serial()**, **arch()**, **normality()** and **stability()** in package **{vars}**.

VAR-based Forecasting: Introduction

Wold decomposition of stable VAR(p) models:

...concept useful for forecasting and IRF construction.

A simplified stable VAR(p) model
$\boldsymbol{y}_t = \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{A}_2\boldsymbol{y}_{t-2} + \cdots + \boldsymbol{A}_p\boldsymbol{y}_{t-p} + \boldsymbol{u}_t$ can be written as:

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_0\boldsymbol{u}_t + \boldsymbol{\Phi}_1\boldsymbol{u}_{t-1} + \boldsymbol{\Phi}_2\boldsymbol{u}_{t-2}\ldots \quad ,$$

where $\boldsymbol{\Phi}_s = \sum_{j=1}^{s} \boldsymbol{\Phi}_{s-j}\boldsymbol{A}_j$ for $s = 1, 2, \ldots$; $\boldsymbol{A}_j = \boldsymbol{0}$ for $j > p$

$\boldsymbol{\Phi}_0 = \boldsymbol{I}_m$ $(m \times m)$ Identity matrix.

See Lütkepohl: "New introduction to multiple time series analysis" for derivation and discussion.

VAR: Forecasting

Wold decomposition examples:

Stable VAR(2) model:
...simplified: no intercept term

$\boldsymbol{y}_t = \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{A}_2\boldsymbol{y}_{t-2} + \boldsymbol{u}_t$
may be written as:

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_0\boldsymbol{u}_t + \boldsymbol{\Phi}_1\boldsymbol{u}_{t-1} + \\ + \boldsymbol{\Phi}_2\boldsymbol{u}_{t-2}\dots$$

where:
$\boldsymbol{\Phi}_0 = \boldsymbol{I}_m$
$\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_0\boldsymbol{A}_1$
$\boldsymbol{\Phi}_2 = \boldsymbol{\Phi}_1\boldsymbol{A}_1 + \boldsymbol{\Phi}_0\boldsymbol{A}_2$
$\boldsymbol{\Phi}_3 = \boldsymbol{\Phi}_2\boldsymbol{A}_1 + \boldsymbol{\Phi}_1\boldsymbol{A}_2$
...
$\boldsymbol{\Phi}_s = \boldsymbol{\Sigma}_{j=1}^{s}\boldsymbol{\Phi}_{s-j}\boldsymbol{A}_j =$
$= \boldsymbol{\Phi}_{s-1}\boldsymbol{A}_1 + \boldsymbol{\Phi}_{s-2}\boldsymbol{A}_2$

($s = 1, 2, \dots$; $\boldsymbol{A}_j = \boldsymbol{0}$ for $j > p$)

Stable VAR(1) model:
...any VAR(p) may be written as VAR(1)

$\boldsymbol{y}_t = \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{u}_t$
may be written as:

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_0\boldsymbol{u}_t + \boldsymbol{\Phi}_1\boldsymbol{u}_{t-1} + \\ + \boldsymbol{\Phi}_2\boldsymbol{u}_{t-2}\dots$$

where:
$\boldsymbol{\Phi}_0 = \boldsymbol{I}_m$
$\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_0\boldsymbol{A}_1 = \boldsymbol{A}_1$
$\boldsymbol{\Phi}_2 = \boldsymbol{\Phi}_1\boldsymbol{A}_1 = \boldsymbol{A}_1\boldsymbol{A}_1 = \boldsymbol{A}_1^2$
$\boldsymbol{\Phi}_3 = \boldsymbol{\Phi}_2\boldsymbol{A}_1 = \boldsymbol{A}_1^3$
...
$\boldsymbol{\Phi}_s = \boldsymbol{A}_1^s$

(note that $\boldsymbol{\Phi}_0 = \boldsymbol{A}_1^0 = \boldsymbol{I}_m$)

VAR: Forecasting

Iterative forecasting

VAR(p) model is estimated using observations $t = 1, 2, \ldots, T$:

$$\boldsymbol{y}_t = \hat{\boldsymbol{A}}_1 \boldsymbol{y}_{t-1} + \hat{\boldsymbol{A}}_2 \boldsymbol{y}_{t-2} + \cdots + \hat{\boldsymbol{A}}_p \boldsymbol{y}_{t-p} + \hat{\boldsymbol{u}}_t$$

and for $t = T$:

$$\boldsymbol{y}_T = \hat{\boldsymbol{A}}_1 \boldsymbol{y}_{T-1} + \hat{\boldsymbol{A}}_2 \boldsymbol{y}_{T-2} + \cdots + \hat{\boldsymbol{A}}_p \boldsymbol{y}_{T-p} + \hat{\boldsymbol{u}}_T.$$

Arbitrarily long forecasts ($t = T+1, T+2, \ldots, T+h$) can be iteratively produced using the estimated $\boldsymbol{A}$-matrices and the observed ($t = 1, 2, \ldots, T$) and predicted ($t = T+1, T+2, \ldots, T+h$) values of $\boldsymbol{y}_t$:

$$\hat{\boldsymbol{y}}_{T+1} = \hat{\boldsymbol{A}}_1 \boldsymbol{y}_T + \hat{\boldsymbol{A}}_2 \boldsymbol{y}_{T-1} + \cdots + \hat{\boldsymbol{A}}_p \boldsymbol{y}_{T-p+1}$$

$$\hat{\boldsymbol{y}}_{T+2} = \hat{\boldsymbol{A}}_1 \hat{\boldsymbol{y}}_{T+1} + \hat{\boldsymbol{A}}_2 \boldsymbol{y}_T + \cdots + \hat{\boldsymbol{A}}_p \boldsymbol{y}_{T-p+2}$$

$$\hat{\boldsymbol{y}}_{T+3} = \hat{\boldsymbol{A}}_1 \hat{\boldsymbol{y}}_{T+2} + \hat{\boldsymbol{A}}_2 \hat{\boldsymbol{y}}_{T+1} + \cdots + \hat{\boldsymbol{A}}_p \boldsymbol{y}_{T-p+3}$$

$\ldots\ldots$

As we move through the prediction time period, predicted values are used as regressors for subsequent periods & predictions ...

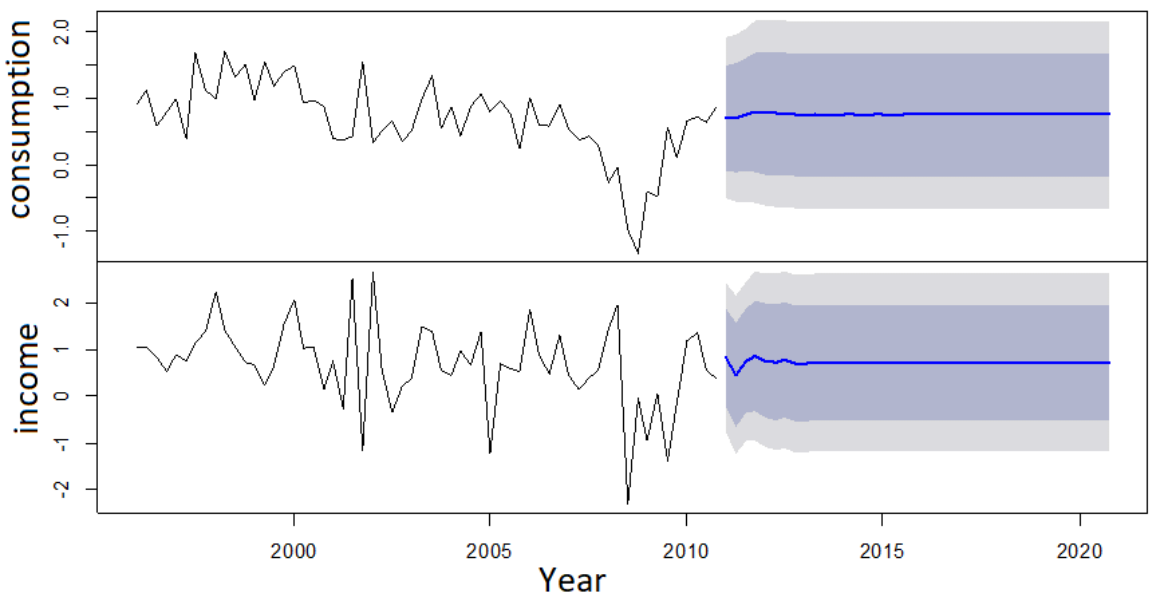
VAR: Forecasting

Forecast error covariance matrix:

$$cov\left(\begin{bmatrix} \boldsymbol{y}_{T+1} - \hat{\boldsymbol{y}}_{T+1} \\ \cdots \\ \boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{\Phi} & \boldsymbol{I} & & \boldsymbol{0} \\ \vdots & & \ddots & \\ \boldsymbol{\Phi}_{h-1} & \boldsymbol{\Phi}_{h-2} & \cdots & \boldsymbol{I} \end{bmatrix} (\boldsymbol{\Sigma}_u \otimes \boldsymbol{I}_h) \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{\Phi}_1 & \boldsymbol{I} & & \boldsymbol{0} \\ \vdots & & \ddots & \\ \boldsymbol{\Phi}_{h-1} & \boldsymbol{\Phi}_{h-2} & \cdots & \boldsymbol{I} \end{bmatrix}'$$

where
$\boldsymbol{\Sigma}_u = cov(\boldsymbol{u}_t)$ is the white noise covariance matrix,
$(\boldsymbol{\Sigma}_u \otimes \boldsymbol{I}_h)$ is a Kroneker product;
$\boldsymbol{I}_h$ is $h \times h$,
$\boldsymbol{\Phi}_i$ are the coefficient matrices of the Wold moving average representation of a stable VAR(p)-process.

Forecast from a VAR(5): example
(confidence levels included)



VAR: Impulse-response functions (IRFs)

Based on Wold decomposition of a stable VAR(p).

IRFs describe the dynamic interactions between endogenous variables. (provided $\boldsymbol{y}_t$ is stationary!)

The $[i, j]$-th elements of the matrices $\boldsymbol{\Phi}_s$ are (interpreted as) the expected response of variable $y_{(i,t+s)}$ to a unit change in variable $y_{j,t}$.

IRF can be cumulated through time: $[i, j]$-th elements of $\boldsymbol{C}_s = \sum_{l=1}^{s} \boldsymbol{\Phi}_l$ measure the accumulated response of variable $y_{i,t+s}$ to a unit change in variable $y_{j,t}$.

IRFs are used for policy analysis: for individual shocks (shocks in different model equations), we can study the dynamic effects on all variables in the model.

Disturbances in different model equations tend to be contemporaneously correlated: we cannot realistically simulate isolated shocks. Solution: model transformation/orthogonalization...

VAR: Impulse-response functions (IRFs)

IRF example:
(from Lütkepohl: "New introduction to multiple time series analysis")

We start with an estimated 3-dimensional VAR(1) system:

$$
\begin{aligned}
y_{1,t} &= 0.5y_{1,t-1} + u_{1,t} \\
y_{2,t} &= 0.1y_{1,t-1} + 0.1y_{2,t-1} + 0.3y_{3,t-1} + u_{2,t} \\
y_{3,t} &= 0.2y_{2,t-1} + 0.3y_{3,t-1} + u_{3,t}
\end{aligned}
$$

This may be re-written as

$$
\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix}
$$

How IRFs work:
Say, at time t, we have a unit disturbance in $y_{2,t}$ (**isolated** contemporaneous disturbance through $u_{2,t}$). At time $t+1$, it causes: $y_{2,t+1}$ changes by 0.1 and $y_{3,t+1}$ changes by 0.2 and $y_{1,t+1}$ is unaffected

Hence, $\boldsymbol{A}_1 = \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix}$ Also, for a VAR(1) model: $\boldsymbol{\Phi}_s = \boldsymbol{A}_1^s$

VAR: Impulse-response functions (IRFs)

IRF example contd.

Using $\boldsymbol{\Phi}_s = \boldsymbol{A}_1^s$, the IRFs may be generated as follows:

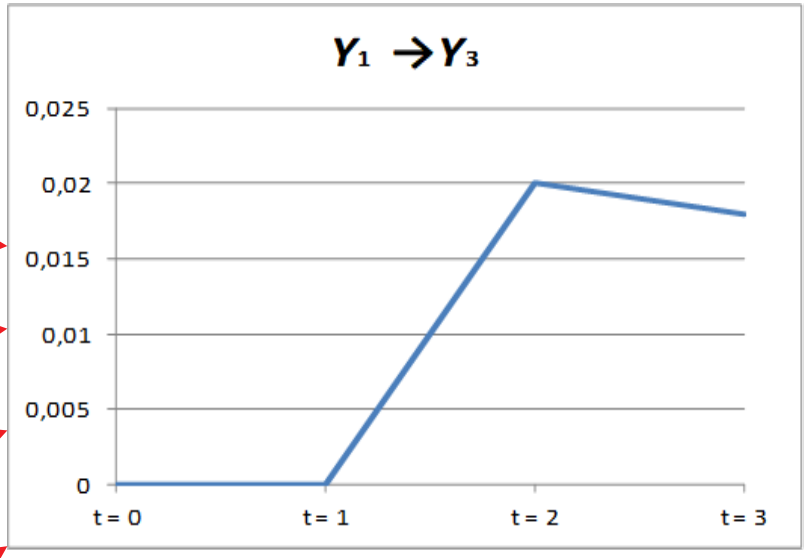
$$\boldsymbol{\Phi}_0 = \boldsymbol{A}_1^0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\boldsymbol{\Phi}_1 = \boldsymbol{A}_1 = \begin{bmatrix} 0.500 & 0 & 0 \\ 0.100 & 0.100 & 0.300 \\ 0 & 0.200 & 0.300 \end{bmatrix},$$

$$\boldsymbol{\Phi}_2 = \boldsymbol{A}_1^2 = \begin{bmatrix} 0.250 & 0 & 0 \\ 0.060 & 0.070 & 0.120 \\ 0.020 & 0.080 & 0.150 \end{bmatrix},$$

$$\boldsymbol{\Phi}_3 = \boldsymbol{A}_1^3 = \begin{bmatrix} 0.125 & 0 & 0 \\ 0.037 & 0.031 & 0.057 \\ 0.018 & 0.038 & 0.069 \end{bmatrix},$$

...



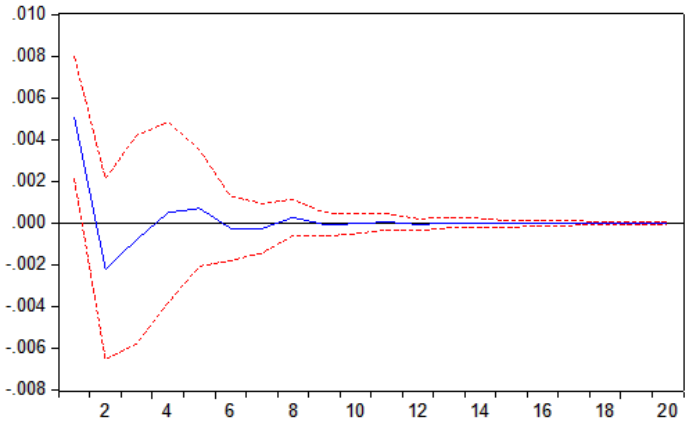
[3,1]-th elements of $\boldsymbol{\Phi}_s = \boldsymbol{A}_1^s$ [VAR(1) model ...] matrices measure the response of variable $y_{3,t+s}$ to a unit change in variable $y_{1,t}$ (a unit $u_{1,t}$ disturbance occurs at $t = 0$).

$\boldsymbol{\Phi}_s$: "Impulses/shocks in columns and responses in rows"

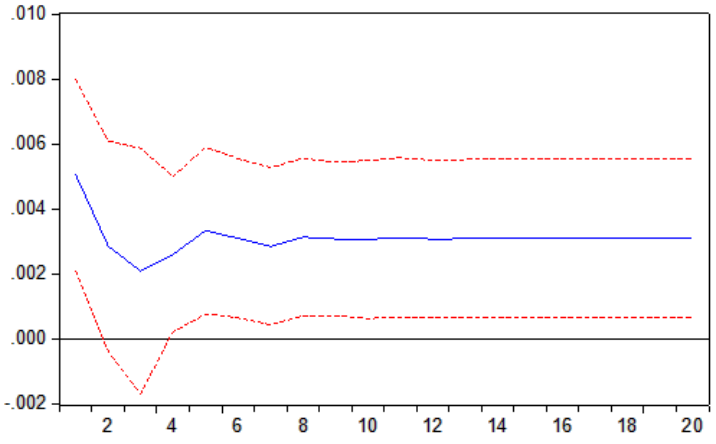
VAR: Impulse-response functions (IRFs)

IRF example contd.
Cumulative IRFs may be easily produced and plotted using $\boldsymbol{C}_s = \sum_{l=1}^{s} \boldsymbol{\Phi}_l$:

IRF (CPI → GDP)



Cumulative IRF (CPI → GDP)



In a stable VAR(p) model, responses to a one-off shock die out over time (shown left). Hence, the accumulated IRF converges to some constant value (shown right).

Important extensions to VARs are based on assumptions (zero restrictions) on the (asymptotic) behavior of $\boldsymbol{C}_s$ (Blanchard-Quah decomposition, see: Lütkepohl: "New introduction to multiple time series analysis")

VAR: Impulse-response functions (IRFs)

IRF – orthogonalization example

Previous example is based on – a very strong – assumption of uncorrelated random effects. Usually, we cannot realistically simulate isolated shocks on observed variables.

2-dimensional VAR(1) model with correlated errors example:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

where $var(u_{1t}) = \sigma_1^2$ and $var(u_{2t}) = \sigma_2^2$,
and, most importantly, $cov(u_{1t}, u_{2t}) = E(u_{1t} \cdot u_{2t}) = cov_{12} \neq 0$.

$\Rightarrow$ It is unrealistic to simulate isolated unit disturbances to $\boldsymbol{u}_t$ as $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

If $cov_{12} \neq 0$, then – for a unit disturbance in u_{1t} – we have:
$dist(\boldsymbol{u}_t) = \begin{bmatrix} 1 \\ E(1 \cdot u_{2t}) \end{bmatrix} = \begin{bmatrix} 1 \\ cov_{12} \end{bmatrix}$

VAR: Impulse-response functions (IRFs)

IRF – orthogonalization example contd.

Base VAR model: $\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$

In our VAR(1) model, responses to a unit disturbance in u_{1t} are:

For a $\boldsymbol{u}_t$ disturbance $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$: $E\Delta \begin{bmatrix} y_{1t+1} \\ y_{2t+1} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix}$

For a $\boldsymbol{u}_t$ disturbance $\begin{bmatrix} 1 \\ cov_{12} \end{bmatrix}$:

$$E\Delta \begin{bmatrix} y_{1t+1} \\ y_{2t+1} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 \\ cov_{12} \end{bmatrix} = \begin{bmatrix} a_{11} + a_{12}\, cov_{12} \\ a_{21} + a_{22}\, cov_{12} \end{bmatrix}$$

It is virtually impossible to study the isolated effects of individual disturbances (analysis is even more complicated for $p > 1$ & $m > 2$)

VAR: Impulse-response functions (IRFs)

IRF – orthogonalization example contd.

Our sample VAR(1) model: $\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$

with $var(u_{1t}) = \sigma_1^2$, $var(u_{2t}) = \sigma_2^2$ and $cov_{12} \neq 0$,
errors may be transformed (orthogonalized) as follows:

$$\begin{bmatrix} y_{1t} \\ (y_{2t} - \delta y_{1t}) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ (a_{21} - \delta a_{11}) & (a_{22} - \delta a_{12}) \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ (u_{2t} - \delta u_{1t}) \end{bmatrix}$$

where $\delta = cov_{12}/\sigma_1^2$ and $cov(u_{1t}, (u_{2t} - \delta u_{1t})) = 0$

Hence, IRFs based on the **transformed model** depict the isolated effects of a given unit disturbance.

VAR: Impulse-response functions (IRFs)

IRF orthogonalization – a generalized approach & notation:

In a VAR(p) model, $\boldsymbol{\Sigma}_u = cov(\boldsymbol{u}_t)$ is the error-term covariance matrix and it may be expressed as $\boldsymbol{\Sigma}_u = \boldsymbol{P}\boldsymbol{P}'$ where $\boldsymbol{P}$ is a lower triangular matrix (assumptions on $\boldsymbol{\Sigma}_u$ apply).

Orthogonalized IRFs from a VAR(p) model
$\boldsymbol{y}_t = \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{A}_2\boldsymbol{y}_{t-2} + \cdots + \boldsymbol{A}_p\boldsymbol{y}_{t-p} + \boldsymbol{u}_t$
may be calculated using the transformed MA representation:

$$\boldsymbol{y}_t = \boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2\boldsymbol{\varepsilon}_{t-2}\ldots,$$

where: $\boldsymbol{\varepsilon}_t = \boldsymbol{P}^{-1}\boldsymbol{u}_t$
$\boldsymbol{\Psi}_i = \boldsymbol{\Phi}_i\boldsymbol{P}$ for $i = 1, 2, \ldots$
$\boldsymbol{\Psi}_0 = \boldsymbol{P}$.

(Use bootstrapped confidence intervals for the orthogonalized IRFs)
. . . see: Lütkepohl: "New introduction to multiple time series analysis" for detailed description.

VAR: Impulse-response functions (IRFs)

IRF orthogonalization – final remarks:

- $\boldsymbol{\Sigma}_u = \boldsymbol{P}\boldsymbol{P}'$ and $\boldsymbol{P}$–based transformation is sensitive to the ordering of equations.
  i.e. different ordering of variables in $\boldsymbol{y}$ often yields different orthogonalized IRFs!

- Orthogonalized innovations are difficult to interpret... Even if y_{1t} and y_{2t} are well defined, the dimension of $(y_{2t} - \delta y_{1t})$ / see previous example / often has no economic interpretation.

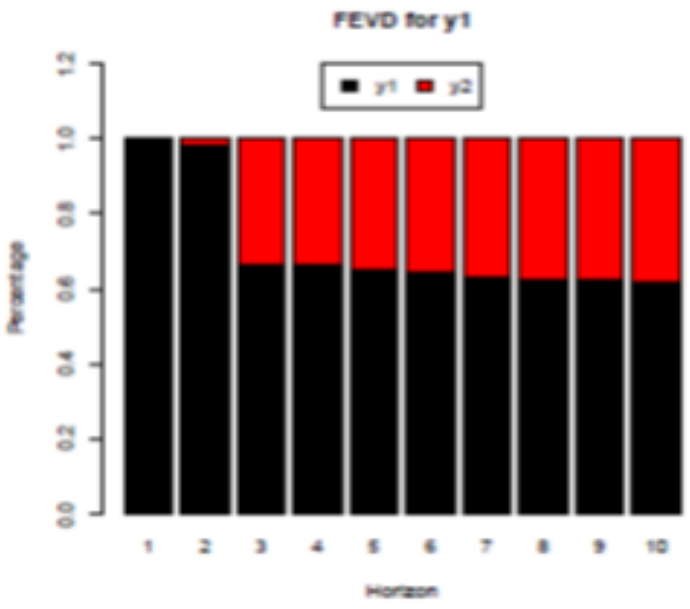
  We treat orthogonalized IRFs as dimensionless series ...

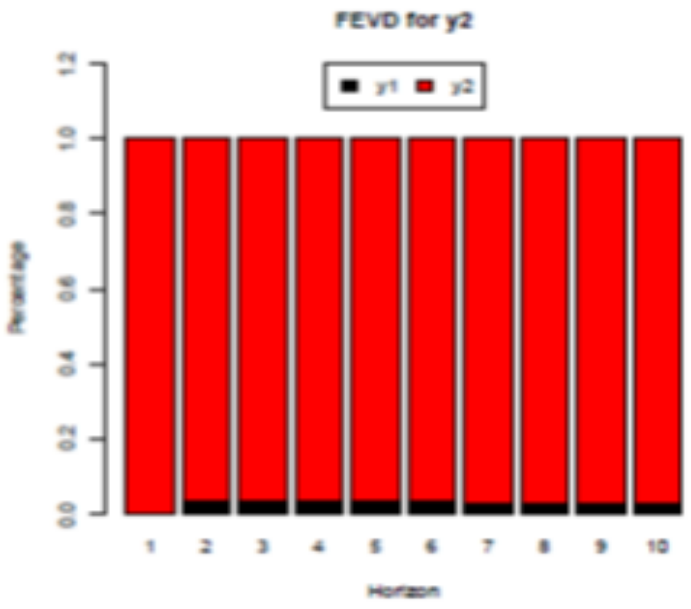
- Generalized orthogonalization approaches exist (IRFs independent of $\boldsymbol{y}$ ordering).
  ...see: Lütkepohl: "New introduction to multiple time series analysis"

VAR: Variance decomposition

Forecast Error Variance Decomposition (FEVD)

- FEVD: based on orthogonalised impulse response coefficient matrices $\mathbf{\Psi}$
- Used to analyse the contribution of variable j to the h-step FEV of variable k.
- **R**: use **fevd()** in **{vars}**





VAR: Final remarks

Are VAR models atheoretical?

- Essentially yes, there are no prior restrictions on parameters in VARs.
- Often, estimated VARs may lead to models consistent with economic theory. For example, for $\boldsymbol{y}_t^T = (Unempl_t, CPI_t)$, VAR($p$) models often generate IRFs consistent with Phillip's theory. For verification, we use causality tests.

IRF critique

- If "important" variables are dropped from a VAR model, IRFs may be significantly distorted.
- In most practical cases (yet, not generally) predictions from such "reduced" VARs might remain largely unaffected.

VAR: Final remarks

Selected VAR-related topics & extensions not covered in this course:

- Structural VAR models: SVARs
- Time-varying (and/or) factor-augmented VARs
- Blanchard-Quah decomposition
- . . . many extensions to VARs exist

VECM: Introduction

- VAR models are well suited to model systems of nonstationary cointegrated variables.
  - Forecasting is possible,
  - IRFs do not converge to zero over time if the underlying series are non-stationary ...
- For cointegrated series, we can use the error correction mechanism (ECM) to model short-time dynamics.
- Such models are named Vector Error Correction Models (VECMs).

  Long term dynamics in the m-dimensional system of variables [given cointegrating relationship(s) exist(s)] is used in a VECM: an ECM-like model in first differences.
- Most of the previous discussion of VAR models can be adequately applied to VECMs.
- VECM-specific topics follow

VECM: Number of cointegrating vectors

For an m-dimensional $I(1)$ / non-stationary / vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_m)'$, there are

- $0 < r < m$ possible cointegrating vectors [$r = 0 \Rightarrow$ non-CI series]
- and $(m - r)$ common stochastic trends.
  (Proof comes from a Beveridge-Nelson decomposition of Δy)

$\boldsymbol{y} = (y_1, y_2)'$: max. 1 linearly independent cointegrating vector: $r \in \{0; 1\}$

- If any $(\alpha y_1 + \beta y_2) \sim I(0)$, we can find ∞ linear combinations of (α, β) that lead to a $I(0)$ processes.
- Hence, it is easy and common to normalize the CI relationship by setting $\alpha = 1$
- If $y_1, y_2 \sim CI(1, 1)$, a cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$ exists, such that $(y_1 - \beta_2 y_2) \sim I(0)$

Example of a CI system for $\boldsymbol{y} = (y_1, y_2)'$ with $\boldsymbol{\beta} = (1, -\beta_2)'$:

[1] $y_{1t} = \beta_2 y_{2t} + u_t,$ $\Leftarrow$ one CI relationship

[2] $y_{2t} = y_{2t-1} + v_t,$ $\Leftarrow$ one common "stochastic trend"

where $u_t, v_t \sim I(0)$

VECM: Number of cointegrating vectors

Ex.: CI system for $\boldsymbol{y} = (y_1, y_2, y_3)'$ with $r = 1$ an $\boldsymbol{\beta} = (1, -\beta_2, -\beta_3)'$:

[1] $y_{1t} = \beta_2 y_{2t} + \beta_3 y_{3t} + u_t$, $\Leftarrow$ one CI relationship

[2] $y_{2t} = y_{2t-1} + v_t$, $\Leftarrow$ first common "stochastic trend"

[3] $y_{3t} = y_{3t-1} + w_t$, $\Leftarrow$ second common "stochastic trend"

where $u_t, v_t, w_t \sim I(0)$

The 1^{st} equation describes the long-run equilibrium, 2^{nd} and 3^{rd} equations describe the $(m - r)$ common stochastic trends.

Possible economic setup for the example:

$y_{1t} \sim I(1)$ nominal F/X rate index for 2 currencies (domestic/foreign),
$y_{2t} \sim I(1)$ domestic inflation index (e.g. CPI),
$y_{3t} \sim I(1)$ foreign inflation.

We can test a hypothesis of stationary combination of the three series:
real exchange rate $\sim I(0)$: under H_0: $\boldsymbol{\beta} = (1, -1, -1)'$
. . . i.e. the 2 identifying CI restrictions imposed are: $\beta_2 = \beta_3 = 1$

VECM: Number of cointegrating vectors

Example of a CI system for $\boldsymbol{y} = (y_1, y_2, y_3)'$ with $r = 2$

[1] $y_{1t} = \beta_{13} y_{3t} + u_t, \quad \Leftarrow$ first CI relationship

[2] $y_{2t} = \beta_{23} y_{3t} + v_t, \quad \Leftarrow$ second CI relationship

[3] $y_{3t} = y_{3t-1} + w_t, \quad \Leftarrow$ a common "stochastic trend"

where $u_t, v_t, w_t \sim I(0)$

Here, we have two CI vectors: $\boldsymbol{\beta}_1 = (1, 0, -\beta_{13})'$ and $\boldsymbol{\beta}_2 = (0, 1, -\beta_{23})'$.

Remember: $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are normalized - yet not unique - stationary combinations of the variables.

Any linear combination: $\boldsymbol{\beta}_3 = c_1 \boldsymbol{\beta}_1 + c_2 \boldsymbol{\beta}_2$ is also a CI vector

$\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ form form the basis (span the space) of cointegrating vectors...

Financial market-based setup for the example:
$y_{1t} \sim I(1)$ Long-term interest rate "1" (say, 3M),
$y_{2t} \sim I(1)$ Long-term interest rate "2" (say, 6M),
$y_{3t} \sim I(1)$ Short-term interest rate (say, overnight).

CI relationships indicate that the spreads between the short-run and long-run rates are stationary, i.e. $I(0)$.

VECM: Construction

Start with a VAR(p) model for an m-dimensional $I(1)$ / non-stationary / vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_m)'$:

$$\boldsymbol{y}_t = \boldsymbol{\Theta} \boldsymbol{D}_t + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{A}_2 \boldsymbol{y}_{t-2} + \cdots + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{u}_t, \qquad t = 1, 2, \ldots, T,$$

where $\boldsymbol{D}_t$ contains deterministic terms (constant, trend, seasonality, etc.)

Even if $\boldsymbol{y}$ series are cointegrated through some CI vector $\boldsymbol{\beta}$, the cointegrating relationship is "hidden" in the VAR(p) representation – it only becomes apparent in first differences-based VECM:

The VECM model is defined as follows:

$$\boldsymbol{\Delta y}_t = \boldsymbol{\Theta} \boldsymbol{D}_t + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \boldsymbol{\Delta y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \boldsymbol{\Delta y}_{t-p+1} + \boldsymbol{\varepsilon}_t,$$

This is the error correction element of the VECM specification

where $\boldsymbol{\Pi} = \boldsymbol{A}_1 + \boldsymbol{A}_2 + \cdots + \boldsymbol{A}_p - \boldsymbol{I}_{m \times m}$ is the long-run impact matrix

$0 \leq \text{rank}(\boldsymbol{\Pi}) < m$ defines r, the number of CI vectors

$\boldsymbol{\Gamma}_k = -\sum_{j=k+1}^{p} \boldsymbol{A}_j \quad k = 1, \ldots, p-1$: short-run impact matrices

Note that VAR(p) is transformed into a "VECM($p-1$)"

VECM: Construction

VAR(1) model for a 2-dimensional vector $\boldsymbol{y} = (y_1, y_2)'$; $\quad y_1, y_2 \sim CI(1,1)$:

$$\boldsymbol{y}_t = \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{u}_t, \qquad t = 1, 2, \ldots, T,$$

For a bivariate VAR, only one cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$ can exist /normalized/. The VECM model is defined as follows:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\Pi}$ is a 2×2 matrix with rank: $r = 1$.

$\boldsymbol{\Pi}$ may be decomposed as follows: $\boldsymbol{\Pi} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (1, -\beta_2) = \begin{bmatrix} \alpha_1 & -\alpha_1\beta_2 \\ \alpha_2 & -\alpha_2\beta_2 \end{bmatrix}$

To understand the decomposition, we may re-write the VECM:

$$\Delta y_{1t} = \alpha_1 (y_{1,t-1} - \beta_2 y_{2,t-1}) + \varepsilon_{1t}$$

1^{st} equation relates changes in y_{1t} to disequilibrium error $(y_{1,t-1} - \beta_2 y_{2,t-1})$

$$\Delta y_{2t} = \alpha_2 (y_{1,t-1} - \beta_2 y_{2,t-1}) + \varepsilon_{2t}$$

2^{nd} equation relates changes in y_{2t} to disequilibrium error $(y_{1,t-1} - \beta_2 y_{2,t-1})$

VECM: Johansen's methodology

1. Specify & estimate the m-dimensional VAR(p) model for $\boldsymbol{y}_t$.
2. Construct (Likelihood ratio) tests to determine the rank of $\boldsymbol{\Pi}$, i.e. to determine the number of cointegrating vectors.
3. Impose normalization and identifying restrictions on the cointegrating vectors (if necessary).
4. Given normalized CI vectors, estimate the resulting cointegrated VECM using maximum likelihood.

✓ IRFs and forecasts may be generated after a VECM is estimated ...

VECM: Johansen's methodology

$$\boldsymbol{\Delta y}_t = \boldsymbol{\Theta D}_t + \boldsymbol{\Pi y}_{t-1} + \boldsymbol{\Gamma}_1 \boldsymbol{\Delta y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \boldsymbol{\Delta y}_{t-p+1} + \boldsymbol{\varepsilon}_t,$$

If $\boldsymbol{y}_t \sim I(1) \quad \Rightarrow \quad \boldsymbol{\Pi}$ is a singular matrix: $0 \leq \text{rank}(\boldsymbol{\Pi}) < m$

1. rank $(\boldsymbol{\Pi}) = 0 \quad \Rightarrow \quad \boldsymbol{\Pi} = \mathbf{0}$
   $\boldsymbol{y}_t$ is not cointegrated and the VECM is a VAR on 1^{st} diffs.

2. $0 < \text{rank}(\boldsymbol{\Pi}) = r < m$
   $\boldsymbol{y}_t$ is cointegrated with r linearly independent vectors
   and $(m - r)$ common stochastic trends.

3. $\ldots \quad \text{rank}(\boldsymbol{\Pi}) = m \quad \Longleftrightarrow \quad \boldsymbol{y}_t \sim I(0)$
   full rank of $\boldsymbol{\Pi}$ means that $\boldsymbol{y}_t$ is in fact stationary.

VECM: Johansen's methodology

Johansen's Trace Statistic & test

- Based on the estimated eigenvalues of the matrix $\mathbf{\Pi}$:
  $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_m$ where $0 \leq \lambda_j < 1$
  (and r is the # of nonzero eigenvalues)
- $H_0(r) \ : \ r = r_0$
  $H_1(r_0) : \ r > r_0$
- $LR_{tr}(r_0) = \ -T \sum_{i=r_0+1}^{m} \log(1 - \hat{\lambda}_i)$
- Under H_0, eigenvectors $\hat{\lambda}_{r_0+1}, \ldots, \hat{\lambda}_m$ should be close to zero (as well as the LR_{tr} statistic).
- Under H_0, LR_{tr} follows a multivariate Dickey-Fuller distribution.

VECM: Johansen's methodology

Testing sequence for Johansen's Trace-Statistic test

1	$H_0: \quad r = 0$	vs.	$H_1: \quad 0 < r \leq m$
2	$H_0: \quad r = 1$	vs.	$H_1: \quad 1 < r \leq m$
3	$H_0: \quad r = 2$	vs.	$H_1: \quad 2 < r \leq m$
	$\ldots$	$\ldots$	$\ldots$
m	$H_0: \quad r = m - 1$	vs.	$H_1: \quad r = m$

We keep increasing r until we no longer reject the null.

VECM: Johansen's methodology

Johansen's Maximum Eigenvalue statistic & test

- Based on the estimated eigenvalues of the matrix $\mathbf{\Pi}$:
  $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_m$ where $0 \leq \lambda_j < 1$
- $H_0(r_0): \; r = r_0$
  $H_1(r_0): \; r = r_0 + 1$
- $LR_{max}(r_0) = -T \log(1 - \hat{\lambda}_{r_0+1})$
- Under H_0, LR_{max} follows a complex multivariate distribution (critical values implemented in SW).
- Testing sequence is analogous

VAR & VECM – Other extensions

Advanced analysis of univariate and multivariate time series (with examples in **R**)

http://faculty.washington.edu/ezivot/econ589/manual.pdf