# Materials for the course
# Financial Econometrics II

Michał Rubaszek:

Financial Markets Modelling Unit

Econometrics Institute

2024

SGH Warsaw School of Economics

# Contents

## Block 1

Forecasting and simulations
with
ARMA, VAR and SVAR models

***TOPICS***

1. Introduction to R
2. ARMA models
3. VAR models
4. Forecasting contests

# Topic 1

# Introduction

This script includes materials for the course *Financial Econometrics II*.

This document is accompanied by materials available at `http://web.sgh.waw.pl/~mrubas/`:

- Scripts in R package
- Data

Additional materials:


1. Alexander C., 2009. Market Risk Analysis, Wiley
2. Bauwens L., Laurent S., Rombouts J., 2006. Multivariate GARCH models: a survey, JAE 21, 79-109
3. Cont R., 2001. Empirical properties of asset returns, Quantitative Finance 1, 223-236
4. Danielsson J., 2011. Financial Risk Forecasting, Wiley
5. Lutkepohl H., Krätzig M. (2004). Applied Time Series Econometrics, Cambridge University Press
6. Luetkepohl H., 2011. Vector Autoregressive Models, Economics WP ECO2011/30, EUI
7. Nelsen R., 2006. An Introduction to Copulas, Springer
8. Rubaszek M., (2012). Modelowanie Polskiej Gospodarki z Pakietem R, Oficyna Wydawnicza SGH
9. Tsay R. (2002). Analysis of Financial Time Series, Wiley

## Content of R codes

1. Operations on vectors and matrices
2. Conditioning, loops, defining functions
3. Importing data (`read.csv`, `quantmod`, `OECD`, `restatapi`)
4. Converting and plotting data (`ts`, `zoo`, `xts`)
5. Simple vs. compound interest rate

## Rates of return / growth rates

**Simple rate of return:**

$$Y_t = (1 + R_t)Y_{t-1} \leftrightarrow R_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

**Compound interest rate** ($m$ is compounding frequency)**:**

$$Y_t = \left(1 + \frac{R_{m,t}}{m}\right)^m Y_{t-1}$$

**Continuously compound interest rate:**

$$Y_t = \lim_{m\to\infty} \left(1 + \frac{R_{m,t}}{m}\right)^m Y_{t-1} = \exp(r_t)\, Y_{t-1}$$

**Logarithmic rate of return:**

$$Y_t = \exp(r_t)Y_{t-1} \leftrightarrow r_t = \ln(Y_t/Y_{t-1})$$

**Notice:** $1 + R = \exp(r) \leftrightarrow r = \ln(1 + R)$

## Rates of return / growth rates

**Simple returns:**

- Easy to calculate for a portfolio of assets: $R_p = \sum_{k=1}^{K} w_k R_k$
- Easy to communicate to non-statisticians
- Not symmetric nor additive...

**Log returns:**

- Symmetric and additive
- Easy to communicate to statisticians
- Difficult to calculate for a portfolio of assets: $r_p \neq \sum_{k=1}^{K} w_k r_k$

**We will work with log returns**

## Exercises

**Exercise 1.1.**

Write an algorithm, which would allow to calculate the roots of the equation:

$$\boldsymbol{e^x - (x+1)^2 = 0}$$

knowing that they are in the interval $< -3,3 >$.

[Hint: make two loops with functions `for` and `while`]

**Exercise 1.2.**

Create a function `invVal(Y,h,R,m)`that will calculate the value of investment $Y$ after $h$ years, given that the annual interest rate is $R$ and compound frequency $m$.

Use the function to calculate the value of 1000PLN after 1 year for $m = \{1,2,4,\infty\}$ and $R_m = 10\%$.

## Exercises

**Exercise 1.3.**

Using the `reststapi` or `OECD` package import to R the annual growth rate of real GDP in Poland (at quarterly frequency). Write a series as a `zoo` object and make a plot. What was the average growth rate over the last 10 years

**Exercise 1.4.**

Import daily data for the WIG index from the Internet to R. After converting the series to a `zoo` object, make a panel of figures for

- historic prices
- logarithmic growth rates
- ACF for levels
- ACF for growth rates

# Topic 2

# ARMA models

- Calculating impulse-response functions
- Testing for unit root
- Estimating ARMA model
- Information criteria
- Testing ARMA specification
- Forecasting with ARMA model

## IRF – impulse response function

**Impulse response function – IRF:**
describes how variable $y_t$ reacts over time to exogenous impulse $\epsilon_t$.

**Moving Average model:**

$$y_t = \mu + \theta_0\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots$$

**Formula for IRF:**

$$IRF_k = \theta_k = \frac{\partial y_t}{\partial \epsilon_{t-k}} = \frac{\partial y_{t+k}}{\partial \epsilon_t}$$

**How to calculate IRF for a model?**
Transform model to a moving average (MA) form

## IRF – example

**Moving Average representation:** $y_t = \mu + \theta_0\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots$

**Formula for IRF:** $IRF_k = \theta_k = \frac{\partial y_t}{\partial \epsilon_{t-k}} = \frac{\partial y_{t+k}}{\partial \epsilon_t}$

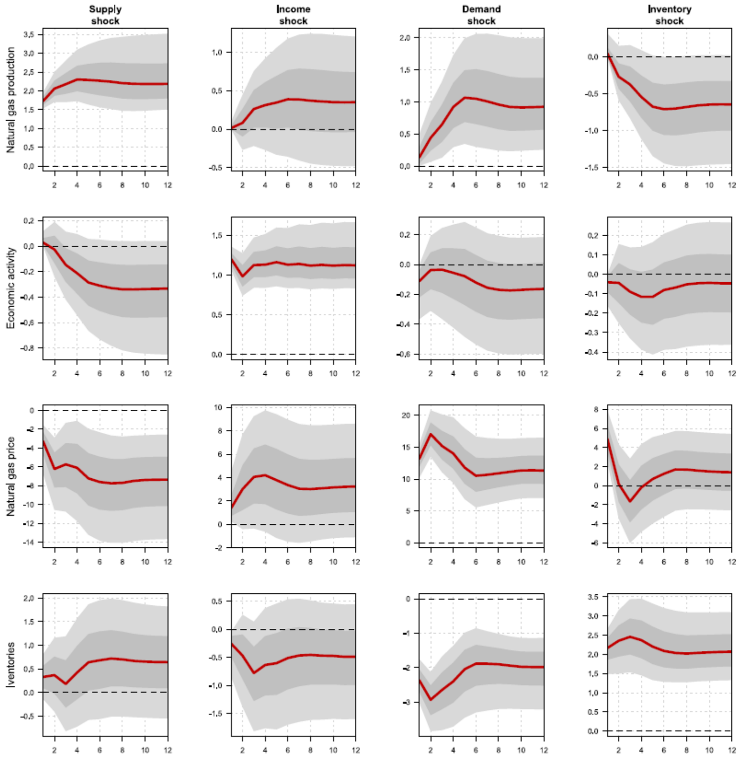
**Example:** A model for GDP growth rate at home and abroad

$$\begin{bmatrix} y_t^* \\ y_t \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.50 \end{bmatrix} + \begin{bmatrix} 0.50 & 0.00 \\ 0.25 & 0.25 \end{bmatrix} \begin{bmatrix} \eta_t^* \\ \eta_t \end{bmatrix} + \begin{bmatrix} 0.25 & 0.00 \\ 0.10 & 0.20 \end{bmatrix} \begin{bmatrix} \eta_{t-1}^* \\ \eta_{t-1} \end{bmatrix} + \begin{bmatrix} 0.125 & 0.00 \\ 0.005 & 0.10 \end{bmatrix} \begin{bmatrix} \eta_{t-2}^* \\ \eta_{t-2} \end{bmatrix} + \cdots$$

what is the interpretation of $\mu$ and $\theta_k$?

12

# Plot of IRF




Source: Rubaszek, Szafranek & Uddin (2020)

13

# Calculating IRF: AR(1) model

**AR model:** $y_t = \rho y_{t-1} + \epsilon_t$

**MA representation:** $y_t = \epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + \cdots$
Hence: $\theta_k = \rho^k$ and $\mu = 0$

**Lag operator:** $Ly_t = y_{t-1}$ and $L^k y_t = y_{t-k}$

**AR model writen with lag operator:** $(1 - \rho L)y_t = \epsilon_t$

$$y_t = (1-\rho L)^{-1}\epsilon_t = \sum_{k=0}^{\infty} \rho^k \epsilon_{t-k} \, [+\lim_{s\to\infty} \rho^s E(y_{t-s})]$$

**Exercise:** Calculate MA representation for $y_t = 0.8y_{t-1} + \epsilon_t$

## Calculating IRF: AR(1) model

**AR model with a constant:**

$y_t = \alpha + \rho y_{t-1} + \epsilon_t$

Substitute:

$z_t = y_t - \mu$, where $\mu = \frac{\alpha}{1-\rho}$

and think in terms of:

$z_t = \rho z_{t-1} + \epsilon_t$

**Exercise:** Calculate MA representation for $y_t = 2 + 0.5y_{t-1} + \epsilon_t$

15

## Calculating IRF: AR(2) model

**AR(2) model:**

$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \epsilon_t$

$(1 - \lambda_1 L)(1 - \lambda_2 L)y_t = \epsilon_t$

**Hence AR(2) as a multiplication of two AR(1) processes**

$y_t = (1 - \lambda_1 L)^{-1} z_t \rightarrow y_t = \lambda_1 y_{t-1} + z_t$

$z_t = (1 - \lambda_2 L)^{-1} \epsilon_t \rightarrow z_t = \lambda_2 z_{t-1} + \epsilon_t$

**Exercise:** Find roots of characteristic equation for $y_t = 1.3y_{t-1} - 0.4y_{t-2} + \epsilon_t$

16

## Calculating IRF: AR(P) model

**AR(P) model:** $y_t = \sum_{p=1}^{P} \rho_p y_{t-p} + \epsilon_t$

$\rho(L)y_t = (1 - \rho_1 L - \cdots - \rho_P L^P)y_t = \Pi_{p=1}^{P}(1 - \lambda_p L)y_t$

$\rho(L)y_t = \epsilon_t$

$y_t = \Pi_{p=1}^{P}(1 - \lambda_p L)^{-1}\epsilon_t$

AR(P) as a multiplication of $P$ AR(1) processes

## Exercises

**Exercise 2.1.**
**Write the IRF $y_{t+k}$ with respect to $\epsilon_t$ for the following processes:**

A. $y_t = 0.8y_{t-1} + \epsilon_t$

B. $y_t = 2 + 0.5y_{t-1} + \epsilon_t$

C. $y_t = 1.3y_{t-1} - 0.4y_{t-2} + \epsilon_t$

D. $y_t = 0.8y_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}$

E. $y_t = y_{t-1} + \epsilon_t$

## Unit root

**For AR model:** $y_t = \rho y_{t-1} + \epsilon_t$

$y_t = \sum_{k=0}^{\infty} \rho^k \epsilon_{t-k} \; [+ \lim_{s \to \infty} \rho^s E(y_{t-s})]$

If $\rho = 1$ then

- $\lim_{s \to \infty} \rho^s E(y_{t-s}) \neq 0$
- $\lim_{s \to \infty} \rho^s = 1$

This means that the impact of a shock is not decaying and that the process is not returning to an equilibrium value. It is non-stationary

For $\rho < 1$ we might calculate so-called half-life:

$HL = \frac{\ln 0.5}{\ln \rho} \leftrightarrow \rho^{HL} = 0.5$

## Unit root: stationarity

**Definition of weak stationarity:**

A process is said to be covariance-stationary, or weakly stationary, if its first and second (unconditional) moments are time invariant, and for each period $t$ are equal to:

$$\begin{aligned} E(Y_t) &= \mu \\ Var(Y_t) &= \gamma_0 = \sigma^2 \\ Cov(Y_t, Y_{t-s}) &= \gamma_s \end{aligned}$$

**Important:**

- unconditional and conditional moments might differ
  $E(Y_t) \neq E(Y_t | Y_{t-1} = y_{t-1})$
- Stationarity can be interpreted as „mean reversion” ,
  i.e. that a series should fluctuate around $\mu$ and its volatility around $\sigma$

## Unit root: stationarity

**Definition of strong stationarity:**

The joint distribution of $Y_{t_1}, Y_{t_2}, \ldots, Y_{t_s}$ is the same as the joint distribution of $Y_{t_1+k}, Y_{t_2+k}, \ldots, Y_{t_s+k}$:

$$f\left(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_s}\right) = f(Y_{t_1+k}, Y_{t_2+k}, \ldots, Y_{t_s+k})$$

- Is not limited to the first two moments
- Implies weak stationarity
- Not particularly useful in practical applications as it cannot be tested...

## Unit root: tests

**Augumented Dickey-Fuller test:**

$$\Delta y_t = [\alpha_0 + \alpha_1] + \delta y_{t-1} + \sum_{p=1}^{P} \gamma_p \Delta y_{t-p} + \epsilon_t$$

$H0: \delta = 0$, i.e. non-stationarity

$H1: \delta < 0$, i.e. stationarity

$$DF_{cal} = \frac{\hat{\delta}}{S_{\hat{\delta}}} \sim DF$$

**Note:** Adding lags of $\Delta y$ is a parametric correction for possible autocorrelation of the error term $\epsilon_t$

## Unit root: tests

**Phillips-Perron test:**

$$y_t = [\alpha_0 + \alpha_1] + \rho y_{t-1} + \epsilon_t$$

$H0: \rho = 1$, i.e. non-stationarity

$H1: \rho < 1$, i.e. stationarity

$$PP_{cal} = \left(\frac{\hat{\gamma}_0}{\hat{\gamma}_\infty}\right)^{0.5} \frac{(\hat{\rho} - 1)}{S_\rho} - \frac{T}{2}(\hat{\gamma}_\infty - \hat{\gamma}_0)\left(\frac{S_\rho}{\sqrt{\hat{\gamma}_\infty \hat{\gamma}_0}}\right) \sim PP$$

where $\hat{\gamma}_0$ and $\hat{\gamma}_\infty$ are variance and long-term variance for residuals $\hat{\epsilon}_t$.

**Note:** If $\hat{\gamma}_0 = \hat{\gamma}_\infty$ then $PP_{cal} = DF_{cal}$. In other case we have non-parametric correction for possible autocorrelation of the error term $\epsilon_t$

## Unit root: tests

**KPSS test:**

$y_t = x_t + z_t$

$x_t = x_{t-1} + v_t, v_t \sim WN(0, \sigma_v^2)$

$z_t = [\alpha_0 + \alpha_1] + \epsilon_t$

$H0: \sigma_v^2 = 0$, i.e. stationarity

$H1: \sigma_v^2 > 0$, i.e. non-stationarity

$$KPSS_{cal} = \frac{1}{T^2} \frac{\sum_{t=1}^{T} S_t^2}{\hat{\gamma}_\infty} \sim KPSS$$

where $\hat{\gamma}_\infty$ is the long-run variance of residuals $\hat{\epsilon}_t$ from regression of $y_t$ on a constant and a trend (depending on a specification) and $S_t = \sum_{s=1}^{t} \hat{\epsilon}_s$

## Unit root: tests

**Important:**

For persistent processes and small sample the power of ADF and PP tests is low, whereas the KPSS test is subject to size distortion

[illustration in the Monte Carlo example in the R file]

**Implication:**

- Be careful while differentiating the data
- Economic knowledge might be better advice that the tests

## Exercises

**Exercise 2.2.**

Import data for the US economy over the years 1860-1970 with commands:

```
> require(urca)
> data(nporg)
```

For each series decide whether to use logs or not. Test for stationarity.

What are the economic reasons of non-stationarity?

**Exercise 2.3.**

Import data for HICP YoY inflation for a selected EU country with the `eurostat` package. Decide on the level of integration of the downloaded variable using ADF, PP and KPSS tests.

## ARMA model introduction

**Specification of ARMA(P,Q) model:**

$$y_t = [\alpha_0 + \alpha_1 t] + \sum_{p=1}^{P} \rho_p y_{t-p} + \sum_{q=0}^{Q} \gamma_q \epsilon_{t-q}$$

$$\rho(L) y_t = [\alpha_0 + \alpha_1 t] + \gamma(L)\epsilon_t$$

**Equilibrium value for stationary ARMA model:**

$$E(y_t) = \mu = \frac{\alpha_0}{1 - \sum_{p=1}^{P} \rho_p}$$

**Specification of ARIMA(P,D,Q) model:**

$$\rho(L)(1-L)^D y_t = [\alpha_0 + \alpha_1 t] + \gamma(L)\epsilon_t$$

## ARMA model

**Why do we need ARMA models?**

- For analysing the properties of univariate time series
- Seasonal adjustment
- Forecasting (see: Nelson Ch.R, 1972. The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy, American Economic Review 62(5), 902-17 - link)

## AR(1) model - estimation

**Let us consider AR(1):**

$y_t = \alpha + \rho y_{t-1} + \epsilon_t,\ \ \epsilon_t \sim N(0, \sigma_\epsilon^2)$

**Conditional likelihood of a single observation:**

$$p(y_t|\alpha, \rho, y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{(y_t - \alpha - \rho y_{t-1})^2}{2\sigma_\epsilon^2}\right)$$

**Likelihood of all observations:**

$$p(y_1, y_2, \dots, y_T|\alpha, \rho) = p(y_1|\alpha, \rho) \times p(y_2|\alpha, \rho, y_1) \times \cdots \times p(y_T|\alpha, \rho, y_{T-1})$$

where $p(y_1|\alpha, \rho) = \frac{1}{\sqrt{\frac{2\pi\sigma_\epsilon^2}{1-\rho^2}}} \exp\left(-\frac{\left(y_t - \frac{\alpha}{1-\rho}\right)^2}{\frac{2\sigma_\epsilon^2}{1-\rho^2}}\right)$

- If we neglect $p(y_1|\alpha, \rho)$ : conditional ML estimator (=LS estimator)
- If we include $p(y_1|\alpha, \rho)$ : full ML estimator
  (the derivation of ML function for ARMA(P,Q) with the Kalman filter - link)

## ARMA model - specification

**Information criteria:**

Akaike: $AIC = -2\frac{\ell}{T} + 2\frac{K}{T}$

Schwarz: $BIC = -2\frac{\ell}{T} + 2\frac{K}{T}\ln(T)$

Hannan-Quinn: $HQIC = -2\frac{\ell}{T} + 2\frac{K}{T}\ln(\ln T)$

where $K$ is the number of estimated parameters and $\ell = \ln \mathcal{L}$ is the log-likelihood.
We choose the model with the lowest IC.

Notice:

- $K(SIC) \leq K(HQIC) \leq K(AIC)$
- IC depends on the fit (log-likelihood) and penalty on the number of params.

## ARMA model - specification

**Likelihood ratio test:**

H0: the fit of big ARMA ($m$ params more) is the same as fit of small ARMA

$LR = -2(\ell_{\mathrm{r}} - \ell_u) \sim \chi^2(m)$

where $m$ is te number of additional parms. and $\ell$ is the log-likelihood for restricted (small) a nd unrestricted (big) models.

**Autocorelation (portmanteau) Ljung-Box test:**

$H0{:}\, \rho_{\epsilon,j} = 0$ for $j = 1,2,\dots,J$

$LB = T^2 \sum_{j=1}^{J} \frac{1}{T-j} \hat{\rho}_{e,j}^2 \sim \chi^2\big(J - (P+Q)\big)$

## ARMA model - forecasting

**ARMA(P,Q):** $y_t = [\alpha_0 + \alpha_1 t] + \sum_{p=1}^{P} \rho_p y_{t-p} + \sum_{q=0}^{Q} \gamma_q \epsilon_{t-q}$

**MA representation:** $y_t = \theta_0 \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots$

**Point forecast (calculated recursively)**:

$y_{T+h}^f = [\alpha_0 + \alpha_1 t] + \sum_{p=1}^{P} \rho_p y_{T+h-p}^f + \sum_{q=0}^{Q} \gamma_q \epsilon_{T+h-q}^f$

where $\epsilon_{T+h}^f = 0$ for $h > 0$

**Forecast error:**

$y_{T+H} - y_{T+H}^f = \theta_0 \epsilon_{T+H} + \theta_1 \epsilon_{T+H-1} + \cdots + \theta_{H-1} \epsilon_{T+1} = \sum_{h=0}^{H-1} \theta_h \epsilon_{T+H-h}$

**Forecast variance (only due to stochastic term):**

$Var(y_{T+H}) = \theta_0^2 + \theta_1^2 + \cdots + \theta_{H-1}^2 = \sum_{h=0}^{H-1} \theta_h^2$

# Exercises

**Exercise 2.4.**

For the ARMA(2,0) model:

$y_t = 1{,}1y_{t-1} - 0{,}3y_{t-2} + \epsilon_t, \epsilon_t \sim N(0{,}1)$

- check if the model is stationary
- Knowing $y_T = 1$ and $y_{T-1} = 2$ calculate the forecast for periods $T + 1$ and $T + 2$.
- Write the model in $MA(\infty)$ form. Calculate the values for the first three coefs.
- Calculate point and 95% interval forecast for $T + 2$

**Exercise 2.5.**

For the series of your choice:

- Choose the specification of the ARMA model with the Schwarz criterion
- Convert the model to $MA(\infty)$ form
- Verify the model for autocorrelation
- Calculate point and density forecasts

# Topic 3

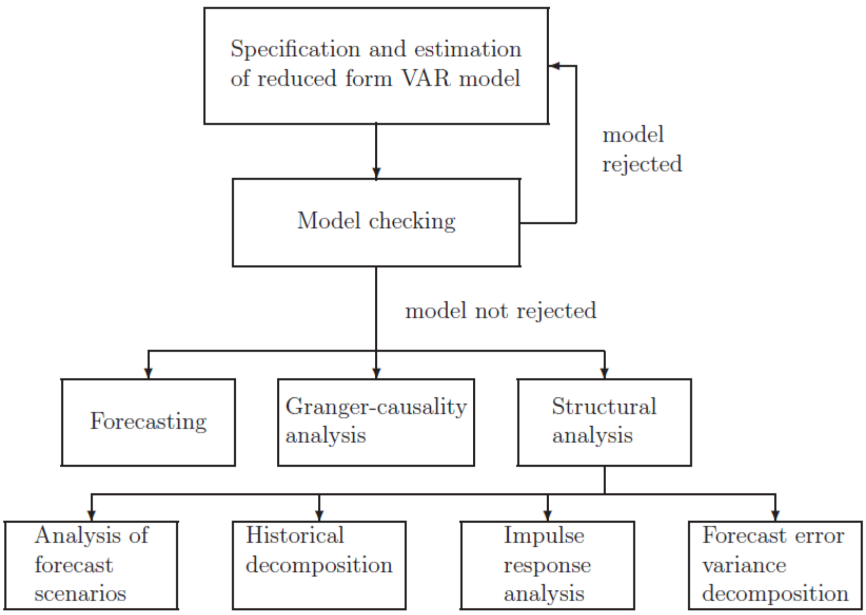
# VAR models

- Estimating VAR model
- Structural of VAR
- Impulse-response function (IRF)
- Forecast error variance decomposition
- Historical decomposition
- Forecasting with VAR model

# Vector autoregression model - VAR

- VAR is an extension of univariate autoregression model to multivariate time series data, in which all variables are treated as endogenous (Sims critique)
- VAR is a useful model for analysing the dynamic behaviour of economic and financial time series. It is also a standard tool used for forecasting
- Structural VAR as an important „story-telling" model

# VAR model




Source: Luetkepohl (2011)

37

## Specification of a VAR model

**VAR(p) model:**

$$y_t = C + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \epsilon_t \, , \epsilon_t \sim N(0, \Sigma)$$

where:

$y_t \; = [y_{1t} \;\; y_{2t} \;\; \dots \;\; y_{nt}]'$ vector of $n$ endogenous variables

$\epsilon_t \; = [\epsilon_{1t} \;\; \epsilon_{2t} \;\; \dots \;\; \epsilon_{nt}]'$ vector of error terms

$C \; = [c_1 \;\; c_2 \;\; \dots \;\; c_n]'$ vector of constants

$A_p = \left[a_{ij,p}\right]_{n \times n}$ matrix of parameters

$\Sigma \; = \left[\sigma_{ij}\right]_{n \times n}$ covariance matrix

## IRF: VAR(1) case

**VAR model:**

$$y_t = A y_{t-1} + \epsilon_t$$

$$y_t = \epsilon_t + A\epsilon_{t-1} + A^2 \epsilon_{t-2} + A^3 \epsilon_{t-3} + \cdots$$

**VMA representation:**

$$y_t = \Psi_0 \epsilon_t + \Psi_1 \epsilon_{t-1} + \Psi_2 \epsilon_{t-2} + \Psi_3 \epsilon_{t-3} + \cdots$$

$\Psi_k = \left[\psi_{ij,k}\right]_{n \times n} = \frac{\partial y_t}{\partial \epsilon_{t-k}} = \frac{\partial y_{t+k}}{\partial \epsilon_t}$: $\quad n \times n$ matrix of IRFs

$$\Psi_k = A^k$$

## Stationarity: VAR(1) case

$y_t = \Psi_0 \epsilon_t + \Psi_1 \epsilon_{t-1} + \Psi_2 \epsilon_{t-2} + \Psi_3 \epsilon_{t-3} + \cdots$

VAR is stationary if: $\lim_{k\to\infty} \Psi_k = \lim_{k\to\infty} A^k = 0$

**Spectral decomposition:** $A = V\Lambda V^{-1}$

$\Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_n)$ eigenvalues

$V = [v_1 \; v_2 \; \dots \; v_n]$ eigenvectors

$y_t = V\Lambda V^{-1} y_{t-1} + \epsilon_t \quad \leftrightarrow \quad \tilde{y}_t = \Lambda \tilde{y}_{t-1} + \tilde{\epsilon}_t$

where $\tilde{y}_t = V^{-1} y_t$ and $\tilde{\epsilon}_t = V^{-1}\epsilon_t$

## Stationarity: VAR(1) case

$\tilde{y}_t = \Lambda \tilde{y}_{t-1} + \tilde{\epsilon}_t$

$$\begin{bmatrix} \tilde{y}_{1t} \\ \tilde{y}_{2t} \\ \dots \\ \tilde{y}_{nt} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} \tilde{y}_{1t-1} \\ \tilde{y}_{2t-1} \\ \dots \\ \tilde{y}_{nt-1} \end{bmatrix} + \begin{bmatrix} \tilde{\epsilon}_{1t} \\ \tilde{\epsilon}_{2t} \\ \dots \\ \tilde{\epsilon}_{3t} \end{bmatrix}$$

We have $n$ univariate AR(1) processes, hence $\tilde{y}_t$ is stationary of all characteristic roots are lower than unity $|\lambda_i| < 1$.

$y_t = V\tilde{y}_t$: our observables as a linear combination of $n$ AR(1) processes

$\lim_{k\to\infty} \Psi_k = \lim_{k\to\infty} A^k = \lim_{k\to\infty} V\Lambda^k V^{-1} = 0$ only if $|\lambda_i| < 1$ for $i$=1,2, … , $n$

## Stationarity: VAR(P) case

**VAR(P) model:**
$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_P y_{t-P} + \epsilon_t$

**Canonical form of VAR(p):**
$y_t^* = A^* y_{t-1}^* + \epsilon_t^*$

$$A^* = \begin{bmatrix} A_1 & A_2 & \dots & A_p \\ I & 0 & \dots & \dots \\ 0 & I & \dots & 0 \\ 0 & 0 & \dots & 0 \end{bmatrix}, y_t^* = \begin{bmatrix} y_t \\ y_{t-1} \\ \dots \\ y_{t-p+1} \end{bmatrix}, \epsilon_t^* = \begin{bmatrix} \epsilon_t \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

- Model is stationary if the roots of characteristic polynomial for $A^*$ are $|\lambda_i^*| < 1$
- VMA form can be calculated for canonical VAR

$y_t^* = \Psi_0^* \epsilon_t^* + \Psi_1^* \epsilon_{t-1}^* + \Psi_2^* \epsilon_{t-2}^* + \Psi_3^* \epsilon_{t-3}^* + \cdots$

→ $\Psi_k$ is the $n \times n$ upper part of $\Psi_k^*$

## Estimating the VAR model

$y_t = C + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \epsilon_t , \epsilon_t \sim N(0, \Sigma)$
$y_t' = C' + y_{t-1}' A_1' + y_{t-2}' A_2' + \cdots + y_{t-p}' A_p' + \epsilon_t'$
$y_t' = x_t' B + \epsilon_t'$

$$x_t = \begin{bmatrix} 1 \\ y_{t-1} \\ y_{t-2} \\ \dots \\ y_{t-p} \end{bmatrix}, \qquad B = \begin{bmatrix} C' \\ A_1' \\ A_2' \\ \dots \\ A_p' \end{bmatrix}$$

$Y = XB + \mathcal{E}$

$$Y = \begin{bmatrix} y_1' \\ y_2' \\ y_3' \\ \dots \\ y_T' \end{bmatrix}, \quad X = \begin{bmatrix} x_1' \\ x_2' \\ x_3' \\ \dots \\ x_T' \end{bmatrix}, \; \mathcal{E} = \begin{bmatrix} \epsilon_1' \\ \epsilon_2' \\ \epsilon_3' \\ \dots \\ \epsilon_T' \end{bmatrix}$$

# Estimating the VAR model

**VAR(p) in matrix notation:**
$Y = XB + \mathcal{E}$

**LS estimates**
$\hat{B} = (X'X)^{-1}XY$

**Residuals**
$\hat{\mathcal{E}} = Y - X\hat{B}$

**Estimate of the covariance matrix**
$\hat{\Sigma} = (T-k)^{-1}(\hat{\mathcal{E}}'\hat{\mathcal{E}})$
Where $k = 1 + np$ is the number of parameters in each equations

More details: see p. 16-18 of Dieppe et al. (2016) - link

# VAR model - specification

**Information criteria:**

Akaike: $AIC = -2\frac{\ell}{T} + 2\frac{K}{T}$

Schwarz: $BIC = -2\frac{\ell}{T} + 2\frac{K}{T}\ln(T)$

Hannan-Quinn: $HQIC = -2\frac{\ell}{T} + 2\frac{K}{T}\ln(\ln T)$

where $K = n(1+np)$ is the number of parameters and $\ell = \ln \mathcal{L}$ is the log-likelihood.

**Ljung-Box (adjusted portmonteau) autocorrelation test:**

$LB_{cal} = T^2 \sum_{j=1}^{J} \frac{1}{T-j} tr(\hat{\Gamma}_j' \hat{\Gamma}_0^{-1} \hat{\Gamma}_j \hat{\Gamma}_0^{-1}) \sim \chi^2(n^2(J-p))$

where $\hat{\Gamma}_j = \frac{1}{T}\sum \epsilon_t \epsilon_{t-j}'$.

## SVAR: structural VAR

**VAR(P) model, in which shocks have no economic interpretation:**

$y_t = C + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \epsilon_t$ , $\epsilon_t \sim N(0, \Sigma)$

**SVAR(p) model, in which shocks have interpretation**

$y_t = C + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + D\eta_t$ , $\eta_t \sim N(0, I)$

[equivalent notation, by multiplying both sides by $D^{-1}$]

$D_0 y_t = F + D_1 y_{t-1} + D_2 y_{t-2} + \cdots + D_p y_{t-p} + \eta_t$ , $\eta_t \sim N(0, I)$

**SVMA representation**

$y_t = \mu + \psi_0 \eta_t + \psi_1 \eta_{t-1} + \psi_2 \eta_{t-2} + \psi_3 \eta_{t-3} \dots$

where $\psi_0 = D$

## SVAR: identification of shocks

**VAR(p) model:**

$y_t = C + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \epsilon_t$ , $\epsilon_t \sim N(0, \Sigma)$

**SVAR(p) model:**

$y_t = C + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + D\eta_t$ , $\eta_t \sim N(0, I)$

**SVMA representation**

$y_t = \mu + \psi_0 \eta_t + \psi_1 \eta_{t-1} + \psi_2 \eta_{t-2} + \psi_3 \eta_{t-3} \dots$

**How to find matrix *D*?**

We need to impose $\frac{n(n-1)}{2}$ restrictions, taking into account that $DD' = \Sigma$

**Short-term restrictions / Cholesky identification:**

We assume that $D$ is lower triangular matrix

**Long-run restrictions / Blanchard-Quah identification:**

We impose restrictions on the matrix of long-term response $\psi = \sum_{i=0}^{\infty} \psi_i = \left(I - A_1 - \cdots - A_p\right)^{-1} D$

## FEVD: forecast error variance decomposition

**SVMA representation**

$$y_t = \mu + \psi_0 \eta_t + \psi_1 \eta_{t-1} + \psi_2 \eta_{t-2} + \psi_3 \eta_{t-3} \dots$$

**Error of forecast for horizon $h$ due to future shocks**

$$y_{t+h} - y_{t+h|t} = \psi_0 \eta_{t+h} + \psi_1 \eta_{t+h-1} + \dots + \psi_{h-1} \eta_{t+1}$$

**Variance of forecast error for horizon $h$ due to future shocks**

$$Var_t(y_{t+h}) = \psi_0 \psi_0' + \psi_1 \psi_1' + \dots + \psi_{h-1} \psi_{h-1}'$$

More details: see p. 101-103 of Dieppe et al. (2016) - link

## FEVD: foreast error variance decomposition

**Variance of forecast error for horizon $h$ due to future shocks**

$$Var_T(y_{T+h}) = \psi_0 \psi_0' + \psi_1 \psi_1' + \dots + \psi_{h-1} \psi_{h-1}'$$

$$\psi_k = \begin{bmatrix} \psi_{s,11} & \cdots & \psi_{s,1n} \\ \vdots & \ddots & \vdots \\ \psi_{s,n1} & \cdots & \psi_{s,nn} \end{bmatrix}, \psi_s \psi_s' = \begin{bmatrix} \psi_{s,11}^2 + \psi_{s,12}^2 + \cdots + \psi_{s,1n}^2 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \psi_{s,n1}^2 + \psi_{s,n2}^2 + \cdots + \psi_{s,nn}^2 \end{bmatrix}$$

Substituting yields:

$$\sigma_i^2(h) = Var_T(y_{i,T+h}) = \sum_{s=0}^{h-1} (\psi_{s,i1}^2 + \psi_{s,i2}^2 + \cdots + \psi_{s,in}^2) = \sum_{j=1}^{n} (\psi_{0,ij}^2 + \psi_{1,ij}^2 + \cdots + \psi_{h-1,ij}^2)$$

**Contribution of shocks $\boldsymbol{\eta_{j,T+s}}$ to total forecast error variance:**

$$\sigma_{ij}^2(h) = \psi_{0,ij}^2 + \psi_{1,ij}^2 + \cdots + \psi_{h-1,ij}^2$$

**so that:**

$$\sigma_i^2(h) = \sum_{j=1}^{n} \sigma_{ij}^2(h)$$

## Historical decomposition

**SVMA representation**

$$y_t = \mu + \psi_0 \eta_t + \psi_1 \eta_{t-1} + \psi_2 \eta_{t-2} + \psi_3 \eta_{t-3} \dots$$

**For a single variable**

$$y_{it} = \mu_i + \sum_{j=1}^{n} (\psi_{0,ij} \eta_{j,t} + \psi_{1,ij} \eta_{j,t-1} + \psi_{2,ij} \eta_{j,t-2} + \psi_{3,ij} \eta_{j,t-3} \dots)$$

**Contribution of shocks $\eta_{j,t-s}$ the value of $y_{it}$:**

$$y_{it,j} = \psi_{0,ij} \eta_{j,t} + \psi_{1,ij} \eta_{j,t-1} + \psi_{2,ij} \eta_{j,t-2} + \psi_{3,ij} \eta_{j,t-3} \dots$$

More details: see p. 101-103 of Dieppe et al. (2016) - link

## Forecesting with VAR models

**VAR(p) model:**

$$y_t = C + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + \epsilon_t \, , \, \epsilon_t \sim N(0, \Sigma)$$

**Point forecast for horizon $h$:**

$$y_{t+h|t} = C + A_1 y_{t+h-1|t} + A_2 y_{t+h-2|t} + \dots + A_p y_{t+h-p|t}$$

**VMA representation:**

$$y_t = \Psi_0 \epsilon_t + \Psi_1 \epsilon_{t-1} + \Psi_2 \epsilon_{t-2} + \Psi_3 \epsilon_{t-3} + \dots$$

**Error of forecast for horizon $h$ due to future shocks**

$$y_{t+h} - y_{t+h|t} = \Psi_0 \epsilon_{t+h} + \Psi_1 \epsilon_{t+h-1} + \dots + \Psi_{h-1} \epsilon_{t+1}$$

**Variance of forecast error for horizon $h$ due to future shocks**

$$Var_t(y_{t+h}) = \Psi_0 \Sigma \Psi_0' + \Psi_1 \Sigma \Psi_1' + \dots + \Psi_{h-1} \Sigma \Psi_{h-1}'$$

# Exercises

**Exercise 3.1.**

For the model ($y$ and $y^*$ denote output at home and abroad):

$$\begin{bmatrix} y_t^* \\ y_t \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.50 \end{bmatrix} + \begin{bmatrix} 0.75 & 0.00 \\ 0.25 & 0.50 \end{bmatrix} \begin{bmatrix} y_{t-1}^* \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1.00 & 0.00 \\ 0.50 & 1.00 \end{bmatrix} \begin{bmatrix} \eta_t^* \\ \eta_t \end{bmatrix}, \eta_t \sim \mathrm{N}(0, I)$$

- Check if the model is stationary
- Calculate the equiibrium value of $y_t$ and $y_t^*$
- Knowing $y_T^* = 1$ and $y_T = 2$ calculate the forecast for periods $T + 1$ and $T + 2$.
- Write the model in $VMA(\infty)$ form. Calculate the values for the first two lags.
- Calculate FEVD for $y_{T+1}$ and $y_{T+2}$
- Calculate point and 95% interval forecast for $y_{T+1}$

**Exercise 3.2.**

For a model VAR(2)

$$\begin{bmatrix} y_t^* \\ y_t \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.50 \end{bmatrix} + \begin{bmatrix} 0.50 & 0.00 \\ 0.25 & 0.25 \end{bmatrix} \begin{bmatrix} y_{t-1}^* \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 0.25 & 0.00 \\ 0.00 & 0.25 \end{bmatrix} \begin{bmatrix} y_{t-1}^* \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1.00 & 0.00 \\ 0.50 & 1.00 \end{bmatrix} \begin{bmatrix} \eta_t^* \\ \eta_t \end{bmatrix}, \eta_t \sim \mathrm{N}(0, I)$$

- build a companion matrix
- calculate [in R] if it is stationary
- compute [in R] VMA representation for the first four lags

# Exercises

**Exercise 3.3.**

Select a variable for a domestic ($y_t$) and foreign ($y_t^*$) economy
(inflation, GDP growth rate, unemployment rate or gov. bond 10Y yield) and:

- Estimate VAR model for $Y_t = [y_t^* \; y_t]'$ (select lags, check for autocorrelation)
- Identify monetary policy shock (Cholesky decomposition)
- Plot IRF
- Plot FEVD
- Calculate FEVD from IRF?
- Calculate historical decomposition
- Make a forecast for the next two years

**Exercise 3.4.**

Select an EU country and download data for changes in output $\Delta y_t$ and the unemployment rate $u_t$:

- Estimate VAR model for $Y_t = [\Delta y_t \; u_t]'$, make BQ long-term structuralization
- Plot IRF, FEVD, historical decomposition
- Calculate output gap and compare it to data from AMECO database (link)

# Topic 4

# Out-of-sample forecast evaluation

- Ex-ante and ex-post forecast
- Point and density forecasts
- Forecasting competition schemes
- Bias-variance trade-off
- Efficiency of forecasts
- Sequential forecasts
- Ex-post forecasts accuracy measures (ME, MAE, RMSE)
- Diebold-Mariano test
- Machine Learning methods (NN, RF)

## About forecasting

***The ultimate goal of a positive science is to develop a theory or hypothesis that yields valid and meaningful predictions about phenomena not yet observed. Theory is judged by its predictive power.***

***A hypothesis can't be tested by its assumptions. What is important is specifying the conditions under which the hypothesis works. What matters is its predictive power, not it's conformity to reality.***

Milton Friedman, 1953. *The Methodology of Positive Economics*.
in *Essays in Positive Economics*: University of Chicago Press.

## About forecasting

- Predicting future economic outcomes is helpful in making appropriate plans, making investment decisions, conducting economic policies.

- We make inference about future outcome using available data (for time series: current and past data) and statistical models.
  We call this process econometric forecasting

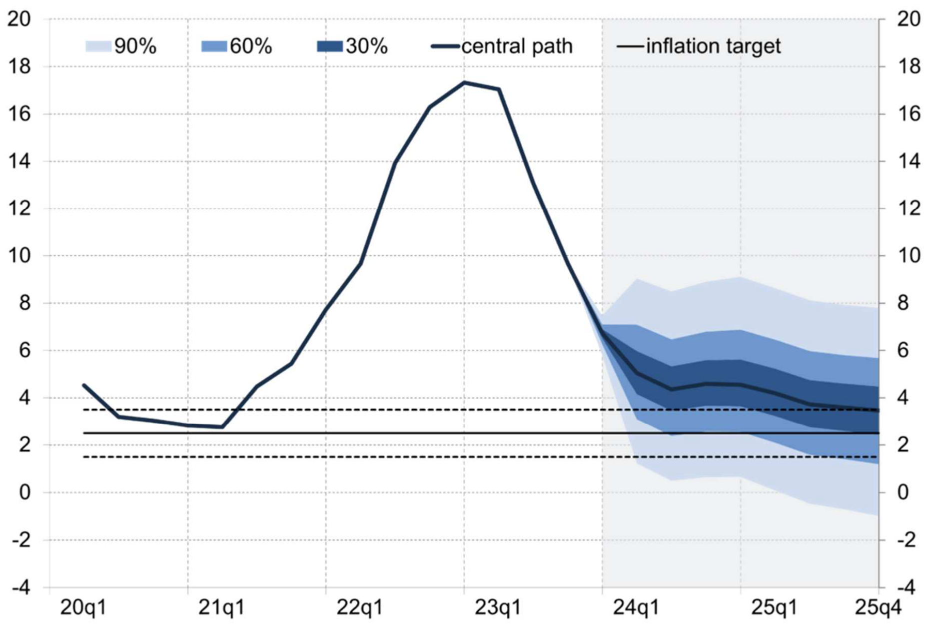
- Point forecast from model $M$, horizon $h$ and information set $\Omega_T$:

$$y^f_{T,h} = y_{T+h|T} = E_T(y_{T+h}|M) = E(y_{T+h}|M, \Omega_T)$$

- Density forecast provides information on all quantiles of the distribution. We focus on the entire distribution (pdf):

$$p^f_{t,h}(u) = p_{t+h|t}(u)$$

## Example of forecast for inflation in Poland




Source: Narodowy Bank Polski, Inflation Report

## About forecasting

### Types of time series forecasts

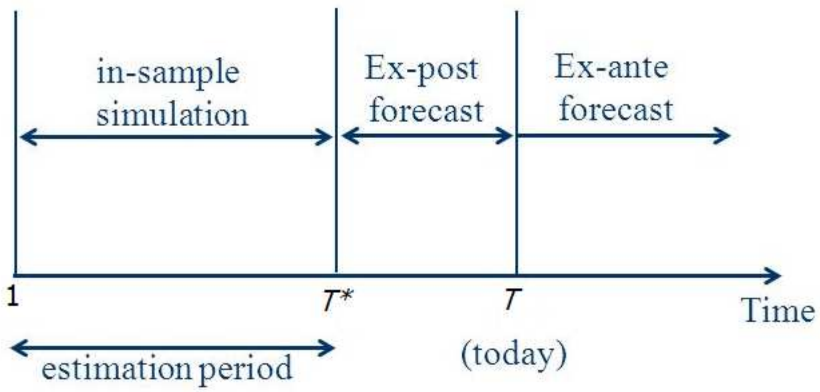
- Qualitative / model-based (e.g. from VAR/DSGE model)
- Quantitative / expert based (e.g. survey forecast, SPF)

### General characteristics of time series forecasts:

- Forecasting is based on the assumption that the past predicts the future
  *Think carefully if the past is related to what you expect about the future*
- Forecasts are always wrong
  *However, some models/methods might work better or worse than the other*
- Forecasts are usually more accurate for shorter time periods
  *But, economic theories are more informative for longer horizon*

## ex-ante vs ex-post forecast

- Ex-ante forecast is a true inference about the future
  It is for periods in which we **don't know the realization**
- Ex-post forecast is to check model reliability
  It is for periods in which we **know the realization**




## Ex-ante forecast:
### error in known ARMA/VAR model

- Assume we know DGP, i.e. the parameters and the specification of ARMA/VAR. We therefore know the parameters of infinite moving average representation

$$y_t = \mu + \psi_0\eta_t + \psi_1\eta_{t-1} + \psi_2\eta_{t-2} + \psi_3\eta_{t-3} \dots \qquad \eta_t \sim N(0, I)$$

- Forecast from known DGP is called **optimum forecast.** We cannot obtain more accurate forecast from another model.

- Forecast error of optimal forecast is solely due to futures shocks (random error):
  $$y_{T+h} - y_{T+h|T} = \psi_0\eta_{T+h} + \psi_1\eta_{T+h-1} + \dots + \psi_{h-1}\eta_{T+1}$$

- The resulting variance of forecast is:
  $$E\left\{\left(y_{T+h} - y_{T+h|T}\right)^2\right\} = \psi_0\psi_0' + \psi_1\psi_1' + \dots + \psi_{h-1}\psi_{h-1}'$$

## Ex-ante forecast
### error in estimated ARMA/VAR model

- Assume that we don't know the true DGP but use a model $M$ instead
- The variance of our forecast is:

$$E\{(y_{T+h} - y^M_{T+h|T})^2\} = E\{(y_{T+h} - y_{T+h|T})^2\} + \\ + E\{(y_{T+h|T} - y^M_{T+h|T})^2\} + \\ + 2E\{(y_{T+h} - y_{T+h|T})(y_{T+h|T} - y^M_{T+h|T})\}$$

Component A: error of "optimum forecast" (see previous slide)

Component B: estimation / misspecification error
we want to minimize this value

Component C: equals to 0 if we cannot forecast future shock

## Ex-ante forecast
### estimation / misspecification error

Let us focus on the estimation / misspecification error and model complexity

$$E\{\left(y_{T+h|T} - y^M_{T+h|T}\right)^2\}$$

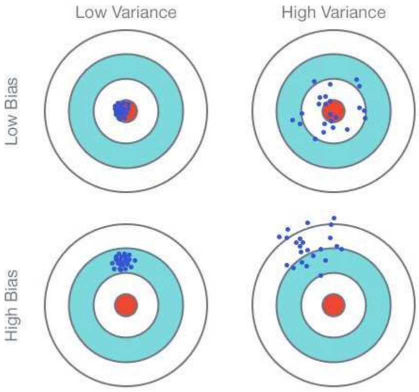
I. **Large / complex models**
   - many parameters = large estimation error (high variance)
   - many explanatory variables = good specification (low bias)

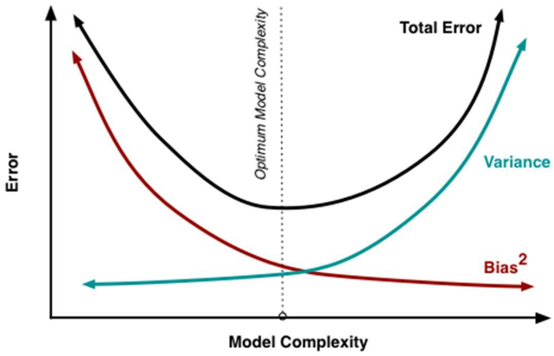
II. **Small / simple models**
   - few parameters = small estimation error (low variance)
   - few explanatory variables = potential misspecification (high bias)

Which effect dominates? We don't know and need to check it

# Ex-ante forecast

## Illustration of the variance / bias trade-off




# Ex-ante forecast

## Illustration of the variance / bias trade-off

- Let as assume that the true DGP is AR(1):

$$y_t = \mu + \rho(y_t - \mu) + \epsilon_t$$

- We have a sample of 180 monthly observations (15 years) for $y_t$ and would like to decide on one of the three competing models:

| | |
|---|---|
| RW, Random walk: | $\mu^{RW} = 0$ and $\rho^{RW} = 1$ |
| HL, 5-year half life model: | $\mu^{HL} = \bar{y}$ and $\rho^{HL} = 0.5^{1/60}$ |
| AR, estimated AR model: | $\mu^{AR}$ and $\rho^{AR}$ are estimated |

- Which model performs best? It depends on the value of $\rho$

# Ex-ante forecast

## Illustration of the variance / bias trade-off

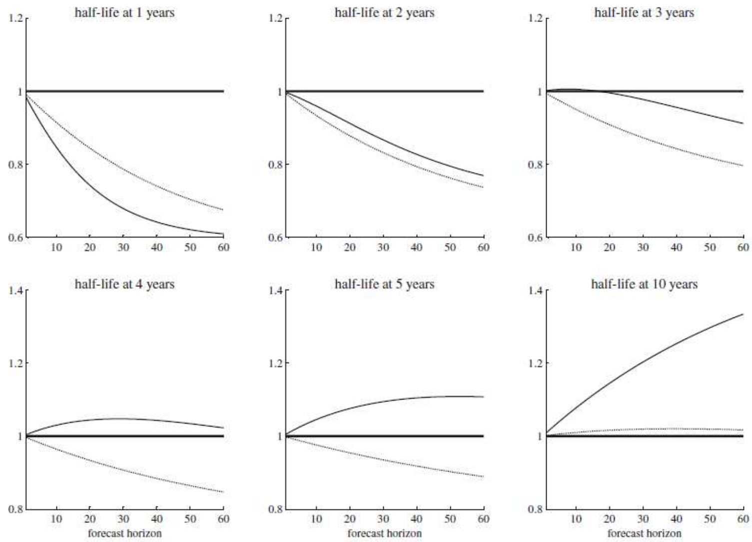
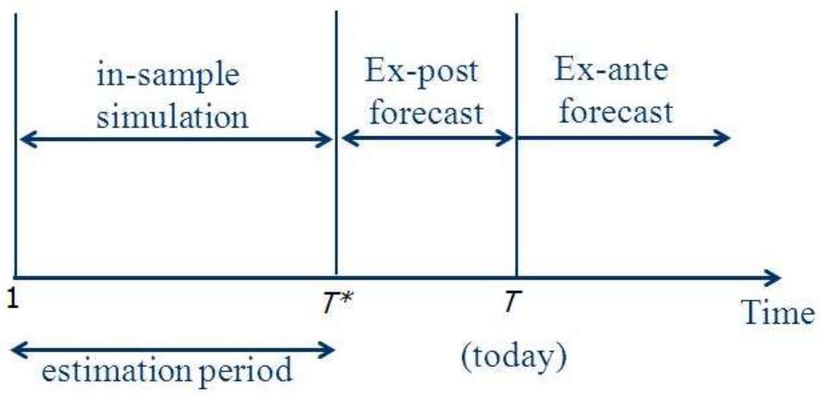



**Fig. 3** Theoretical mean squared forecast errors. Notes: Each *line* represents the ratio of MSFE from a given method to MSFE from the random walk, where values below unity indicate better accuracy of point forecasts. The *straight* and *dotted lines* stand for AR and HL, respectively. The forecast horizon is expressed in months

*Source: Ca' Zorzi M., Mućk J., Rubaszek M., 2016. RER forecasting and PPP: This time the Random Walk loses, Open Economies Review*

# About ex-post forecast

- We usually work with models that performed well in the past
- In ex-post forecast we ask a question :
  *how accurate forecasts the model would deliver if it was used in the past*
- We evaluate *ex-post* forecasts to be sure about model reliability
- An important issue is the use of "real time data, RTD"



## About *ex-post* forecast

- We compare forecast $y^{f}_{t,h}$ from model $M_i$ to realization $y_{t+h}$ to assess:
  - the absolute quality of forecasts from model $M_i$
    MFE, effciency/unbiasedness tests, sequential forecasts, PIT
  - the relative quality of forecasts from models $M_i$ and $M_j$
    RMSFE/MSFE/MAFE, log predictive scores
- Various forecasting schemes
  - rolling scheme
  - recursive schemes
  - fixed schemes
- A very important choice relates to the split of the sample into estimation and evaluation subsamples

## Forecasting schemes - illustration

**Figure 1(a). Recursive estimation scheme**

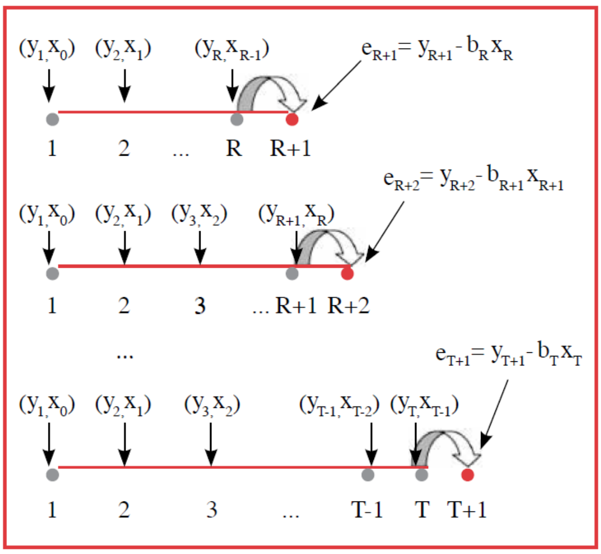
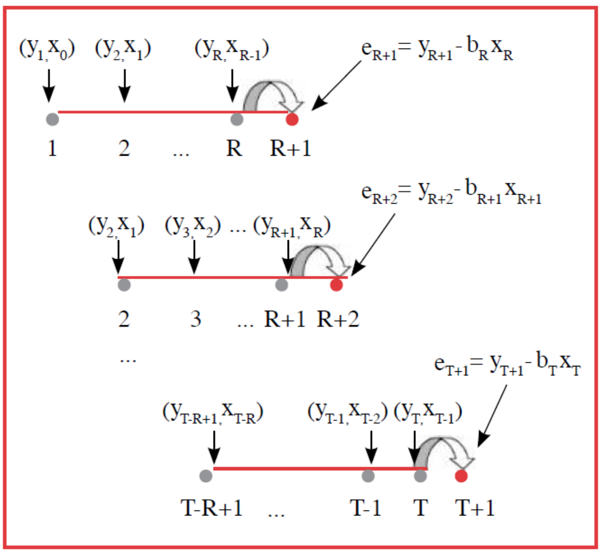


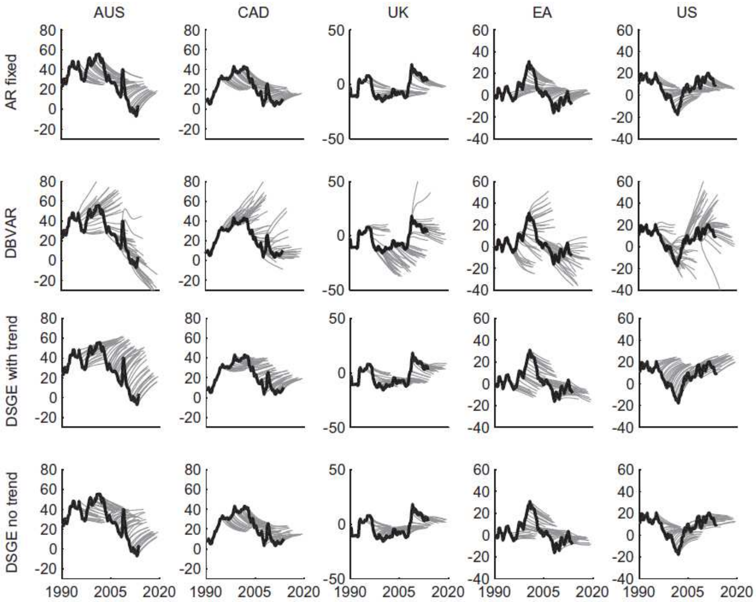

**Figure 1(b). Rolling estimation scheme**




*Source: Barbara Rossi, 2014. Density forecasts in economics and policymaking, CREI WP 37*

# Point forecasts accuracy measures

**Fig. 3.** Sequential real exchange rate forecasts.




*Source: Ca' Zorzi M. & Kolasa M. & Rubaszek M., 2017. Exchange rate forecasting with DSGE models, Journal of International Economics*

69

# Point forecasts accuracy measures

**Mean forecasts error for horizon $h$:** $MFE_h \quad = \frac{1}{T_h}\sum_{t=T_1+1}^{T-h}(y_{t+h} - y_{t,h}^f)$

**Root mean squared forecast error:** $RMSFE_h \; = \sqrt{\frac{1}{T_h}\sum_{t=T_1+1}^{T-h}\left(y_{t+h} - y_{t,h}^f\right)^2}$

where $\mathrm{T}_h = T - T_1 - h + 1$

**Diebold-Mariano test for equal forecast accuracy:**

Forecast errors from two competing models $e_{1t,h} = y_{t+h} - y_{1t,h}^f$ and $e_{2t,h} = y_{t+h} - y_{2t,h}^f$

The quadratic loss differential $d_{t,h} = e_{1t,h}^2 - e_{2t,h}^2$

The null of equal forecast accuracy (RMSFE) $H_0 \colon E(d_{t,h}) = 0$

Test statistic: $DM = \frac{\bar{d}_{t,h}}{\sqrt{S/T_h}} \sim N(0,1)$

where $S = \sum_{i=-(h-1)}^{h-1} \hat{\gamma}(i)$ is the ``long-term'' variance

# Point fct. accuracy measures: illustration

Mean Forecast Errors (MFEs) of Unconditional Forecasts

| $h$ | DSGE | SPF | DSGE-VAR($\hat{\lambda}$) $p=2$ | $p=4$ | $p=6$ | $p=8$ | DSGE-VAR($\infty$) $p=16$ |
|---|---|---|---|---|---|---|---|
| | | | Output growth (real GDP, QoQ SAAR) | | | | |
| 0 | −0.57** | 0.38 | −0.98*** | −0.98*** | −0.86*** | −0.79*** | −0.54* |
| 1 | −0.24 | 0.17 | −1.01*** | −0.98*** | −0.87** | −0.70** | −0.39 |
| 2 | −0.03 | 0.12 | −1.05*** | −0.94** | −0.74* | −0.55 | −0.31 |
| 3 | 0.11 | 0.02 | −1.07*** | −0.86** | −0.66* | −0.45 | −0.28 |
| 4 | 0.07 | −0.19 | −1.18*** | −0.88** | −0.72* | −0.51 | −0.37 |

Root Mean Squared Forecast Errors (RMSFEs) of Unconditional Forecasts

| $h$ | DSGE | SPF | DSGE-VAR($\hat{\lambda}$) $p=2$ | $p=4$ | $p=6$ | $p=8$ | DSGE-VAR($\infty$) $p=16$ |
|---|---|---|---|---|---|---|---|
| | | | Output growth (real GDP, QoQ SAAR) | | | | |
| 0 | 1.95 | 0.98 | 1.10 | 1.17** | 1.15 | 1.16 | 1.04 |
| 1 | 1.99 | 1.07 | 1.18 | 1.23** | 1.23* | 1.23* | 1.10 |
| 2 | 1.83 | 1.21** | 1.23* | 1.29** | 1.26* | 1.25* | 1.13* |
| 3 | 1.89 | 1.19** | 1.23** | 1.23** | 1.21* | 1.19* | 1.14** |
| 4 | 2.10 | 1.16** | 1.21** | 1.16** | 1.17* | 1.18* | 1.13** |

Source: Kolasa, Rubaszek, Skrzypczynski (2012, JMBC)

# Point fct. accuracy measures: efficiency

**Efficiency / unbiasedness test**

A relatively good forecast accuracy does not imply that they are satisfactory in the absolute sense! Absolute performance include ME and efficiency/unbiasedness test. For regression:

$$y_{t+h} = \alpha_0 + \alpha_1 y^f_{t,h} + \epsilon_{t,h}$$

we test whether $\alpha_0 = 0$ and $\alpha_1 = 1$.

[ the alternative specification is $e_{t,h} = \alpha_0 + \alpha_1 y^f_{t,h} + \epsilon_{t,h}$ in which we test $\alpha_0 = 0$ and $\alpha_1 = 0$ ]

Efficiency Test for Unconditional Forecasts

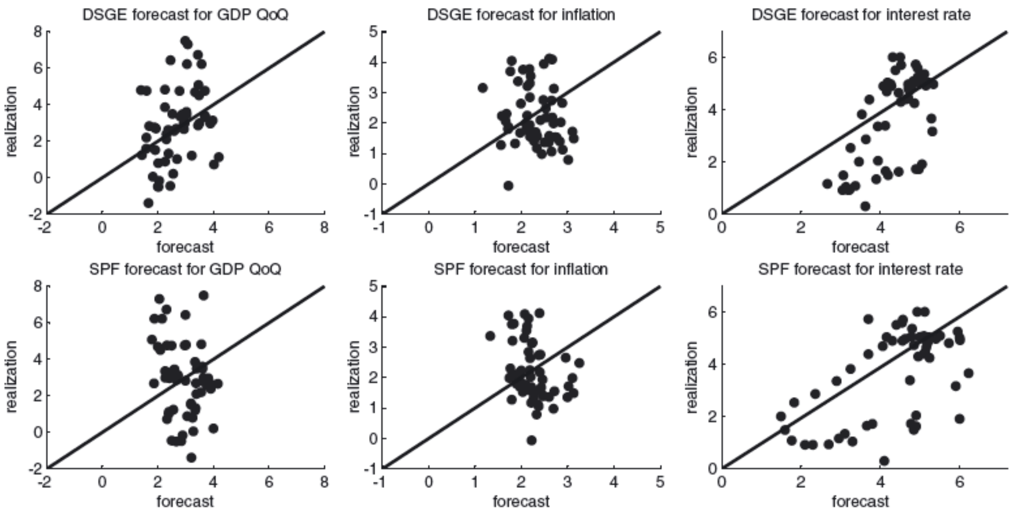
| $h$ | DSGE $\hat{\alpha}_0$ $(S_{\hat{\alpha}_0})$ | $\hat{\alpha}_1$ $(S_{\hat{\alpha}_1})$ | $R^2$ | $\chi^2$ prob | SPF $\hat{\alpha}_0$ $(S_{\hat{\alpha}_0})$ | $\hat{\alpha}_1$ $(S_{\hat{\alpha}_1})$ | $R^2$ | $\chi^2$ prob | DSGE-VAR4 $\hat{\alpha}_0$ $(S_{\hat{\alpha}_0})$ | $\hat{\alpha}_1$ $(S_{\hat{\alpha}_1})$ | $R^2$ | $\chi^2$ prob |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Output growth (real GDP, QoQ SAAR) | | | | | | | |
| 0 | 0.59 | 0.68 | 0.17 | 10.3 | 1.09 | 0.74 | 0.15 | 3.29 | 1.49 | 0.39 | 0.05 | 39.9 |
| | (0.88) | (0.21) | | *0.01* | (0.61) | (0.19) | | *0.19* | (0.94) | (0.19) | | *0.00* |
| 2 | 0.52 | 0.82 | 0.17 | 0.47 | 5.03 | −0.73 | 0.05 | 18.6 | 2.73 | 0.06 | 0.00 | 20.0 |
| | (0.77) | (0.27) | | *0.79* | (1.17) | (0.42) | | *0.00* | (1.31) | (0.31) | | *0.00* |
| 4 | 0.12 | 0.98 | 0.11 | 0.06 | 5.04 | −0.78 | 0.05 | 15.2 | 2.94 | −0.05 | 0.00 | 13.2 |
| | (1.14) | (0.43) | | *0.97* | (1.55) | (0.49) | | *0.00* | (2.62) | (0.67) | | *0.00* |

Source: Kolasa, Rubaszek, Skrzypczynski (2012, JMBC)

# Point forecasts accuracy measures

## Efficiency / unbiasedness test – graphical illustration

FIG. 2. Actuals and Four-Quarter-Ahead Forecasts.




*Source: M. Kolasa & M. Rubaszek & P. Skrzypczyński, 2012. Putting the New Keynesian DSGE Model to the Real-Time Forecasting Test, Journal of Money, Credit and Banking*

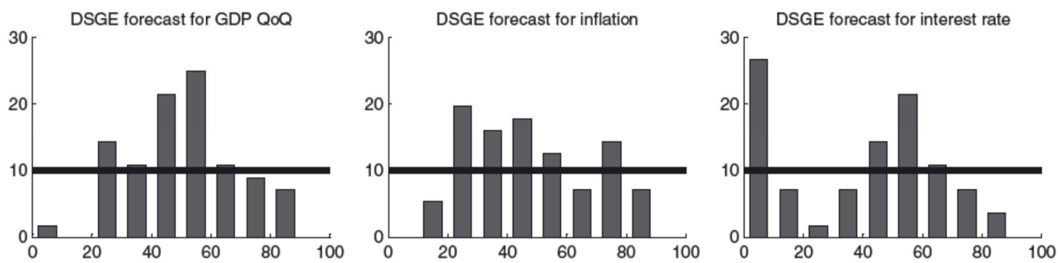
# Density forecasts accuracy: PIT

**PIT – probability Integrat Transform**

$$PIT_{t,h} = \int_{-\infty}^{y_{t+h}} p_{t,h}^{f}(u)du$$

where $p_{t,h}^{f}()$ is the forecast for density distribution.

For a well calibrated model the series $PIT_{t,h}$ should be drawn from $IID\ U(0,1)$




Source: Kolasa, Rubaszek, Skrzypczynski (2012, JMBC)

## Density forecasts accuracy: predictive scores

**LPS – log predictive score**

$$LPS_{t,h} = \log(p_{t,h}^{f}(y_{t+h}))$$

where $p_{t,h}^{f}()$ is the forecast for density distribution.

**We can compare density forecasts from two models with the Amisano and Giacomini (2007) test of equal forecast accuracy:**

The loss differential $L_{t,h} = LPS_{1t,h} - LPS_{2t,h}$

The null of equal forecast accuracy $H_0: E(L_{t,h}) = 0$

Test statistic: $GA = \frac{\bar{L}_{t,h}}{\sqrt{S/T_h}} \rightarrow N(0,1)$

where $S$ is the HAC (Newey and West) estimator of the ``long-term'' variance for $L_{t,h}$

* Amisano, G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. Journal of Business & Economic Statistics 25, 177-190.

## LPS: illustration

Table : Log Predictive Scores (LPS) for DSGE models

| H | United Kingdom | | | | Canada | | | | Australia | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LS | JP- | JP | JP+ | LS | JP- | JP | JP+ | LS | JP- | JP | JP+ |
| | | | | | | Output | | | | | | |
| 1 | -0.15$^{\bullet}$ | 0.04$^{*}$ | -0.07$^{\bullet}$ | -0.08$^{\bullet}$ | -0.39$^{\bullet}$ | -0.05 | -0.18$^{\circ}$ | -0.06 | 0.01 | 0.03$^{\bullet}$ | -0.06$^{\bullet}$ | -0.09$^{\bullet}$ |
| 2 | -0.13$^{\circ}$ | 0.05 | -0.06 | -0.08$^{\bullet}$ | -0.26$^{\bullet}$ | -0.03 | -0.10 | 0.00 | 0.03 | 0.04$^{\bullet}$ | -0.06$^{\bullet}$ | -0.12$^{\bullet}$ |
| 4 | -0.14 | 0.03 | -0.10 | -0.12$^{\bullet}$ | -0.07 | 0.01 | -0.11$^{\circ}$ | 0.02 | 0.09$^{\circ}$ | 0.07$^{\bullet}$ | -0.06$^{\bullet}$ | -0.19$^{\bullet}$ |
| 6 | -0.17 | 0.02 | -0.17 | -0.16$^{\bullet}$ | 0.06 | 0.02 | -0.16$^{\circ}$ | 0.02 | 0.15$^{*}$ | 0.11$^{\bullet}$ | -0.07$^{\bullet}$ | -0.24$^{\bullet}$ |
| 8 | -0.22$^{*}$ | 0.02 | -0.24 | -0.20$^{\bullet}$ | 0.12$^{*}$ | 0.03 | -0.21$^{*}$ | 0.01 | 0.22$^{\bullet}$ | 0.14$^{\bullet}$ | -0.08$^{\bullet}$ | -0.28$^{*}$ |
| 12 | -0.34$^{\bullet}$ | 0.01 | -0.37$^{\bullet}$ | -0.29$^{\bullet}$ | 0.19$^{\bullet}$ | 0.04 | -0.30$^{*}$ | -0.01 | 0.33$^{\bullet}$ | 0.20$^{\bullet}$ | -0.12$^{\bullet}$ | -0.31$^{*}$ |
| | | | | | | Three variables | | | | | | |
| 1 | -0.44$^{\bullet}$ | -0.01 | -0.31$^{\bullet}$ | -0.11$^{\bullet}$ | -0.51$^{\bullet}$ | -0.04 | -0.31$^{*}$ | -0.17$^{*}$ | -0.13$^{\bullet}$ | 0.06$^{\bullet}$ | -0.15$^{\bullet}$ | -0.17$^{\bullet}$ |
| 2 | -0.41$^{*}$ | 0.07$^{\bullet}$ | -0.09$^{\bullet}$ | -0.06$^{*}$ | -0.50$^{\bullet}$ | -0.03 | -0.23$^{*}$ | -0.11 | -0.12$^{*}$ | 0.12$^{\bullet}$ | -0.12$^{\bullet}$ | -0.24$^{\bullet}$ |
| 4 | -0.40 | 0.16$^{\bullet}$ | 0.02 | -0.04 | -0.49$^{\bullet}$ | 0.02 | -0.18$^{*}$ | 0.01 | -0.12 | 0.19$^{\bullet}$ | -0.13$^{\bullet}$ | -0.35$^{\bullet}$ |
| 6 | -0.42 | 0.23$^{\bullet}$ | -0.01 | -0.06 | -0.40$^{*}$ | 0.00 | -0.18$^{*}$ | 0.10 | -0.13 | 0.23$^{\bullet}$ | -0.15$^{\bullet}$ | -0.43$^{\bullet}$ |
| 8 | -0.43 | 0.29$^{\bullet}$ | -0.05 | -0.08 | -0.28 | -0.01 | -0.17$^{*}$ | 0.16 | -0.14 | 0.27$^{\bullet}$ | -0.16$^{\bullet}$ | -0.47$^{\bullet}$ |
| 12 | -0.45 | 0.36$^{\bullet}$ | -0.16 | -0.15$^{\circ}$ | 0.04 | -0.01 | -0.16$^{\circ}$ | 0.23$^{\circ}$ | -0.19 | 0.34$^{\bullet}$ | -0.21$^{\bullet}$ | -0.50$^{\bullet}$ |

Notes: The figures in the table represent the differences of the LPS from a given model in comparison to the NK benchmark so that positive values indicate that forecasts from a given NOEM variant are more accurate than from the benchmark. Asterisks $^{\bullet}$, $^{*}$ and $^{\circ}$ denote, respectively, the 1%, 5% and 10% significance levels of the two-tailed Amisano and Giacomini (2007) test, where the long-run variance is calculated with the Newey-West method.

Source: Kolasa and Rubaszek (2018, IJF)

## Forecasting competition participants

In the forecasting competition we will include:

**Models discussed during lectures:**

- RW
- ARMA
- VAR

**Two machine learning tools:**

- Autoregressive feedforward neural network
  `nnetar()`function in `forecast` package

- Random forest
  `randomForest()` function in `randomForest` package

## ML techniques in time series forecasting

- Most ML methods, have no awareness of time: they take observations to be independent and identically distributed. As time series data are usually characterized by serial dependence, ML methods might encounter difficulties to predict a trend or cycle

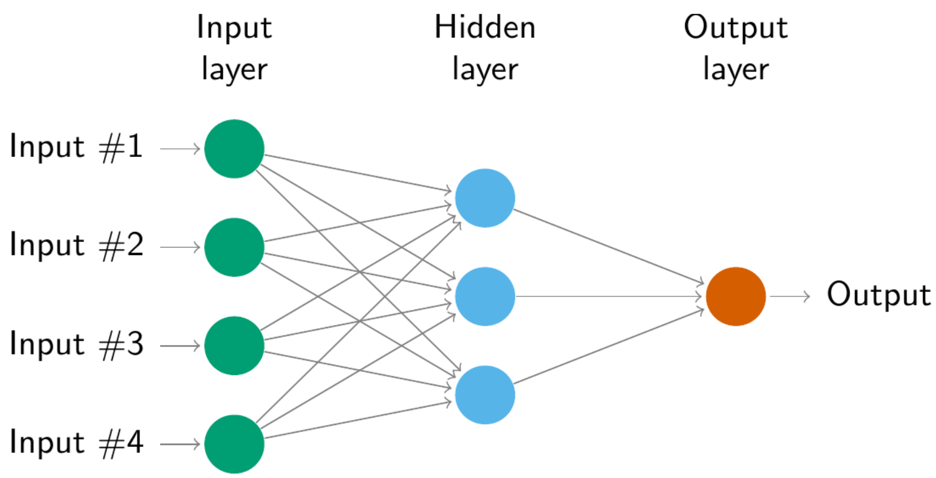
- ML methods encounter difficulties to predict values that fall outside the range of values of the target in the training set. Hence, preparation of data matters (transformations – stationarity is essential, etc.)

- Training ML models usually requires stable/deterministic relationships and a large number of data. Macroeconomic/financial time series are usually short and describe unstable relations

- For the above reasons the success of ML techniques in Macroeconomic Time series forecasting is limited

## Autoregressive feedforward neural network

$NNAR(P, K)$ is AR NN with $P$ lagged inputs and $K$ neurons in the hidden layer, so that $NNAR(P, 0)$ is equivalent to an $ARMA(p, 0)$




Source: https://otexts.com/fpp3/nnetar.html

## Autoregressive feedforward neural network

Inputs into each hidden neuron are combined linearly:

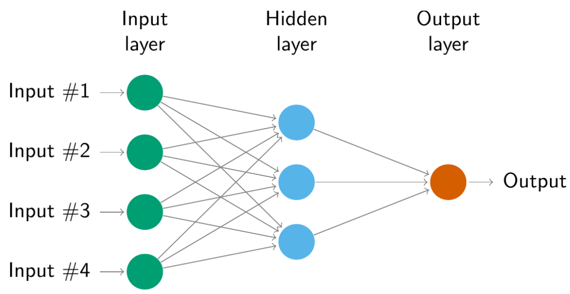
$$z_{jt} = b_j + \sum_{i=1}^{p} w_{ij} y_{t-p}$$

Next, they are transformed with the sigmoid function:

$$u_{jt} = \frac{1}{1 + e^{-z_{jt}}}$$

So that perdiction for output variable is:

$$y_{t|t-1} = c + \sum_{k=1}^{K} v_k u_{jt}$$




Source: https://otexts.com/fpp3/nnetar.html

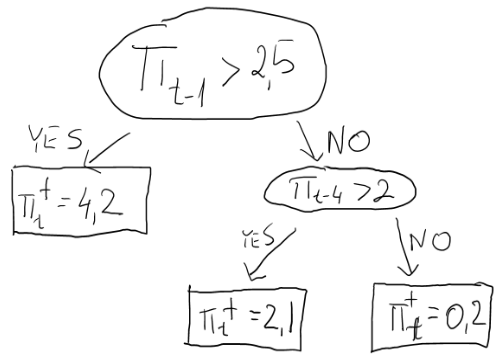
- Parameters $b, c, w$ *and* $v$ are "learned" (or estimated) from the data.
- Learning proces starts with random values, which are then updated using the data. The number of iterations/epochs is a hyperparameter.
- There is an element of randomness in the predictions produced by a neural network → it is usually trained several times using different random starting points and averaged.

## Autoregressive random forest

Autoregressive **random forest** is an ensambling method based on averaging $M$ autoregressive decision trees

**To grow an autoregressive tree we need to set:**

- Number of lags
- Maximum depth (number of nodes)
- Number of variables randomly sampled as candidates at each split
- Bagging method
  (split based on a random „bag" of observations)




## Autoregressive random forest

**Growing the tree:**

- The aim is to divide the entire sample into $J$ terminal leaves to minimize:

$$RSS = \sum_{j=1}^{J} \sum_{t=1}^{T} \left(y_t - \bar{y}_j\right)^2$$

- Computationally infeasible to consider every possible partition, hence we often resort to *recursive binary splitting* (into 2 new branches top-down). In this method we choose the best split at a given node by selecting the „split variable" and threshold to minimize $RSS$ after the division:

$$RSS^* = \sum_{t \in S_1} (y_t - \bar{y}_1)^2 + \sum_{t \in S_2} (y_t - \bar{y}_2)^2$$

- The alternative is to look many steps ahead (a bad split early on might be followed by an excellent split later on). In this approach we grow a large tree and *prune* it back

- Post estimation: we can look at importance of regressors

## Exercises

**Exercise 4.1.**

Select an EU country / a variable of interest (inflation, unemployment, output growth) and:

- Calculate recursive point forecasts from RW, ARMA, VAR models over the last 3 years
- Calculate MFE and RMSFE for the 3 methods
- Compare the accuracy of forecasts from 3 models to RW with DM test
- Conduct efficiency test and draw a scatter-plot for forecasts from the VAR
- Make a plot for sequential forecasts from VAR and BVAR models
- Discuss the results

# Block 1 presentation

Select an EU country and a variable (e.g. inflation, GDP growth rate, unemployment rate or gov. bond 10Y yield).

**a.** **<1.0p>** Describe the variable. To show: time series plot, ACF, UR test

**b.** **<2.0p>** Estimate the ARMA model. To show: information criteria, IRF.

**c.** **<3.0p>** Estimate the VAR model for a vector $(y_t^* \; y_t)'$, where $y_t^*$ is the value of the variable for the euro area, and perform Cholesky structuralization. To show: IRF, FEVD, historical decomposition

**d.** **<2.0p>** Compare the accuracy of forecasts from RW, ARMA and VAR models. To show: MFE, RMSFE, DM test, sequential forecasts graph

**e.** **<2.0p>** Plot forecast from ARMA, VAR and for the next two years and from European Comission (to be found on the webpage). To show: A table with forecasts, a graph with three forecasts

Additionally, I attribute up to 2p for the quality of the presentation (1p. for the .pdf and 1p. for the speech / interpretation of the results). Presentation should take around 7 minutes.

# Block 2

## Forecasting investment risk

## with

## GARCH, MGARCH and copula

***TOPICS***

5. Risk of a univariate portfolio: GARCH model
6. Risk of a multivariate portfolio: MGARCH model
7. Risk of a multivariate portfolio: Copula
8. Backtesting

# Topic 5

# Univariate portfolio. GARCH models

- Downloading data from stooq.pl to `R`
- Descriptive stats: moments, ACF, density plot, QQ plot
- Value at Risk (VaR) and Expected Shortfall (ES)
- Stylized facts for asset returns
- Normal and t-Student distribution
- Historical simulation
- Exponentially Weighted Moving Average (EWMA) model
- GARCH models
- GARCH extensions (GJR-GARCH, EGARCH, GARCH-in-mean)
- Monte Carlo simulations
- Calculating VaR/ES from GARCH models

## *Rates of return: reminder*

**Simple returns:**

- Easy to calculate for a portfolio of assets: $R_p = \sum_{k=1}^{K} w_k R_k$
- Easy to communicate to non-statisticians
- Not symmetric nor additive...

**Log returns:**

- Symmetric and additive
- Easy to communicate to statisticians
- Difficult to calculate for a portfolio of assets: $r_p \neq \sum_{k=1}^{K} w_k r_k$

**Even though the latter, we will work with log returns**

## *Financial series characteristics*

1. No autocorrelation of returns (which is not equivalent to independence):

$$cor(r_t, r_{t-s}) = 0$$

2. Fat tails:

$$kurtosis(r_t) > 3$$

3. Volatility clustering:

$$cor(r_t^2, r_{t-s}^2) > 0$$

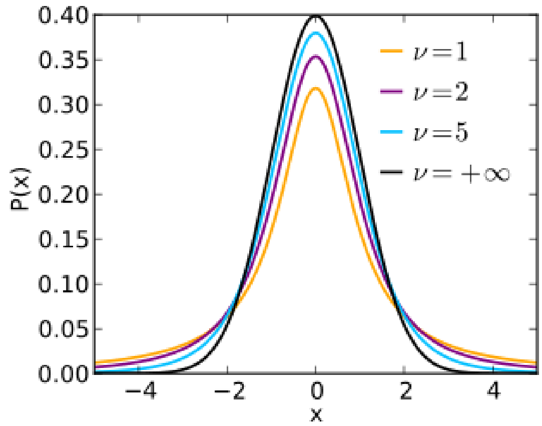
4. Leverage effect (sometimes):

$$cor(r_t^2, r_{t-s}) < 0$$

## *Fat tails: t-Student distribution*

**t-Student distribution $t_v$:**

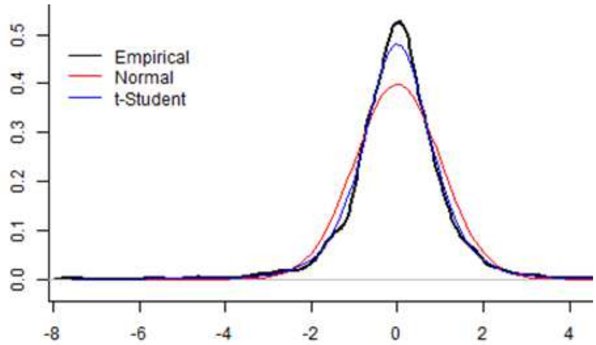
- for $v = \infty$ equivalent to $N(0,1)$
- for $v \leq 2$ no variance (which is equal to $\frac{v}{v-2}$)
- Kurtosis: $K = 3 + \frac{6}{v-4}$
- For stocks usually $v \sim 5$
- One can check with QQ plot / density plot

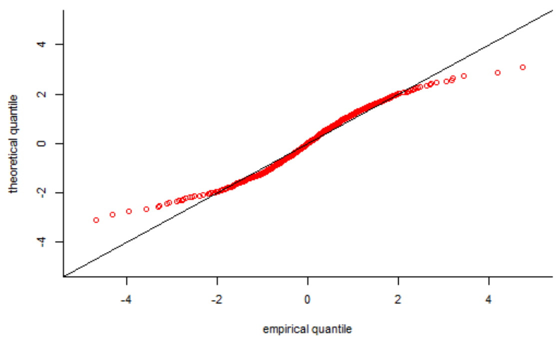



5% critical values for two-tailed t-Student $t_v$

| v | 1 | 2 | 3 | 5 | 10 | 50 | ∞ |
|---|---|---|---|---|---|---|---|
| t* | -12.71 | -4.30 | -3.18 | -2.57 | -2.23 | -2.01 | -1.96 |

## *Fat tails: illustration*

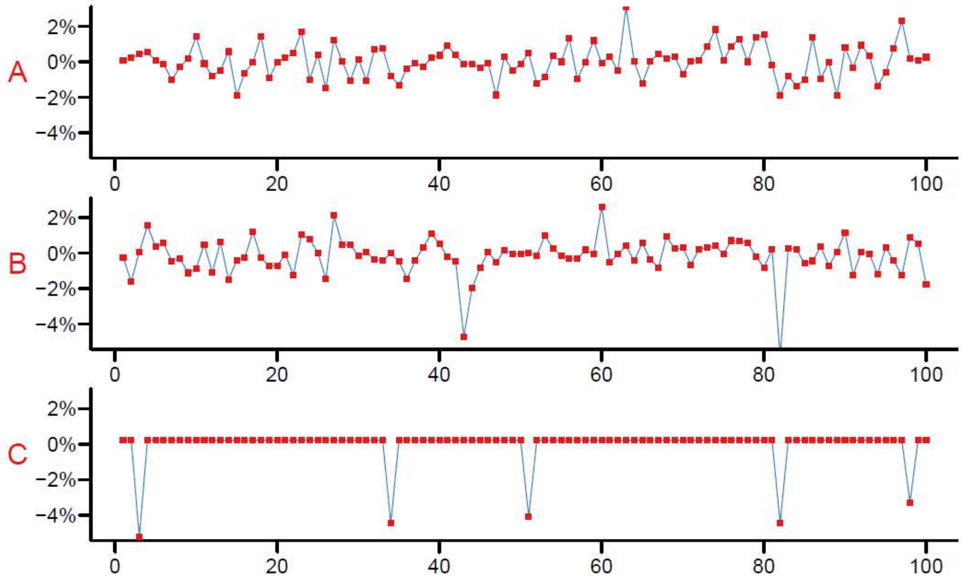



```
|N     |  250.063|
|mu    |    7.534|
|sig   |   32.134|
|min   |  -15.352|
|max   |   12.607|
|skew  |   -0.197|
|kurt  |    7.577|
|JB    |  2199.87|
```

## *Risk is not equivalent to variance!!!*

Three series with $E(r) = 0$ and $Var(r) = 1$ (see Danielson, 2012)




## *VaR and ES: definitions*

**Value at Risk (VaR) for tolerance level *p*:**

$$p = \int_{-\infty}^{VaR_p} f(r)dr$$

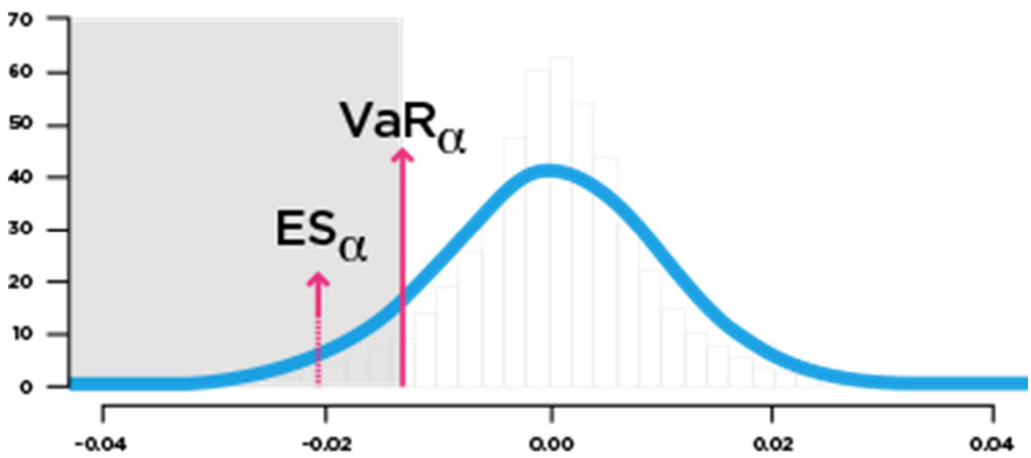
$$P(r \leq VaR_p) = p$$

**Expected shortfall (ES) = Conditional VaR (CVaR):**

$$ES_p = E(r | r \leq VaR_p)$$

$$ES_p = \frac{1}{p}\int_{-\infty}^{VaR_p} rf(r)dr$$

## *VaR and ES: illustration*




## *VaR and ES: calculation stages*

1. Setting tolerance level $p$
2. Setting horizon $H$
3. Choosing estimation sample period $1{:}\,T$
4. Choosing a model
5. VaR/ES computation (for period $T+1$)

   \+ Model validation

**Basel ii/iii:** VaR as a risk measure

**Basel iv:** plans to change into ES

## *VaR: Basel II*

***Quantitative standards Basel II***

a. 99th percentile V*aR* must be computed on a daily basis
b. In calculating VaR the minimum *"holding period"* will be ten trading days. Banks may use VaR numbers calculated according to shorter holding periods scaled up to ten days by the square root of time
c. The choice of sample period for calculating VaR is constrained to a minimum length of one year.
d. banks will be free to use models based, for example, on variance-covariance matrices, historical simulations, or Monte Carlo simulations
e. The multiplication factor will be set by individual supervisory authorities on the basis of their assessment of the quality of the bank's risk management system, subject to an absolute minimum of 3. Banks will be required to add to this factor a "plus" directly related to the ex-post performance of the model, thereby introducing a built in positive incentive to maintain the predictive quality of the model. The plus will range from 0 to 1 based on the outcome of so-called "backtesting."

## *VaR and ES calculation methods*

A. Parametric / non-parametric models

B. Analytical formula / Monte-Carlo simulations

C. Conditional / unconditional volatility

## Parametric models: normal distribution

Analytical formula for $r \sim N(\mu, \sigma^2)$:

$$VaR_p = \mu + \sigma\Phi^{-1}(p)$$
$$ES_p = \mu - \sigma\frac{\phi\left(\Phi^{-1}(p)\right)}{p}$$

where $\phi$ and $\Phi$ are normal distribution pdf and cdf.

Numerical integral formula

$$ES_p = \mu + \sigma\frac{\int_0^p \Phi^{-1}(s)ds}{p}$$

Tables for $r \sim N(0,1)$:

| $p$ | 0.5 | 0.1 | 0.05 | 0.025 | 0.01 | 0.001 |
|---|---|---|---|---|---|---|
| VaR | 0 | 1.282 | 1.645 | 1.960 | 2.326 | 3.090 |
| ES | 0.798 | 1.755 | 2.063 | 2.338 | 2.665 | 3.367 |

## Parametric models: t-Student distribution

Formula for VaR = quantile $p$:

$$VaR_p = \mu + \sigma T_v^{-1}(p)\sqrt{\frac{v-2}{v}}$$

where $T_v$ is the cdf of t-Student with $v$ degrees of freedom

Numerical integral formula for ES

$$ES_p = \mu + \sigma\frac{\int_0^p T_v^{-1}(s)ds\sqrt{\frac{v-2}{v}}}{p}$$

Notes:

- The variance of $X \sim t_v$: $Var(X) = \frac{v}{v-2}$
- In R functions relate to $t_v$ (e.g. `rt`) or scaled $t_v$ (e.g. `rdist` in `rugarch`)

## Non-parametric model: historical simulation

- We assume that the distribution of returns is well approximated by past/historical returns

- We sort past $T$ returns from the lowest to highest: $rs_1 \leq rs_2 \ldots \leq rs_T$ and calculate VaR as $p^{th}$ quantile. For $m = floor(pT)$:

$$VaR_p = rs_m$$

- $ES$ is equal to the average of the worst returns lower than VaR

$$ES_p = \frac{1}{m}\sum_{1}^{m} rs_i$$

## VaR and ES for further horizons

- To measure risk of investment for horizons $H > 1$ we need to approximate the distribution of:

$$y_H = \sum_{h=1}^{H} r_h$$

- Two kind of methods:
  - analytical (square root of time, SQRT)
  - numerical (Monte Carlo, bootstraping)

## *Square root of time method*

Let us assume that $r_t \sim N(\mu, \sigma^2)$ and $r_t$ are *IID*. Then:

$$y_H = \sum_{h=1}^{H} r_h \sim N(H\mu, H\sigma^2)$$

In this case:

$$VaR_H = H\mu + \sqrt{H} \times \sigma\Phi^{-1}(p)$$
$$ES_H = H\mu - \sqrt{H} \times \sigma \frac{\phi(\Phi^{-1}(p))}{p}$$

For $\mu = 0$ this simplifies to:

$$VaR_H = \sqrt{H} \times VaR \quad \text{and} \quad ES_H = \sqrt{H} \times ES$$

This is why we call this method **square root of time**

Note: this method applied only for IID returns with normal distribution

## *Monte Carlo simulations*

Let us assume that returns are t-Student
(or any other distr. for which we don't know analytical formula for the sum of vars.)
In this case we resort to Monte Carlo simulations

MC steps to calculate VaR/ES for any horizon $H$:

1. Draw a path $r_1, r_2, \ldots, r_H$ of returns over horizon $H$ and calculate $y_H = \sum_{h=1}^{H} r_h$
2. Repeat step 1 "$N$" Times. Save $y_H^{(n)}$ for $n = 1,2, \ldots, N$
3. Sort cumulated returns $ys_H^{(1)} \leq ys_H^{(2)} \leq \ldots$
4. Set $M = floor(pN)$
5. Use formulas :

$$VaR_H = ys_H^{(M)}$$

$$ES_H = \frac{1}{M}\sum_{1}^{M} ys_H^{(i)}$$

## Bootstrap

- When we use historical simulation method, an equivalent to MC simulations is Bootstrap
- Bootstrap steps to calculate VaR/ES for any horizon $H$:
  1. Draw $H$ times with replacement from sample $r_{1:T}$. Use draws $r_1, r_2, \dots, r_H$ to calculate $y_H = \sum_{h=1}^{H} r_h$
  2. Repeat step 1 "$N$" Times. Save $y_H^{(n)}$ for $n = 1,2, \dots, N$
  3. Sort cumulated returns $ys_H^{(1)} \leq ys_H^{(2)} \leq \dots$
  4. Set $M = floor(pN)$
  5. Use formulas :

$$VaR_H = ys_H^{(M)}$$

$$ES_H = \frac{1}{M} \sum_{1}^{M} ys_H^{(i)}$$

## Exercises

**Exercise 5.1.**

The rate of return of a portfolio is t-Student distributed, where the number of degrees of freedom is equal to 5 (critical values are provided in table below). Moreover, it is known that the expected rate of return is 5% and standard deviation is 20%.

| $p$ | 1% | 2% | 3% | 4% | 5% | 6% | 7% | 8% | 9% | 10% |
|---|---|---|---|---|---|---|---|---|---|---|
| t-Student | -3.36 | -2.76 | -2.42 | -2.19 | -2.02 | -1.87 | -1.75 | -1.65 | -1.56 | -1.48 |
| scaled t-Student | -2.61 | -2.14 | -1.88 | -1.70 | -1.56 | -1.45 | -1.36 | -1.28 | -1.21 | -1.14 |

a. Select the tolerance level $p$
b. Calculate VaR with pen and paper for $H = 1$ and $H = 4$ (with SQRT)
c. Calculate VaR and ES with R (for $H = 1$ and $H = 4$)
d. Compare the results from points b and c

Is SQT justified?

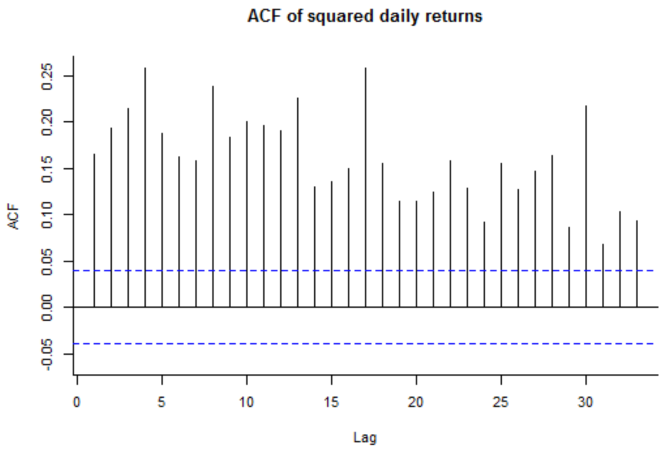
**Exercise 5.2.**

The rate of return has an IID uniform distribution $r \sim U(-0.05; 0.05)$.

a. Calculate VaR and ES for $p = 0.05$ or $0.10$
b. Can you find distribution for horizon $H = 2$?
c. Calculate VaR and ES for $p = 0.05$ or $0.10$ for horizon $H = 2$. Compare the results with SQRT.

## Volatility clustering

Financial series characteristics: $cor(r_t^2, r_{t-s}^2) > 0$




This implies that conditional variance is not constant in time:

$$\sigma_t^2 = E(r_t^2 | r_{t-1}^2, r_{t-2}^2, \dots) \neq \sigma^2$$

## Volatility clustering, EWMA

- **Moving Average (MA):**

$$\sigma_t^2 = \frac{1}{S}\sum_{s=1}^{S}(r_{t-s} - \mu)^2$$

- **Exponentially Weighted Moving Average (EWMA),**
  A model proposed by JP Morgan in 1993, known also as RiskMetrics

$$\sigma_t^2 = \sum_{s=1}^{\infty}\lambda^s r_{t-s}^2$$

- Simplified formula for EWMA model, i.e. IGARCH(1,1):

$$\sigma_t^2 = (1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2$$

- JP Morgan calibrate $\lambda = 0.94$ for daily observations

## *Volatility clustering, GARCH(1,1)*

- **EWMA …**

$$\sigma_t^2 = (1-\lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2$$

- **… as a specific version of GARCH(1,1):**

$$r_t = \mu + \epsilon_t, \;\; \epsilon_t \sim D(0, \sigma_t^2\,)$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

$$\omega > 0, \alpha, \beta \geq 0.$$

- **EWMA restrictions:**

$$\mu = 0$$
$$\omega = 0$$
$$\alpha = 1 - \lambda$$
$$\beta = \lambda$$

## *Volatility clustering, GARCH(1,1)*

- **MA …**

$$\sigma_t^2 = \frac{1}{S} \sum_{s=1}^{S} (r_{t-s} - \mu)^2$$

- **… as a specific version of GARCH(S,0):**

$$r_t = \mu + \epsilon_t, \;\; \epsilon_t \sim D(0, \sigma_t^2\,)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_S \epsilon_{t-S}^2$$

$$\omega > 0, \alpha, \beta \geq 0.$$

- **MA restrictions:**

$$\omega = 0$$
$$\alpha_s = 1/S \text{ for } s{=}1{,}2{,}\ldots S$$

## *Volatility clustering, GARCH(1,1)*

- **GARCH(1,1):**

$$r_t = \mu + \epsilon_t, \qquad \epsilon_t \sim D(0, \sigma_t^2)$$
$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$
$$\omega > 0, \alpha, \beta \geq 0.$$

- **Other notation:**

$$\sigma_t^2 = (1 - \alpha - \beta)\,\bar{\sigma}^2 + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

  where

$$\bar{\sigma}^2 = \frac{\omega}{1 - (\alpha + \beta)}$$

  is the equilibrium value of the variance.

- If $\alpha + \beta < 1$ then the variance is mean reverting (stationary model). For EWMA $\alpha + \beta = 1$: Integrated GARCH, IGARCH model

## *GARCH: estimation*

- The joint probability of all observations:

$$p(y_1, y_2, \dots, y_T) = p(y_1|\Omega_0) \times p(y_2|\Omega_1) \times \cdots \times p(y_T|\Omega_{T-1})$$

  where $\Omega_t$ is information set available till moment $t$

- If we assume that:

$$\epsilon_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$$

  then the likelihood is:

$$\mathcal{L}(\theta|y_{1:T}) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{(y_t - \mu_t)^2}{2\sigma_t^2}\right)$$

  where $\theta$ is the vector of model parameters

## GARCH: estimation

- In many cases the conditional distribution of returns is also characterised by excess kurtosis or skewness.
- In this case we can assume that conditional distribution has t-Student distribution or skewed t-Student distribution.
- For t-Student distribution the likelihood is:

$$\mathcal{L}(\theta|y_{1:T}) = \prod_{t=1}^{T} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi(v-2)}\sigma_t} \left(1 + \frac{(y_t - \mu_t)^2}{(v-2)\sigma_t^2}\right)^{-\left(\frac{v+1}{2}\right)}$$

- For skewed t-Student distribution, see rugarch vignette (link, page 19)

## GARCH(P,Q): specification selection

**Specification selection stages:**

1. Select the specification for levels ($\mu_t$), usually a constant
2. Select the specification for the variance ($\sigma_t^2$), usually GARCH(1,1)
3. Decide on the conditional distribution, usually t-Student

**Criteria:**

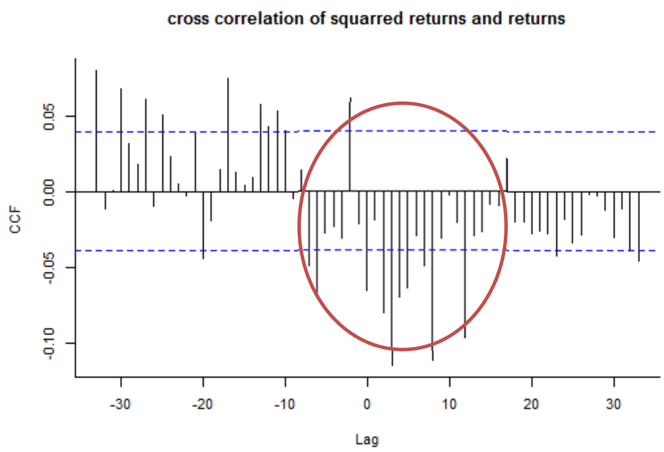
A. No autocorrelation for levels and squares of standarized residuals

$$u_t = \epsilon_t / \sigma_t$$

B. Minimization of information criteria (AIC, BIC, HQ)

## Leverage effect

One of financial series characteristics: $cor(r_t^2, r_{t-s}) < 0$




This implies that conditional variance depends on the sign of past returns

## Asymetric GARCH models

**GJR-GARCH(1,1) by Glosten-Jagannathan-Runkle (1993):**

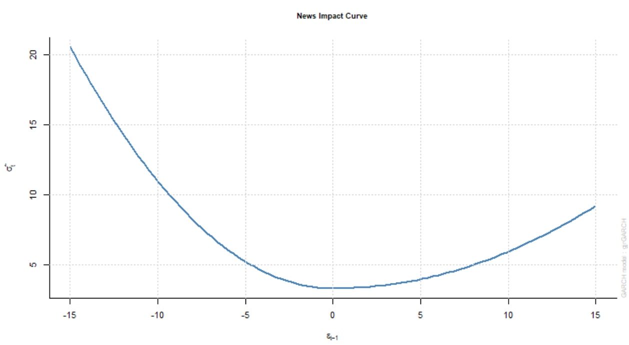
$$r_t = \mu + \epsilon_t,$$
$$\sigma_t^2 = \omega + (\alpha + \gamma I(\epsilon_{t-1} < 0))\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$
$$\omega > 0, \alpha, \beta, \gamma \geq 0.$$

where:

$$\mathrm{I}(x) = \begin{cases} 1 \text{ if } x \text{ is TRUE} \\ 0 \text{ if } x \text{ is FALSE} \end{cases}$$

so that:

$$\sigma_t^2 = \begin{cases} \omega + (\alpha + \gamma)\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 & \text{for } \epsilon_{t-1} < 0 \\ \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 & \text{for } \epsilon_{t-1} \geq 0 \end{cases}$$



## Asymetric GARCH models

**E(xponential)GARCH(1,1) by Nelson (1991):**

$$r_t = \mu + \epsilon_t,$$
$$\ln(\sigma_t^2) = \omega + \alpha u_{t-1} + \gamma |u_{t-1}| + \beta \ln(\sigma_{t-1}^2)$$

where $u_t = \epsilon_t / \sigma_t$ is a standarized error term

As a result:

$$\ln(\sigma_t^2) = \begin{cases} \omega + (\alpha - \gamma) u_{t-1} + \beta \ln(\sigma_{t-1}^2) \text{ for } \epsilon_{t-1} < 0 \\ \omega + (\alpha + \gamma) u_{t-1} + \beta \ln(\sigma_{t-1}^2) \text{ for } \epsilon_{t-1} \geq 0 \end{cases}$$

## GARCH in Mean

- If investors are risk averse then expected return of risky (volatile) assets should be higher than the rate of return of stable assets (e.g. eturn on SP500 was on average 5% higher than from 3M TB)

- GARCH-M (GARCH in Mean, Engle, Lilien i Ronbins, 1987) :

$$r_t = \mu + \delta \sigma_t + \epsilon_t,$$
$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

- Alternative specifications

$$r_t = \mu + \delta \sigma_t^2 + \epsilon_t$$
$$r_t = \mu + \delta \ln \sigma_t + \epsilon_t$$

## Forecasting volatility with GARCH(1,1)

Variance forecast from GARCH model:

$$r_t = \mu + \epsilon_t, \qquad \epsilon_t \sim D(0, \sigma_t^2)$$
$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

Given information set $\Omega_T$, i.e. $\epsilon_{1:T}$ and $\sigma_{1:T}^2$, we can compute that:

$$\sigma_{T+1|T}^2 = \omega + \alpha\epsilon_T^2 + \beta\sigma_T^2$$

For futher horizons we need to notice that:

$$E\left(\epsilon_{T+h}^2\right) = \sigma_{T+h|T}^2$$

Hence:

$$\sigma_{T+2|T}^2 = \omega + (\alpha + \beta)\sigma_{T+1|T}^2$$

Notice that for $\alpha + \beta < 1$ the forecast converges towards:

$$\lim_{H\to\infty} \sigma_{T+H|T}^2 = \frac{\omega}{1 - (\alpha + \beta)}$$

## Simulating future returns form a GARCH

Steps to simulate a single path of returns over horizon $H$ from GARCH model:

$$r_t = \mu + \epsilon_t, \qquad \epsilon_t \sim D(0, \sigma_t^2)$$
$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

1. Given information set $\Omega_T$ calculate $\sigma_{T+1}^2$
2. Draw $\epsilon_{T+1}$ from distribution $D(0, \sigma_{T+1}^2)$
3. Calculate $r_{T+1}$
4. Conditional on the draw for $\epsilon_{T+1}$ calculate $\sigma_{T+2}^2$

... continue until you have the path for $r_1, r_2, \dots, r_H$

## Calculating VaR/ES with GARCH models

MC steps to calculate VaR/ES for any horizon $H$ from GARCH model:

1. Simulate a path $r_1, r_2, \dots, r_H$ of returns over horizon $H$ and cumulate $y_H = \sum_{h=1}^{H} r_h$
2. Repeat step 1 "$N$" Times. Save $y_H^{(n)}$ for $n = 1,2,\dots,N$
3. Sort cumulated returns $ys_H^{(1)} \leq ys_H^{(2)} \leq$ ...
4. Set $M = floor(pN)$
5. Use formulas :

$$VaR_H = ys_H^{(M)}$$

$$ES_H = \frac{1}{M}\sum_{1}^{M} ys_H^{(i)}$$

## Exercises

**Exercise 5.3.**

Let $r_t$ be weekly log-return (expressed as %) for a portfolio. The estimates of the GARCH(1,1) model as as follows:

$$\begin{aligned} r_t &= 0.08 + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_t^2) \\ \sigma_t^2 &= 0.025 + 0.10\epsilon_{t-1}^2 + 0.80\sigma_{t-1}^2 \end{aligned}$$

a. What is the average annual rate of return (assume that a year is 52 weeks)?
b. Calculate the unconditional variance (and standard deviation) for weekly data
c. Knowing that $\epsilon_T^2 = 0.15$ and $\sigma_T^2 = 0.4$ calculate the forecast $\sigma_{T+1|T}^2$
d. Select the tolerance level $p$ and calculate VaR and ES using the values from table below

| $p$ | 0.5 | 0.1 | 0.05 | 0.025 | 0.01 | 0.001 |
|---|---|---|---|---|---|---|
| VaR | 0 | 1.282 | 1.645 | 1.960 | 2.326 | 3.090 |
| ES | 0.798 | 1.755 | 2.063 | 2.338 | 2.665 | 3.367 |

## Exercises

**Exercise 5.4.**

Build an equally weighted portfolio consisting of two assets.

a. Make a graph of historical time series. If the history is shorter than 5 years, select other stocks
b. Select the tolerance level $p$
c. Calculate VaR/ES for horizons $H = 1$ and $H = 10$using parametric models (normal, t-Student); historical simulation; EWMA; GARCH(1,1) model
d. Fill in the table below

| | **VaR**<br>**H=1** | **H=10** | **ES**<br>**H=1** | **H=10** |
|---|---|---|---|---|
| Normal | | | | |
| t-Student | | | | |
| Historical simulation | | | | |
| EWMA | | | | |
| GARCH | | | | |

## Exercises

**Exercise 5.5.**

Build an equally weighted protfolio consisting of two assets.

Construct the best GARCH model:

a. Specify lags P and Q of GARCH(P,Q) as well as the error term distribution with the BIC criterion
b. Check for the autocorrelation of standardized residuals
c. Check for the leverage effect (GJR-GARCH / EGARCH)
d. Check for in-Mean effect
e. Calculate VaR/ES for horizons $H = 1$ and $H = 10$
f. Compare the results from point e to GARCH(1,1) – see Exercise 5.4
g. Calculate the forecast for standard deviation $\sigma_t$ at horizon $H = 1000$ and compare it to sample standard deviation for portfolio returns. Are the differences sizeable?

# Topic 6

# Multivariate portfolio. MGARCH models

- Direct generalizations of the univariate GARCH (VEC GARCH and BEKK)
- Linear combinations of univariate GARCH model (Factor-GARCH and GO-GARCH)
- Nonlinear combinations of univariate GARCH models (CCC-GARCH and DCC-GARCH)
- Calculating VaR and ES from MGARCH models

## MGARCH: general specification

Let $y_t$ be a vector of returns for individual assets entering the investment portfolio. For the joint distribution, let us assume that:

$$y_t = \mu_t + \epsilon_t$$
$$Cov(\epsilon_t) = H_t$$

where:

| | |
|---|---|
| $y_t = (y_{1t}, y_{2t}, \dots, y_{Nt})'$ | is the vector of returns |
| $\mu_t = (\mu_{1t}, \mu_{2t}, \dots, \mu_{Nt})'$ | is the conditional mean |
| $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{Nt})'$ | is the error term |
| $H_t = [h_{ij}]_{N\times\mathrm{N}}$ | is the conditional covariance matrix |

In MGARCH model we model the dynamics of $H_t$ as a function of:

- past values of the covariance matrix $H_{t-q}$ for $q = 1,2,\dots,Q$
- realization of the error term $\epsilon_{t-p}\epsilon'_{t-p}$ for $p = 1,2,\dots,P$

## MGARCH: classification

MGARCH models can be classified depending on the specification of the dynamics for the covariance matrix $H_t$ into (see Bauwens et al. 2006, JAE):

1. Direct generalizations of the univariate GARCH (VEC GARCH or BEKK)
2. Linear combinations of univariate GARCH model (GO-GARCH)
3. Nonlinear combinations of univariate GARCH models (DCC-GARCH)

## VEC-GARCH

VEC-GARCH(1,1) proposed by Bollerslev, Engle and Wooldridge (1988)

$$
\begin{aligned}
y_t &= \mu_t + \epsilon_t, \epsilon_t \sim N(0, H_t) \\
h_t &= \omega + Ae_{t-1} + Bh_{t-1} \\
h_t &= vech(H_t) \\
e_t &= vech(\epsilon_t \epsilon_t')
\end{aligned}
$$

where $vech(\cdot)$ denotes the operator that stacks the lower triangular portion of a $N \times N$ matrix as a $\frac{(N+1)N}{2} \times 1$ vector.

Problems:

- Large number of parameters: $A$ and $B$ are $\frac{(N+1)N}{2} \times \frac{(N+1)N}{2}$ matrices
- Difficulties in ensuring that $\mathrm{H}_t$ is positive definite

## VEC-GARCH

Bivariate example of VEC GARCH:

$$h_t = \omega + Ae_{t-1} + Bh_{t-1}$$

$$
\begin{bmatrix} h_{11,t} \\ h_{22,t} \\ h_{12,t} \end{bmatrix} = \begin{bmatrix} \omega_{11} \\ \omega_{22} \\ \omega_{12} \end{bmatrix} + \begin{bmatrix} \alpha_{11,1} & \alpha_{11,2} & \alpha_{11,3} \\ \alpha_{22,1} & \alpha_{22,2} & \alpha_{22,3} \\ \alpha_{12,1} & \alpha_{12,2} & \alpha_{12,3} \end{bmatrix} \begin{bmatrix} e_{11,t-1} \\ e_{22,t-1} \\ e_{12,t-1} \end{bmatrix} + \begin{bmatrix} \beta_{11,1} & \beta_{11,2} & \beta_{11,3} \\ \beta_{22,1} & \beta_{22,2} & \beta_{22,3} \\ \beta_{12,1} & \beta_{12,2} & \beta_{12,3} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{12,t-1} \end{bmatrix}
$$

where $e_{ij} = \epsilon_i \times \epsilon_j$

## DVEC-GARCH

To limit the number of parameters in VEC-GARCH model, Bollerslev et al (1988) proposed its restriction version, in which matrices $A$ and $B$ from

$$h_t = \omega + Ae_{t-1} + Bh_{t-1}$$

are assumed to be diagonal, so that equation changes into:

$$h_{ij,t} = \omega_{ij} + \alpha_{ij}\epsilon_{i,t-1}\epsilon_{j,t-1} + \beta_{ij}h_{ij,t-1}$$

Even though the number of parameters decreases, the problem of ensuring that $H_t$ is positive definite remains

Important: multivariate EWMA from Riskmetrics is calibrated DVEC-GARCH:

$$h_{ij,t} = (1-\lambda)\epsilon_{i,t-1}\epsilon_{j,t-1} + \lambda h_{ij,t-1}$$

## DVEC-GARCH

Bivariate example of DVEC GARCH:

$$h_t = \omega + Ae_{t-1} + Bh_{t-1}$$

$$\begin{bmatrix} h_{11,t} \\ h_{22,t} \\ h_{12,t} \end{bmatrix} = \begin{bmatrix} \omega_{11} \\ \omega_{22} \\ \omega_{12} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{12} \end{bmatrix} \begin{bmatrix} e_{11,t-1} \\ e_{22,t-1} \\ e_{12,t-1} \end{bmatrix} + \begin{bmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{22} & 0 \\ 0 & 0 & \beta_{12} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{12,t-1} \end{bmatrix}$$

where $e_{ij} = \epsilon_i \times \epsilon_j$

## BEKK-GARCH

Engle and Kroner (1995) proposed BEKK-GARCH model, in which $H_t$ is always positive definite

$$y_t = \mu_t + \epsilon_t, \epsilon_t \sim N(0, H_t)$$
$$H_t = \Omega\Omega' + A\epsilon_{t-1}\epsilon_{t-1}'A' + BH_{t-1}B'$$

where $\Omega$ is lower triangular matrix, whereas $A$ and $B$ are $N \times N$ matrices.

$$\begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} \omega_{11} & 0 \\ \omega_{12} & \omega_{22} \end{bmatrix} \begin{bmatrix} \omega_{11} & \omega_{12} \\ 0 & \omega_{22} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} e_{11,t-1} & e_{21,t-1} \\ e_{12,t-1} & e_{22,t-1} \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{21} \\ \alpha_{12} & \alpha_{22} \end{bmatrix}$$
$$+ \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} h_{11,t-1} & h_{21,t-1} \\ h_{12,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \end{bmatrix}$$

where $e_{ij} = \epsilon_i \times \epsilon_j$

Note: Due to a large number of params BEKK model is rarely used when $N > 3$

## Exercises

**Exercise 6.1.**

The BEKK GARCH model describing the dynamics of a bivariate vector $r = (r_1, r_2)'$ :

$r_t \sim N(0, H_t)$

$$\begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.4 & 0.3 \end{bmatrix} \begin{bmatrix} 0.5 & 0.4 \\ 0 & 0.3 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.0 \\ 0.1 & 0.2 \end{bmatrix} \begin{bmatrix} e_{11,t-1} & e_{12,t-1} \\ e_{12,t-1} & e_{22,t-1} \end{bmatrix} \begin{bmatrix} 0.2 & 0.1 \\ 0.0 & 0.2 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.0 \\ 0.5 & 0.6 \end{bmatrix} \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{12,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 \\ 0.0 & 0.6 \end{bmatrix}$$

a. Assuming that $\epsilon_{1,T} = 0$; $\epsilon_{2,T} = 0$; $h_{11,T} = 3$; $h_{12,T} = 1$; $h_{22,T} = 5$; make a forecast for $H_{T+1}$
b. Calculate the variance $\sigma^2_{T+1}$ of a portfolio with weights $w = (0.5, 0.5)'$
c. Calculate the $VaR_{5\%}$ of a portfolio with weights $w = (0.5, 0.5)'$ knowing that $\Phi^{-1}(0.05) = -1.64$
d. Repeat points b and c for $w = (0.25, 0.75)'$

## Factor-GARCH

Engle, Ng and Rothschild (1990) proposed a factor specification of MGARCH model, in which the dynamics of $H_t$ is described by $K$ factors:

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, H_t)$$

$$\epsilon_t = \Lambda f_t + \eta_t, \quad \eta_t \sim N(0, \Gamma), \qquad f_t \sim N(0, D_t),$$

$$\Gamma = \mathrm{diag}(\gamma_1^2, \gamma_2^2, \ldots, \gamma_N^2), \quad D_t = \mathrm{diag}(d_{1t}, d_{2t}, \ldots, d_{Kt})$$

$$H_t = \Lambda D_t \Lambda' + \Gamma$$

$$d_{kt} = \omega_k + \alpha_k f_{k,t-1}^2 + \beta_k d_{k,t-1}$$

**Note:** This specification allows to transform the problem of finding the dynamics for multidimentional matrix $H_t$ into a problem of finding the dynamics of $K$ univariate processes

## Factor-GARCH

Bivariate Factor-GARCH(1,1,2) with no idiosyncratic term

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, H_t)$$

$$\epsilon_t = \Lambda f_t, \qquad f_t \sim N(0, D_t),$$

$$d_{1t} = \omega_1 + \alpha_1 f_{1,t-1}^2 + \beta_1 d_{1,t-1}$$

$$d_{2t} = \omega_2 + \alpha_2 f_{2,t-1}^2 + \beta_2 d_{2,t-1}$$

$$\begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \begin{bmatrix} d_{1,t} & 0 \\ 0 & d_{2,t} \end{bmatrix} \begin{bmatrix} \lambda_{11} & \lambda_{21} \\ \lambda_{12} & \lambda_{22} \end{bmatrix}$$

## Exercises

**Exercise 6.2.**

On the basis of the below relationship:

$$\begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \begin{bmatrix} d_{1,t} & 0 \\ 0 & d_{2,t} \end{bmatrix} \begin{bmatrix} \lambda_{11} & \lambda_{21} \\ \lambda_{12} & \lambda_{22} \end{bmatrix}$$

write the formula for the value of conditional correlation in Factor-GARCH model as a function of $d_{1t}$ and $d_{2t}$. Is the dynamic for the calculated formula transparent?

Note that: $\rho_{12,t} = \frac{h_{12,t}}{\sqrt{h_{11,t}}\sqrt{h_{22,t}}}$

## GO-GARCH (Generalized Orthogonal)

Van der Weide (1990) proposed a specific verion of Factor-GARCH model, based on spectral decomposition of population cov. matrix, combined with rotation:

$$H \quad = PLP^{-1}, P \ -\text{eigenvectors matrix}, L - \text{eigenvalues matrix}$$
$$\Lambda \quad = PL^{0.5}U$$
$$UU' = I, \qquad U - \text{orthonormal, rotation matrix}$$

GO-GARCH model:

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, H_t)$$
$$\epsilon_t = \Lambda f_t, \qquad f_t \sim N(0, D_t),$$
$$D_t = \text{diag}(d_{1t}, d_{2t}, \dots, d_{Nt})$$
$$H_t = \Lambda D_t \Lambda'$$
$$d_{it} = \omega_i + \alpha_i f_{i,t-1}^2 + \beta_i d_{i,t-1}$$

## *CCC-GARCH (Constant Conditional Correlation)*

- Bollerslev (2002) proposed the CCC GARCH of the following form:

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, H_t)$$
$$H_t = D_t P D_t'$$
$$D_t = \mathrm{diag}(\sqrt{d_{1t}}, \sqrt{d_{2t}}, \dots, \sqrt{d_{Nt}})$$
$$d_{it} = \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i d_{i,t-1} \text{ for } i = 1,2,\dots,N$$

where $P$ is the unconditional correlation matrix

- E(xtended)CCC-GARCH by Jeantheau (1998)

$$d_{it} = \omega_i + \sum_{j=1}^{N} \alpha_{ij} \epsilon_{j,t-1}^2 + \sum_{j=1}^{N} \beta_{ij} d_{j,t-1} \text{ for } i = 1,2,\dots,N$$

## *DCC-GARCH (Dynamic Conditional Correlation)*

Engle (2002) proposed the DCC GARCH of the following form:

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, H_t)$$
$$H_t = D_t P_t D_t'$$
$$D_t = \mathrm{diag}(\sqrt{d_{1t}}, \sqrt{d_{2t}}, \dots, \sqrt{d_{Nt}})$$
$$d_{it} = \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i d_{i,t-1} \text{ for } i = 1,2,\dots,N$$

where $P_t$ is the conditional correlation matrix with the law of motion:

$$z_t = D_t^{-1} \epsilon_t$$
$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha z_{t-1} z_{t-1}' + \beta Q_{t-1}$$
$$P_t = (Q_t \odot I)^{-0.5} Q_t (Q_t \odot I)^{-0.5}$$

where $\odot$ is the Hadamard operator (element-by-element multiplication)

## Exercises

**Exercise 6.3.**

The CCC GARCH model describing the dynamics of a bivariate vector $r = (r_1, r_2)'$ is:

$$r_t = \begin{bmatrix} 0.8 \\ 1.2 \end{bmatrix} + \epsilon_t; \ \ \epsilon_t \sim N(0, H_t)$$
$$H_t = D_t \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} D_t'$$
$$\begin{bmatrix} d_{1t} \\ d_{2t} \end{bmatrix} = \begin{bmatrix} 0.9 \\ 1.1 \end{bmatrix} + \begin{bmatrix} 0.1 & 0.0 \\ 0.0 & 0.1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1}^2 \\ \epsilon_{2t-1}^2 \end{bmatrix} + \begin{bmatrix} 0.8 & 0.0 \\ 0.0 & 0.7 \end{bmatrix} \begin{bmatrix} d_{1,t-1} \\ d_{2,t-1} \end{bmatrix}$$

a. Assuming that $\epsilon_{1T}^2 = 23; \epsilon_{2T}^2 = 16; \ d_{1T} = 16; d_{2T} = 9$; make a forecast for $H_{T+1}$
b. Calculate the parameters of the distribution of returns a portfolio with weights $w = (0.25, 0.75)'$ for the period $T+1$
c. Calculate the $VaR_{2.5\%}$ of a portfolio with weights $w = (0.25, 0.75)'$ knowing that $\Phi^{-1}(0.025) = -1.96$
d. Calculate the equilibrium value of $H_t$?

## Exercises

**Exercise 6.4.**

For a protfolio consisting of two assets:

a. Make a graph of historical time series. If the history is shorter than 5 years, select other stocks
b. Select the tolerance level $p$
c. Estimate GO-GARCH(1,1) and DCC-GARCH(1,1) model. Which is better fitted to the data?
d. Make a graph of conditional std. dev. for two vars. from both models (one chart per. variable)
e. Make o graph of conditional correlation dynamics from both models (one chart)
f. Make a graph of conditional std. dev. of a portfolio from both models (one chart)
g. Calculate VaR/ES for horizons $H = 1$ and $H = 10$ and fill in the table below

| | **VaR** | | **ES** | |
|---|---|---|---|---|
| | **H=1** | **H=10** | **H=1** | **H=10** |
| GO-GARCH | | | | |
| DCC-GARCH | | | | |

# Topic 7

# Multivariate portfolio. Copulas

- Multivariate normal distribution
- Non-linear dependencies
- Copula function: an intuition
- Sklar theorem
- Empirical copula
- Elliptic copulas
- Archimedean copulas
- Kendall $\tau$ correlation
- Fitting copulas to the data
- Calculating VaR and ES using copulas

## Multivariate normal distribution

- In many applications it is convenient to assume that multivariate returns have multivariate normal distribution:

$$r \sim N(\mu, \Sigma)$$

- For a portfolio of assets with weights $w$ the rate of return:

$$r_p = w' r \sim N\left(\mu_p, \sigma_p^2\right)$$

where $\mu_p = w'\mu$ and $\sigma_p^2 = w'\Sigma \mathrm{w}$

## Multivariate normal distribution

- Multivariate normal distribution implies that the relationship between variables $Y$ and $X$ is linear

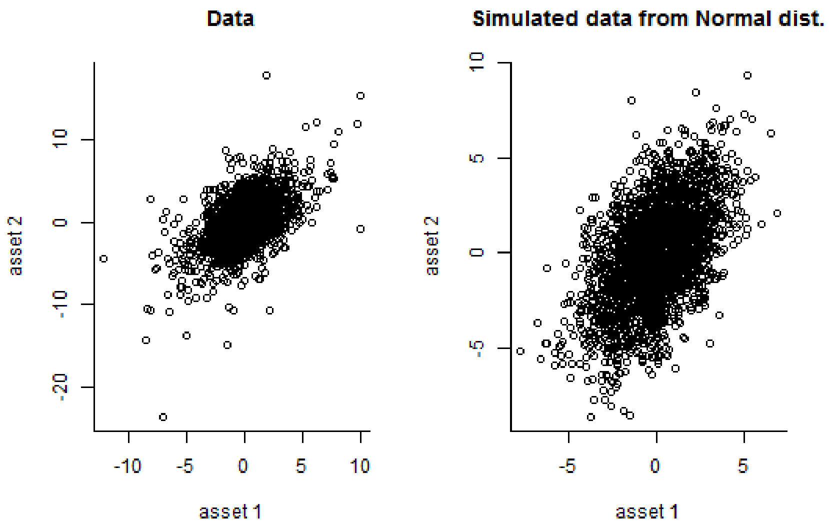
$$Y = a + bX + \epsilon, \qquad b = \frac{cov(Y,X)}{var(X)}, \qquad a = \bar{Y} - b\bar{X}$$

- In other words, the relationship is always the same and does not depend on the scale of change …

- … but at financial markets dependences tend to be stronger during crashes that in normal times, which leads to risk undervaluation!!!

## Non-linear dependencies: illustration

The scale of nonlinearity of dependencies between two variables can be illustrated by comparing a scatter plot of realized and simulated series




## Copula: a general idea

- Copula as a convenient method to model non-linear dependencies between variables $X$ and $Y$ (we will discuss only a bivariate case)
- A general idea is to decompose a (sophisticated) joint distribution of $X$ and $Y$ into:
  - Univariate marginal distributions for $X$ and $Y$
  - Copula function, which combines both marginal distributions

## Copula function

**SKLAR THEOREM:**

For:
$H(X,Y)$ multivariate/bivariate joint cdf
$F(X)$ and $G(Y)$ univariate marginal cdf
there exists a copula $C()$ for which:

$$H(X,Y) = C\big(F(X), G(Y)\big)$$

If $F$ and $G$ are continuous, then $C$ is unique

## Copula function: notation

$f(X); g(Y)$: pdf of marginal distributions
$U = F(X); V = G(Y)$: cdf of marginal distributions
$h(X,Y)$: pdf of joint distribution
$H(X,Y)$: cdf of joint distribution
$C(U,V)$: copula function

$$H(X,Y) = C(F(X), G(Y)) = C(U,V)$$

$$h(X,Y) = f(X)g(Y)C(F(X), G(Y))$$

To draw from the joint distribution, we need to decide on:

- the shape of marginal distributions $F(X)$ and $G(Y)$
- the shape of copula function $C(U,V)$

# Most popular copulas

## *Empirical copula*

- Let $\hat{F}$ and $\hat{G}$ be the empirical distribution functions for $x_t$ and $y_t$:
$$\hat{F}(x) = \#(x_t < x)/T \text{ and } \hat{G}(y) = \#(y_t < y)/T$$

  and $u_t = \hat{F}(x_t)$ and $v_t = \hat{G}(y_t)$, where $t = 1, 2, \dots, T$

- Definition of empirical copula:
$$C(u, v) = \frac{\#(u_t \leq u \ \wedge v_t \leq v)}{T}$$

- (discrete) probability density of empirical copula is:
$$c(u, v) = \frac{\#(u_t = u \ \wedge v_t = v)}{T}$$

where $\#(z)$ is the number of observations $t$ that fulfill condition $z$

## Exercises

**Exercise 7.1.**

For a sample of observations:

| $t$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $x$ | 1 | 2 | 3 | 4 | 6 |
| $y$ | 7 | 5 | 1 | 3 | 4 |

calculate:

- The values of $u_t$ and $v_t$ for $t = 1{,}2{,}3{,}4{,}5$
- Empirical copula
- Density of empirical copula

**Exercise 7.2.**

Let $(X, Y)$ be the random variables describing the outcome of rolling two dices.

- What is the marginal pdf/cdf for $X$ and $Y$?
- What is the joint pdf for $(X, Y)$
- What is the density copula for $(X, Y)$
- Roll two dices 10 times to create your sample for $(x, y)$ and calculate the empirical copula / density of empirical copula. Use function `x <- sample(1:6,10,replace=TRUE)`

## Elliptic copulas

**Normal copula:**

$$C(U,V) = \Phi_\Sigma(\Phi^{-1}(U), \Phi^{-1}(V))$$

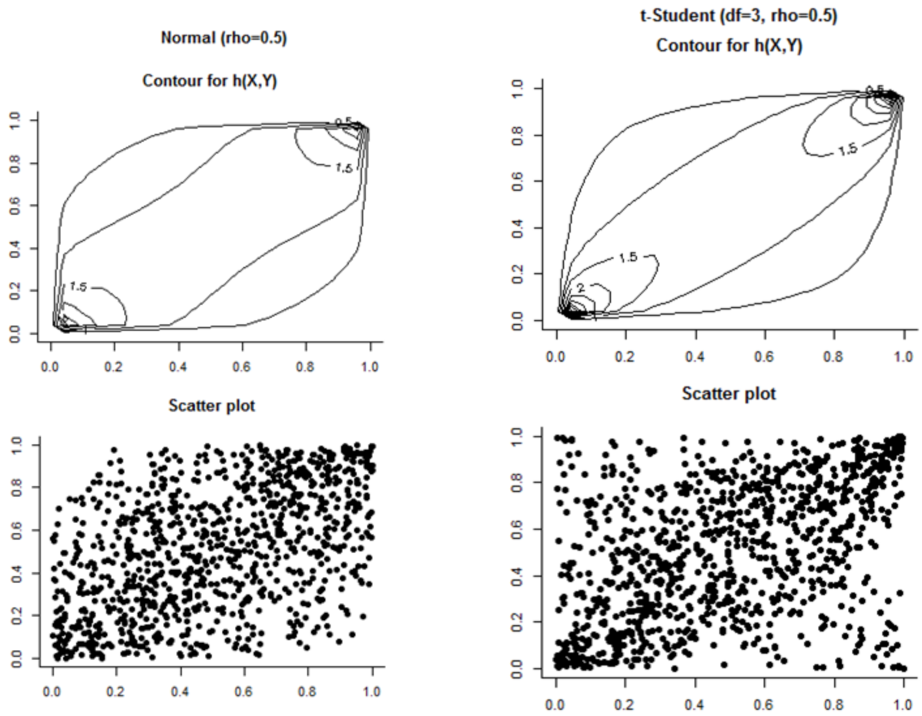
where $\Phi/\Phi_\Sigma$ is univariate/multivariate normal cdf and $\Sigma$ is the covariance matrix

**t-Student copula:**

$$C(U,V) = \mathrm{T}_{v,\Sigma}(\mathrm{T}_v^{-1}(U), \mathrm{T}_v^{-1}(V))$$

where $\mathrm{T}_v/\mathrm{T}_{v,\Sigma}$ is univariate/multivariate t-Student cdf with $v$ degrees of freedom

## Elliptic copulas




## Archimedean copulas

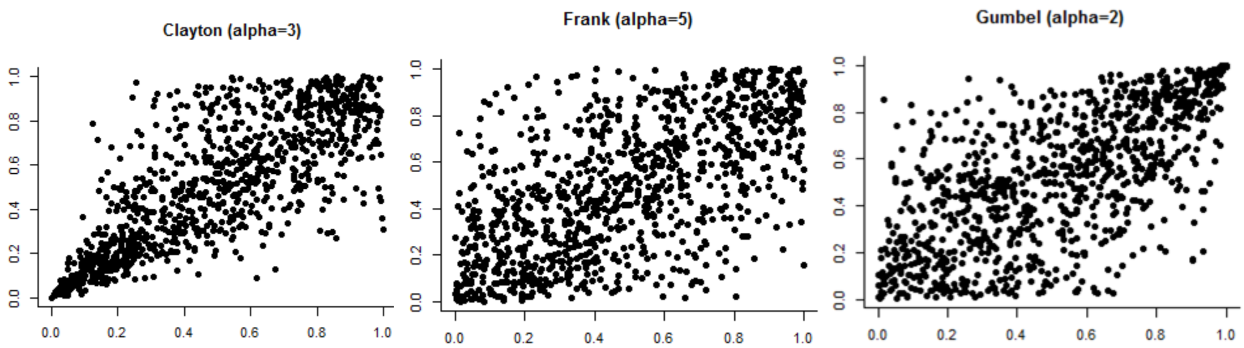
The formula for copula:

$$C(U,V) = \varphi^{-1}(\varphi(U) + \varphi(V))$$

where $\varphi$ is the generator

| copula | $C(u,v)$ | $\varphi(t)$ | $\varphi^{-1}(t)$ |
|---|---|---|---|
| Clayton | $C(u,v) = \left(\max\left\{u^{-\theta} + v^{-\theta} - 1; 0\right\}\right)^{-1/\theta}$ | $\varphi(t) = (1+t)^{-1/\theta}$ | $\varphi^{-1}(t) = t^{-\theta} - 1$ |
| Gumbel | $C(u,v) = \mathrm{e}^{-((-\log(u))^{\theta} + (-\log(v))^{\theta})^{1/\theta}}$ | $\varphi(t) = \mathrm{e}^{(-t^{1/\theta})}$ | $\varphi^{-1}(t) = (-\log(t))^{\theta}$ |
| Frank | $C(u,v) = \frac{-1}{\theta}\log\left(1 + \frac{(\mathrm{e}^{-\theta u}-1)(\mathrm{e}^{-\theta v}-1)}{\mathrm{e}^{-\theta}-1}\right)$ | $\varphi(t) = \frac{-1}{\theta}\log(\mathrm{e}^{-t}(\mathrm{e}^{-\theta}-1)+1)$ | $\varphi^{-1}(t) = -\log\left(\frac{\mathrm{e}^{-\theta t}-1}{\mathrm{e}^{-\theta}-1}\right)$ |

## Archimedean copulas




## Tail dependence

- Copulas are useful tool for panic/crisis scenarios, in which the dependencies between assets tend to be stronger
- A useful measure for these scenarios is downward:

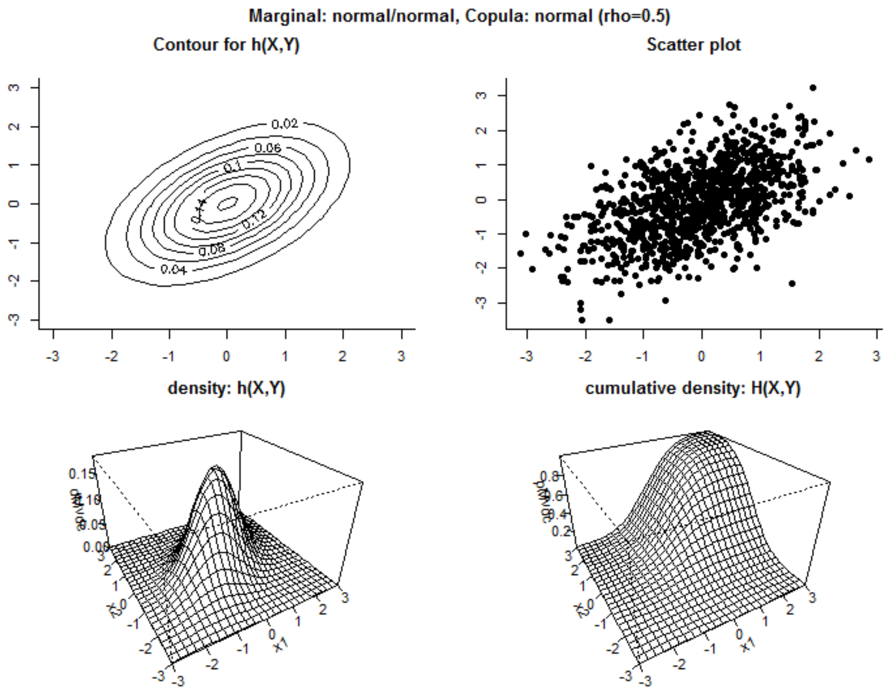
$$\lambda^L = \lim_{q \to 0} \frac{C(q, q)}{q}$$

or upward tail dependence:

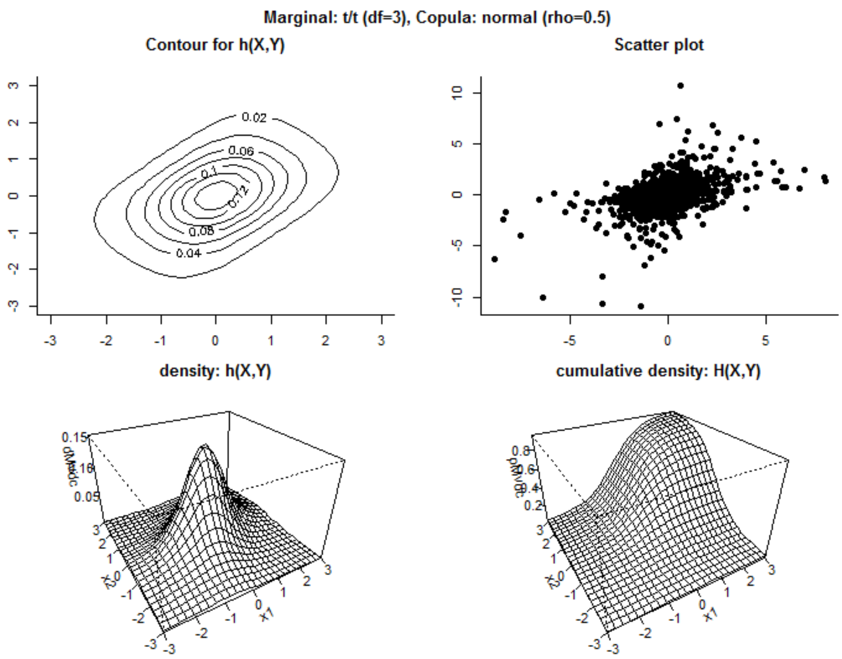
$$\lambda^U = \lim_{q \to 0} \frac{1 - C(1-q, 1-q)}{q}$$

- The values for copulas:
  - Normal / t-Student: $\lambda^L = 0$ $\lambda^U = 0$ if $\rho < 1$
  - Clayton: $\lambda^L = 2^{-\frac{1}{\theta}}$ $\lambda^U = 0$
  - Gumbel: $\lambda^L = 0$ $\lambda^U = 2 - 2^{\frac{1}{\theta}}$
  - Frank: $\lambda^L = 0$ $\lambda^U = 0$

## *Joint distribution from copulas*




## *Joint distribution from copulas*



# Fitting to the data

## *Kendall tau correlation*

**Definition:**

$$\tau = \mathrm{P}\{(x_i - x_j)(y_i - y_j) > 0\} - \mathrm{P}\{(x_i - x_j)(y_i - y_j) < 0\}$$

**Sample estimate:**

$$\hat{\tau} = \frac{P - Q}{N(N-1)/2}$$

$P$ – number of concordant pairs: $\mathrm{N}\{(x_i - x_j)(y_i - y_j) > 0\}$

$Q$ – number of discordant pairs: $\mathrm{N}\{(x_i - x_j)(y_i - y_j) < 0\}$

## Fitting copula to the data: method of moments

For copula $C(U,V|\theta)$ we are looking parameter $\theta$ for which:

$$\tau_\alpha = 4\int\int C(U,V|\theta)dC(U,V|\theta) - 1$$

is closest to $\hat{\tau}$

**Formulas:**

Normal copula / t-Student: $\rho = \sin(\frac{\pi}{2}\tau)$

Clayton copula: $\theta = 2\tau(2-\tau)^{-1}, \ \ \theta > 0, \tau > 0$

Gumbel copula: $\theta = (1-\tau)^{-1} \quad , \ \ \theta \geq 0, \tau \geq 0$

Frank copula: $\tau = 1 - \frac{4}{\theta}\left[1 - \int_0^1 \frac{\ln(1-t^\theta)(1-t^\theta)}{t^{\theta-1}} dt\right]$

## Exercises

**Exercise 7.3.**

For a sample of observations:

| $t$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $x$ | 1 | 2 | 3 | 4 | 6 |
| $y$ | 3 | 5 | 1 | 6 | 7 |

calculate:

- Kandal tau
- Estimate of the parameter for normal copula
- Estimate of the parameter for Clayton copula
- Estimate of the parameter for Gumbel copula

## *Fitting copula to the data: maximum likelihood*

**One-step procedure (full ML)**

We are looking for parameter $\theta$ which maximizes

$$l(\theta|X,Y) = \sum_{t=1}^{T} \log c[F(x_t|\,\theta), G(y_t|\,\theta)] + \log f(x_t|\theta) + \log g(y_t|\theta)$$

**Two-step procedure**

Step 1: Estimate marginal distribution parameters

Step 2: Estimate copula function parameters

## *VaR and ES from copula (over horizon H)*

1. Draw $(u, v)$ from $C(U, V)$
2. Calculate $x = F^{-1}(u)$ and $y = G^{-1}(v)$
3. Calculate the simulate rate of return of the portfolio $r = w_x x + w_y y$
4. Repeat steps 1-3 $H$ times to simulate a path $r_1, r_2, \dots, r_H$ of returns over horizon $H$ and calculate $y_H = \sum_{h=1}^{H} r_h$
5. Repeat steps 1-4 $N$ times. Save $y_H^{(n)}$ for $n = 1,2, \dots, N$
6. Sort cumulated returns $ys_H^{(1)} \leq ys_H^{(2)} \leq \dots$
7. Set $M = floor(pN)$
8. Use formulas :

$$VaR_H = ys_H^{(M)}$$

$$ES_H = \frac{1}{M}\sum_{1}^{M} ys_H^{(i)}$$

## Exercises

**Exercise 7.4.**

For a protfolio consisting of 2 assets.

a. Make a graph of historical time series. If the history is shorter than 5 years, select other stocks
b. Select the tolerance level $p$
c. Estimate 5 copulas listed in the table below. Which is the best fitted to the data?
d. Calculate VaR/ES for horizons $H = 1$ and $H = 10$
e. Fill in the table below

| | **VaR** | | **ES** | |
|---|---|---|---|---|
| | **H=1** | **H=10** | **H=1** | **H=10** |
| Normal | | | | |
| t-Student | | | | |
| Clayton | | | | |
| Gumbel | | | | |
| Frank | | | | |

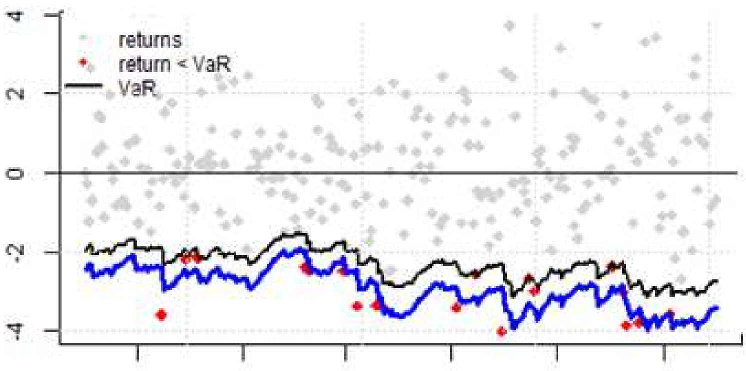
# Topic 8

# Backtesting risk models

- Backtesting settings
- Binomial distribution
- Traffic lights approach
- Conditional coverage and independence of VaR exceedances
- Kupiec test
- Christoffersen tests
- McNeil-Frey test

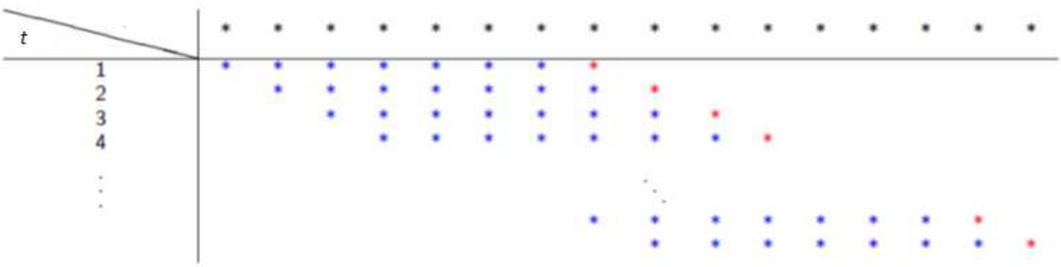
## What is backtesting

- ”backtesting” in finance = ”out-of-sample evaluation” in economics
- backtesting allows to assess model performance if it was used in the past
- for VaR we compare the share of VaR exceedances to tolerance level
- for ES we check if the scale of exceedances is correctly calibrated




## Backtesting procedure for VaR

1. Set observation for the start backtesting $T^*$(usually $T-250$)
2. Use data until period $t=T^*$to calculate VaR for period $t+1$: $VaR_{t+1|t}$
3. Compare $VaR_{t+1|t}$ to realization $r_{t+1}$ to assess if VaR was exceeded
4. Repeat steps 2 and 3 for $t=T^*+1, T^*+2, \ldots, T-1$




Note: We used a similar procedure for out-of-sample forecast evaluation in Block 1

## VaR exceedances

- Using the series $VaR_{t|t-1}$ and $r_t$ for $t = T^* + 1, T^* + 2, \dots, T$ we can construct the series of exceedances

$$e_t = \begin{cases} 1 \text{ if } r_t \leq VaR_{t|t-1} \\ 0 \text{ if } r_t > VaR_{t|t-1} \end{cases}$$

- And calculate the number of exceedances ($n_1$) / no exceedances ($n_0$)

$$n_1 = \sum_{t=T^*+1}^{T} e_t$$
$$n_0 = n - n_1$$

where $n = T - T^*$ is the number of observations with which we evaluate VaR

## Distribution for the number of VaR exceedances

- How many exceedances should we expect?
- For a well specified model $e_t$ should be IID with:

$$e_t \sim B(1,p)$$
$$n_1 \sim B(n,p)$$

where $B(n,p)$ is binomial distribution with $n$ trials and probability $p$

Distribution for $n_1$ if $n = 250$

| | $p = 5\%$ | | $p = 1\%$ | |
|---|---|---|---|---|
| $n_1$ | pdf | cdf | pdf | cdf |
| 0 | 0.0 | 0.0 | 8.1 | 8.1 |
| 1 | 0.0 | 0.0 | 20.5 | 28.6 |
| 2 | 0.0 | 0.0 | 25.7 | 54.3 |
| 3 | 0.1 | 0.1 | 21.5 | 75.8 |
| 4 | 0.3 | 0.5 | 13.4 | 89.2 |
| 5 | 0.9 | 1.3 | 6.7 | 95.9 |
| 6 | 1.8 | 3.1 | 2.7 | 98.6 |
| 8 | 5.4 | 11.9 | 0.3 | 99.9 |
| 10 | 9.6 | 29.1 | 0.0 | 100.0 |
| 12 | 11.6 | 51.8 | 0.0 | 100.0 |
| 14 | 10.0 | 72.9 | 0.0 | 100.0 |
| 16 | 6.4 | 87.5 | 0.0 | 100.0 |
| 18 | 3.1 | 95.3 | 0.0 | 100.0 |
| 20 | 1.2 | 98.5 | 0.0 | 100.0 |

## *Basel Committee „Traffic lights" approach*

| Zone | Exceedances | Plus Factor $k$ | cdf |
|---|---|---|---|
| ■ | 0 | 0.00 | 8.11 |
| ■ | 1 | 0.00 | 28.58 |
| ■ | 2 | 0.00 | 54.32 |
| ■ | 3 | 0.00 | 75.81 |
| ■ | 4 | 0.00 | 89.22 |
| ■ | 5 | 0.40 | 95.88 |
| ■ | 6 | 0.50 | 98.63 |
| ■ | 7 | 0.65 | 99.60 |
| ■ | 8 | 0.75 | 99.89 |
| ■ | 9 | 0.85 | 99.97 |
| ■ | 10+ | 1.00 | 99.99 |

***Quantitative standards Basel II***

e. The multiplication factor will be set by individual supervisory authorities on the basis of their assessment of the quality of the bank's risk management system, subject to an absolute minimum of 3. Banks will be required to add to this factor a "plus" directly related to the ex-post performance of the model, thereby introducing a built in positive incentive to maintain the predictive quality of the model. The plus will range from 0 to 1 based on the outcome of so-called "backtesting."

## *Exercises*

**Exercise 8.1.**

A $VaR_p$ model was evaluated with a backtest using $n$ observations. Let $\pi = n_1/n$ be the share of VaR exceedances, where $n_1$ is the number of VaR exceedances.

Calculate the 95% interval (left tailed and centered) for $n_1$ and $\pi$ using `dbinom/pbinom/qbinom` functions in R, assuming that the VaR model is well specified and that:

a. $\mathrm{n} = 250, \mathrm{p} = 1\%$
b. $\mathrm{n} = 250, \mathrm{p} = 5\%$
c. $\mathrm{n} = 100, \mathrm{p} = 5\%$
d. $\mathrm{n} = 100, \mathrm{p} = 5\%$

Discuss the results

## Backtesting: what we verify?

- For a model $e_t = \alpha + v_t$ we test for **unconditional coverage:**

$$H_0: \alpha = p$$

- For a model $e_t = \alpha + \rho e_{t-1} + v_t$ we test for **independence:**

$$H_0: \rho = 0$$

and **unconditional coverage:**

$$H_0: \alpha = p \wedge \rho = 0$$

**Why shouldn't we use LS regression to test the above hypotheses?**

## Kupiec test: unconditional coverage

- Let $e_t$ are $IID\ B(\alpha)$ so that the likelihood of $\alpha$ given $n_1$ exceedeances in sample $n = n_0 + n_1$ is:

$$\mathcal{L}(\alpha|n_0, n_1) = \binom{n_0 + n_1}{n_1} \alpha^{n_1}(1-\alpha)^{n_0}$$

- The formula for ML estimator:

$$\pi = \hat{\alpha} = n_1/n$$

- We can test the null of **unconditional coverage (Kupiec) test**:

$$H_0: \alpha = p$$

By calculating the likelihood ratio:

$$LR_{UC} = \frac{\mathcal{L}(p|n_0, n_1)}{\mathcal{L}(\pi|n_0, n_1)} = \frac{p^{n_1}(1-p)^{n_0}}{\pi^{n_1}(1-\pi)^{n_0}}$$

and the likelihood ratio test statistic:

$$-2\ln(LR_{UC}) \sim \chi^2(1)$$

## Christoffersen test: independence

- Let's assume that the distribution of $e_t$ depends on history:

$$e_t \sim \begin{cases} B(0,\alpha_0) \text{ if } e_{t-1} = 0 \\ B(0,\alpha_1) \text{ if } e_{t-1} = 1 \end{cases}$$

- The likelihood for $i \in \{0,1\}$ is:

$$\mathcal{L}_i(\alpha_i | n_{i0}, n_{i1}) = \binom{n_{i0} + n_{i1}}{n_{i1}} \alpha_i^{n_{i1}} (1-\alpha_i)^{n_{i0}}$$

where $n_{ij} = \#(e_{t-1} = i \wedge e_t = j)$.

- ML estimator of $\alpha_i$:

$$\pi_i = \hat{\alpha}_i = n_{i1}/(n_{i0}+n_{i1})$$

## Christoffersen test: independence

- The null of **independence Christofersen test**:

$$H_0\colon \alpha_0 = \alpha_1 = \alpha$$

- Under the null the ML estimate for a single probability is :

$$\pi = \hat{\alpha} = (n_{01}+n_{11})/(n_{00}+n_{01}+n_{10}+n_{11})$$

- The likelihood ratio is:

$$LR_{IND} = \frac{\mathcal{L}_0(\pi | n_{00}, n_{01}) \times \mathcal{L}_1(\pi | n_{10}, n_{11})}{\mathcal{L}_0(\pi_0 | n_{00}, n_{01}) \times \mathcal{L}_1(\pi_1 | n_{10}, n_{11})} = \frac{\pi^{(n_{01}+n_{11})}(1-\pi)^{(n_{00}+n_{10})}}{\pi_0^{n_{01}}(1-\pi_0)^{n_{00}} \times \pi_1^{n_{11}}(1-\pi_1)^{n_{10}}}$$

with the likelihood ratio test statistic:

$$-2\ln(LR_{IND}) \sim \chi^2(1)$$

## Christoffersen test: conditional coverage

- **Conditional coverage Christofersen test** is a joint test of of unconditional coverage and independence
- For the null of the test:

$$H_0 : \alpha_0 = \alpha_1 = p$$

the likelihood ratio is:

$$LR_{CC} = \frac{\mathcal{L}_0(p|n_{00}, n_{01}) \times \mathcal{L}_1(p|n_{10}, n_{11})}{\mathcal{L}_0(\pi_0|n_{00}, n_{01}) \times \mathcal{L}_1(\pi_1|n_{10}, n_{11})} = \frac{p^{(n_{01}+n_{11})}(1-p)^{(n_{00}+n_{10})}}{\pi_0^{n_{01}}(1-\pi_0)^{n_{00}} \times \pi_1^{n_{11}}(1-\pi_1)^{n_{10}}}$$

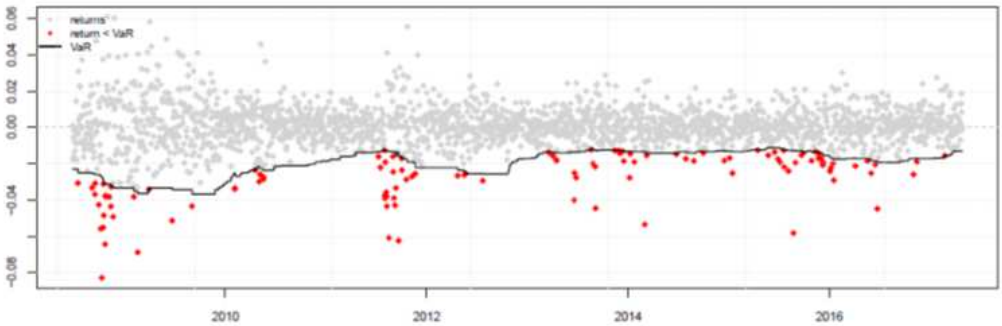
$$LR_{CC} = LR_{IND} \times LR_{UC}$$

with the likelihood ratio test statistic:

$$-2\ln(LR_{CC}) \sim \chi^2(2)$$

## Backtesting - illustration

- Number of observations: $n = 2500$
- Expected number of exceedances: $np = 125$
- Realized number of exceedances: $n_1 = 124$
- Kupiec UC test decission: $H_0$
- Christofersen CC test decission $H_1$




**How can we improve the VaR model?**

## Backtesting ES: McNeila and Frey test

- Let $\tau = t_1, t_2, \ldots, t_{n_1}$ be the periods of VaR exceedances ($r_\tau < VaR_{\tau|\tau-1}$). Given the definition of ES:

$$ES_{\tau|\tau-1} = E\left(r_\tau \middle| r_\tau < VaR_{\tau|\tau-1}\right)$$

  for a well specified ES model the variable:

$$z_\tau = \frac{r_\tau - ES_{\tau|\tau-1}}{\sigma_{\tau|\tau-1}}$$

  should have $D(0,1)$, where $\sigma_{\tau|\tau-1}$ is the conditional standard deviation
- The null of **McNeil and Frey test**:

$$H_0 \colon E(z_t) = 0$$

  can be thereby verified with the standard $t$ test (or bootstrapped version):

$$t = \frac{\bar{z}}{\hat{\sigma}_z / \sqrt{n_1}} \sim t_{v=n_1-1}$$

## Exercises

**Exercise 8.2.**

Build a portfolio consisting of two assets. Backtest risk models for this portfolio in the following steps:

a. Make a graph of historical time series.
b. Select tolerance level ($p = 5\%$ or $p = 1\%$) and evaluation sample ($n = 250$)
c. Calculate the share of VaR exceedances for univariate models (nomal, HS, EWMA)
d. Backtest univariate models with Kupiec / Christofersen / McNnail-Frey tests
e. [Difficult] Try to perform points C and D for multivariate normal and compare with univariate normal. Are the results the same?
f. [Difficult] Try to perform points C and D for more sophisticated methods (GRACH, MGARCH, Copula)

# Block 2 presentation

Select a portfolio of two assets (stocks, exchange rates, commodities)

**a. <1.0p>** Describe the characteristics of returns from the portfolio. To show: time series plot, moments, QQ plot, ACF, ACF of squares

**b. <2.0p>** Estimate the best univariate GARCH model. To show: parameter estimates, conditional standard deviation, model selection methods

**c. <1.0p>** Estimate the best copula. To show: comparison of copulas (LL values), simulation from copula vs realizations (scatter plots)

**d. <3.0p>** (Single) table with VaR/ES for $H = 1$ and $H = 10$ and $p$ =1% and 5% using the following methods

- Historical simulation / t-Student / Normal / EWMA / GARCH
- GO-GARCH / DCC-GARCH
- Copula Eliptic / Archimeadean

**e. <3.0p>** Backtesting with Kupiec / Christoffersen / Frey-McNail (a table whether model passed the test + number of exceedances) for

- Historical simulation and EWMA
- GARCH
- Copula / MGARCH

Additionally, I attribute up to 2p for the quality of the presentation (1p. for the .pdf and 1p. for the speech / interpretation of the results). Presentation should take around 7 minutes.