The R Project and Applied Statistics

Harald Schmidbauer & Vehbi Sinan Tunalıoğlu

5. LİNUX VE ÖZGÜR YAZILIM ŞENLİĞİ

Ankara, 11-14 Mayıs 2006

Before we begin. . .

What are we going to present?

This presentation has two parts.

I. Some statistical background, together with several applications, without a focus on the command structure of R.

II. An applied part, including a brief introduction to R and the R programming environment — in a live session!

1. Statistics and R

Statistics is the science of reasoning with numbers.

Statistics is concerned with

- detecting the structure in data sets.
- facilitating the communication between people.
- making well-founded decisions.
- forecasting the future.
- determining what (numerical) information is needed to solve a given problem.

1. Statistics and R

- Although some techniques can be done using paper and pencil, statistics is a hi-tech science: It needs powerful software to be effective.
- We recommend: R. Please visit:

www.R-project.org

- R is a language and environment for statistical computing and graphics. It is a GNU project with contributors from all over the world.

1. Statistics and R

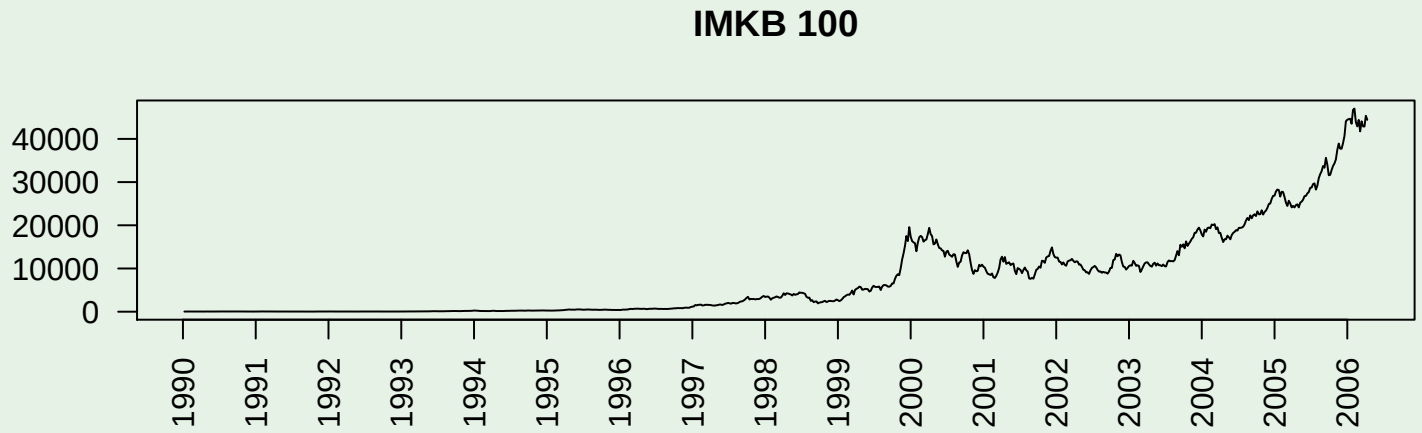
Advantages of R:

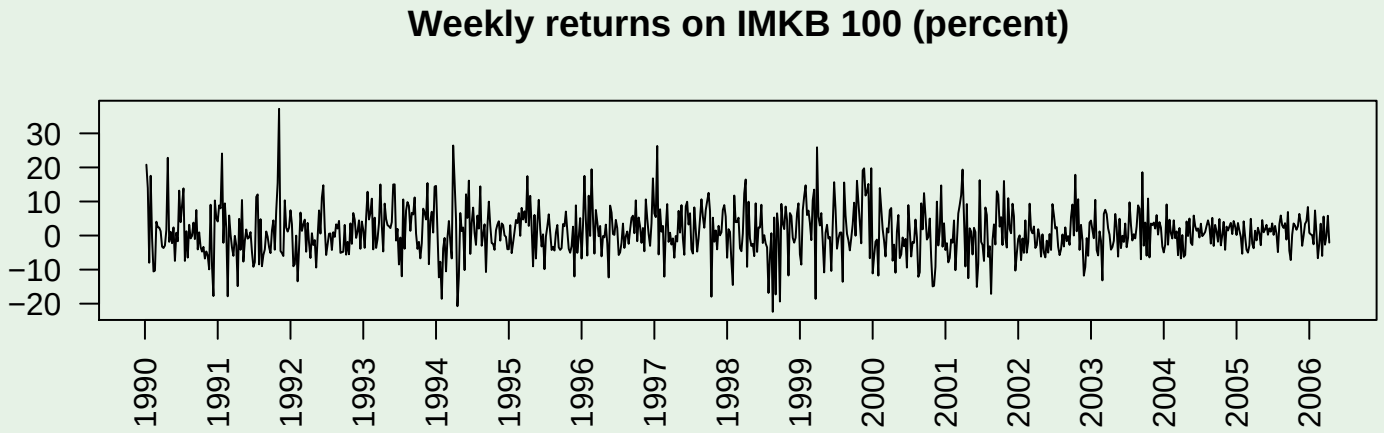
- extremely powerful, very professional
- constantly new functionality added
- GNU, open-source
- fairly easy to contribute
- support from user community

Disadvantages of R:

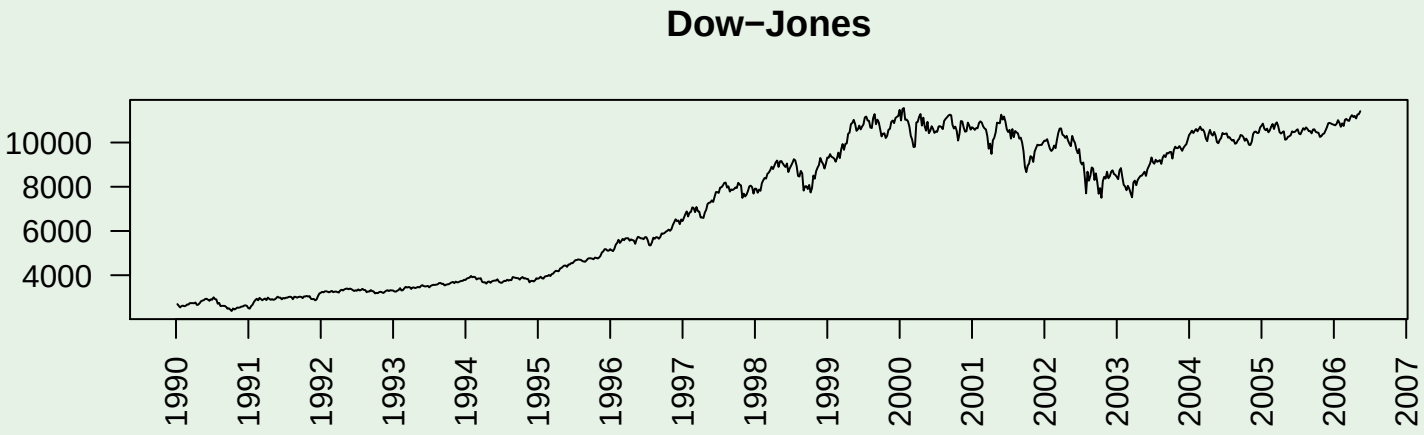
- no guarantee it works, but. . .
- no GUI, but. . .

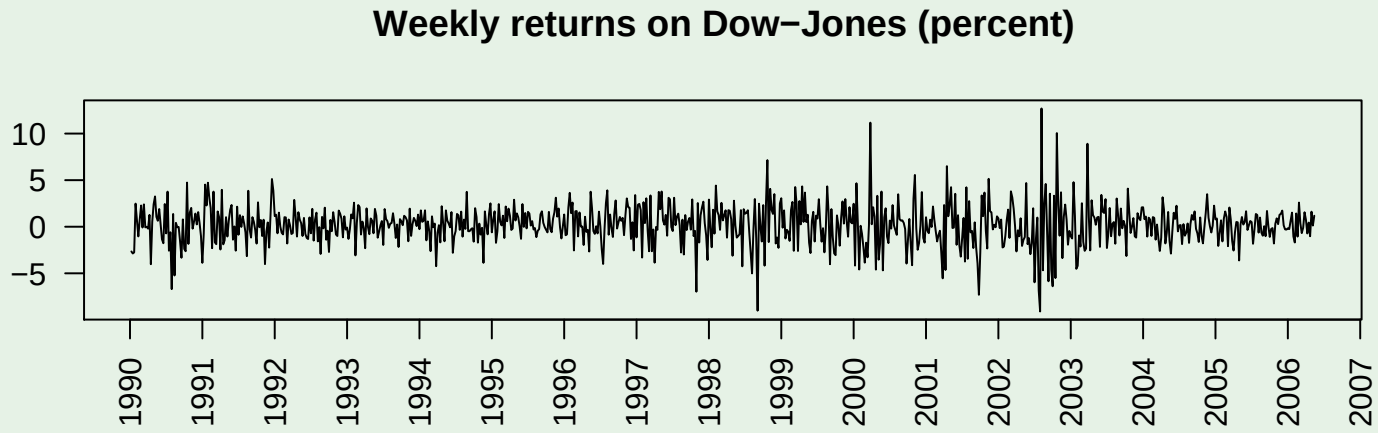
2. An Elementary Example



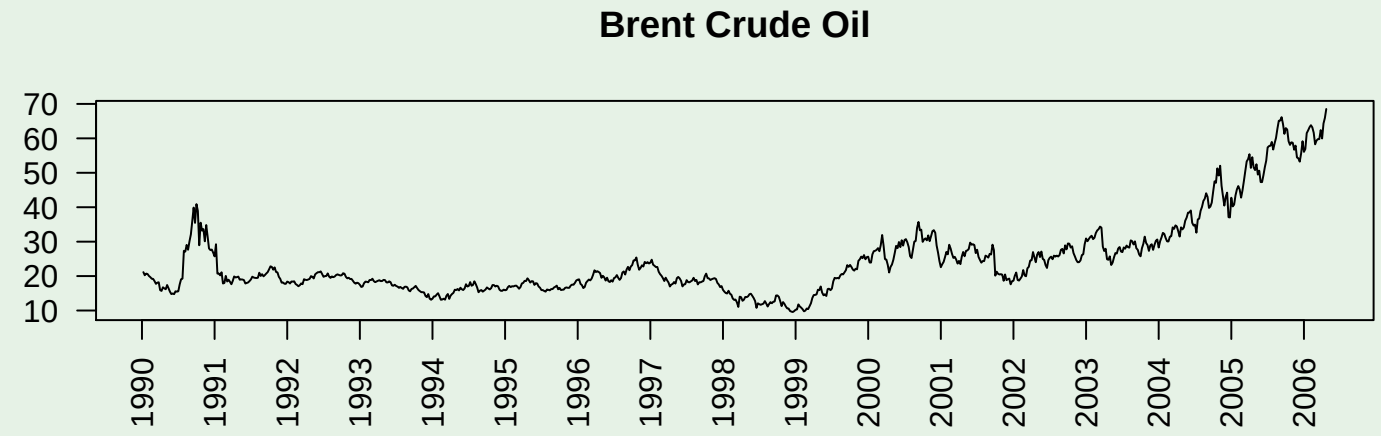


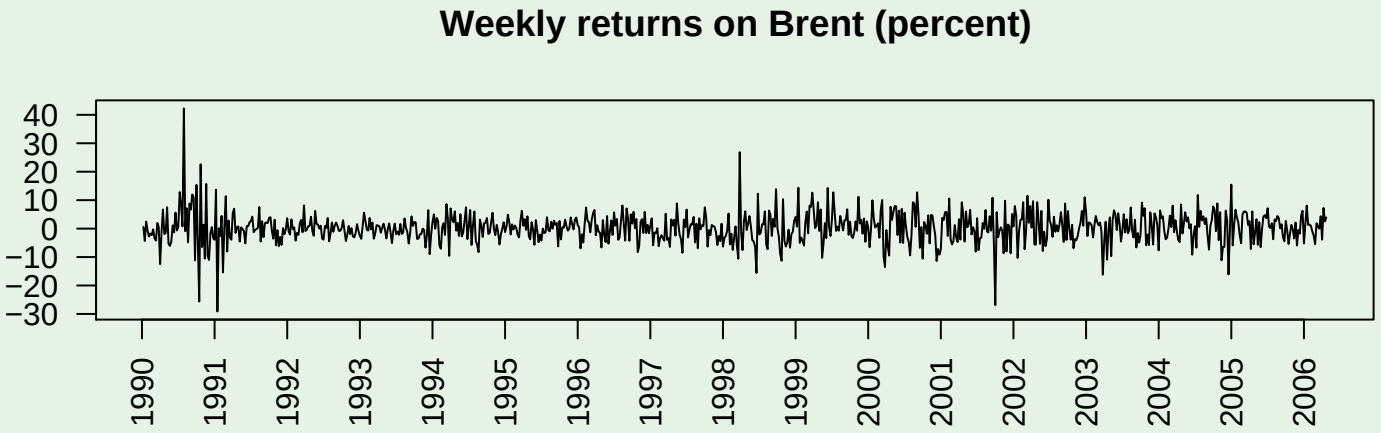
2. An Elementary Example (ctd.)





2. An Elementary Example (ctd.)





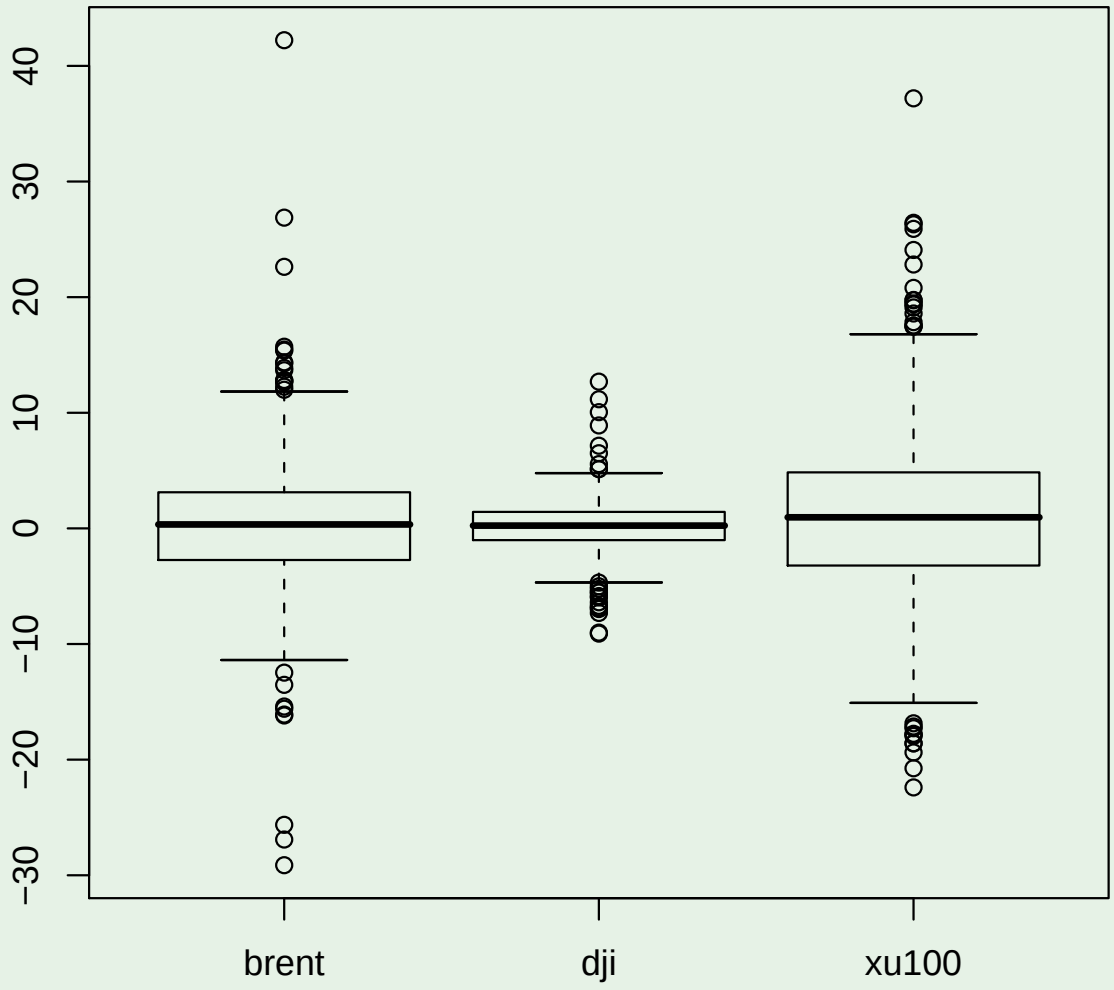
2. An Elementary Example (ctd.)

	xu100	dji	brent
first day	1990-01-09	1990-01-09	1990-01-09
last day	2006-05-02	2006-05-02	2006-04-11
observations	831	852	849
NAs	21	0	0
mean	1.16456	0.18826	0.28539
std error	0.24054	0.07570	0.19264
var	50.09061	4.73443	29.51176
std deviation	7.07747	2.17587	5.43247
skewness	0.35404	0.10626	0.32862
std error	0.17780	0.27521	0.52050
kurtosis	1.60865	3.36157	6.80703
std error	0.56330	0.87830	2.77051

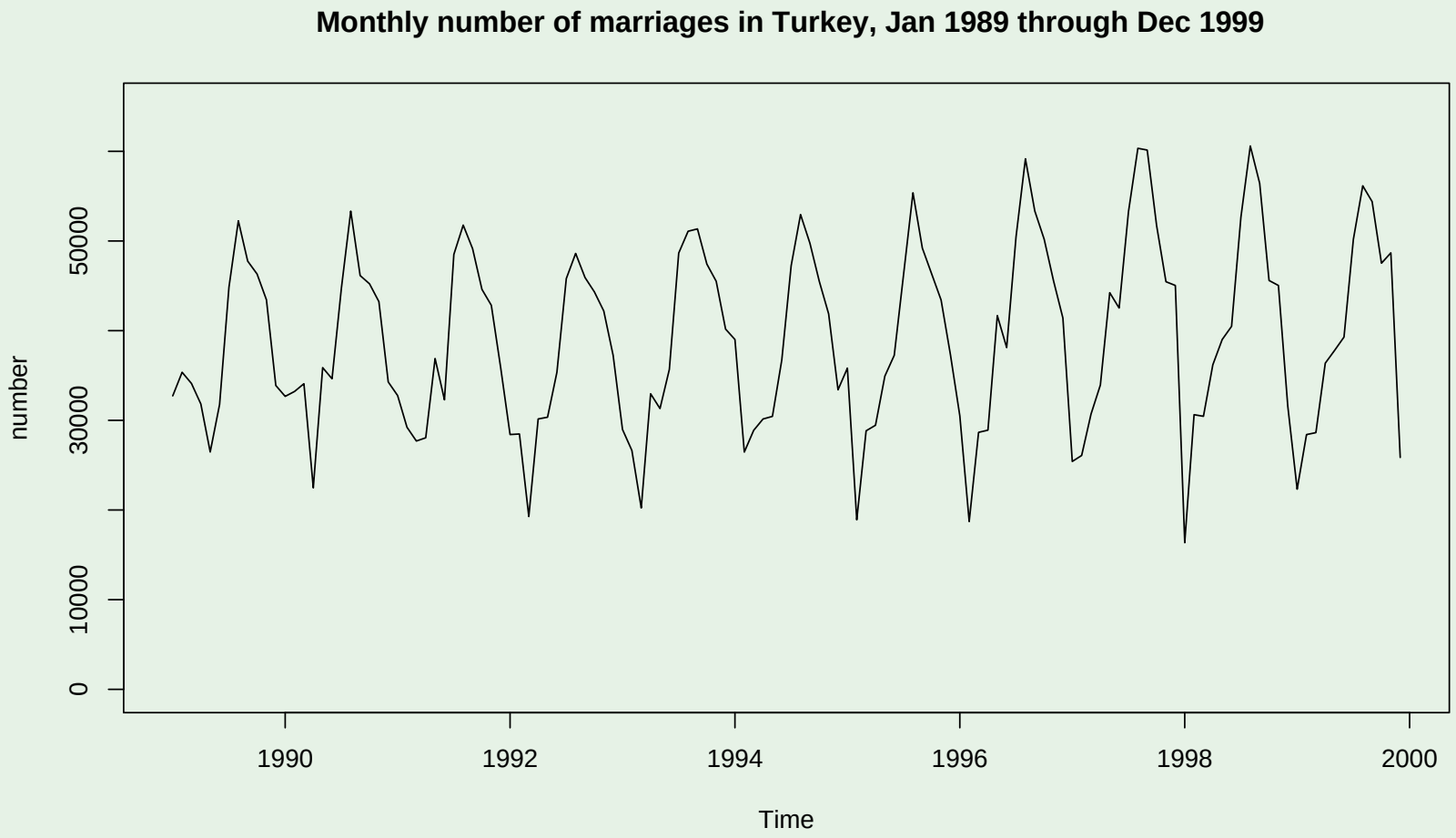
2. An Elementary Example (ctd.)

	xu100	dji	brent
min	−22.39859	−9.09702	−29.12821
lower quartile	−3.21090	−1.01842	−2.70595
median	0.95726	0.23886	0.31192
upper quartile	4.85772	1.42617	3.11213
max	37.19807	12.69361	42.22569
week of min	1998-09-01	2002-07-23	1991-01-22
week of max	1991-11-26	2002-07-30	1990-08-07

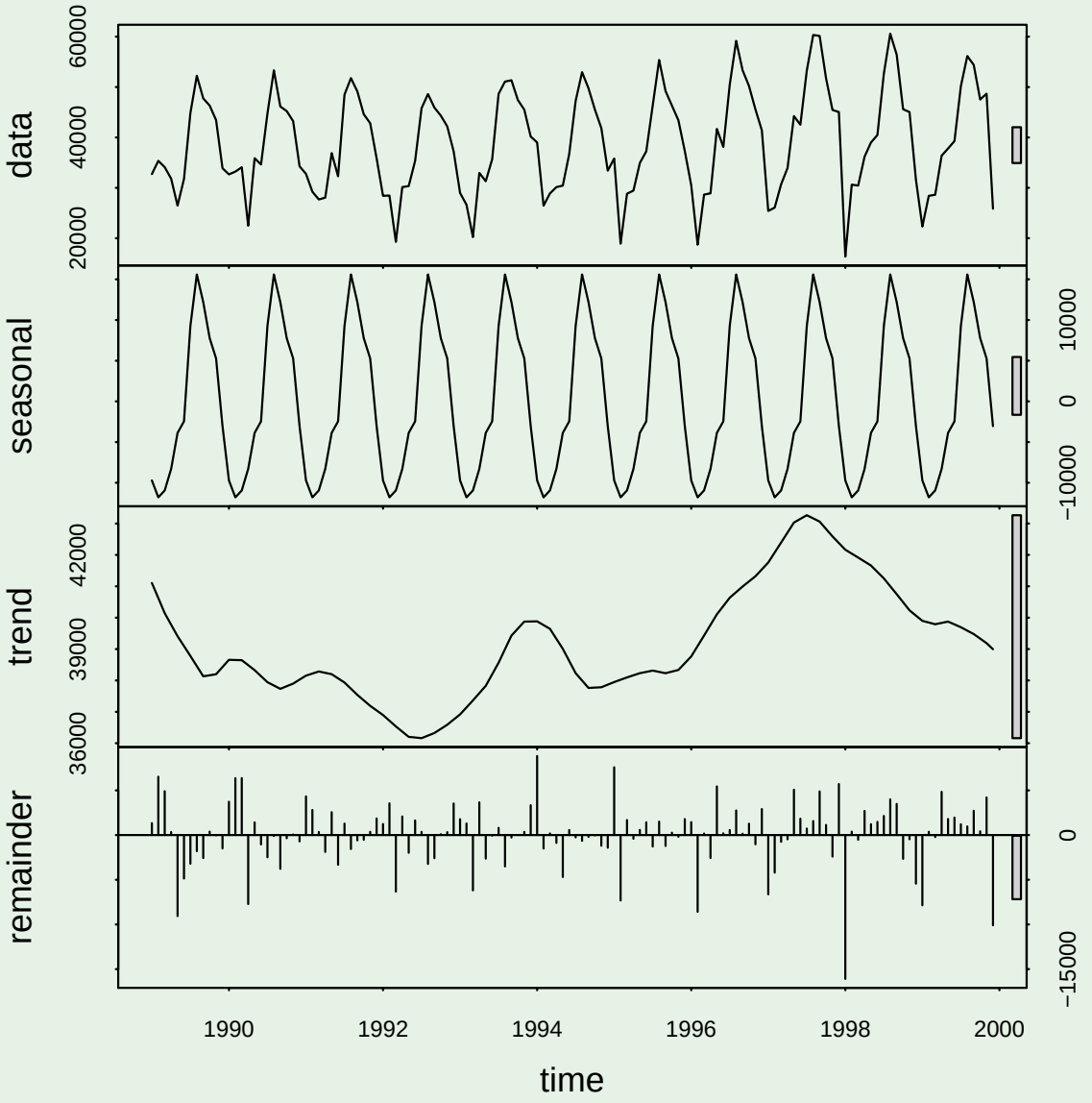
2. An Elementary Example (ctd.)



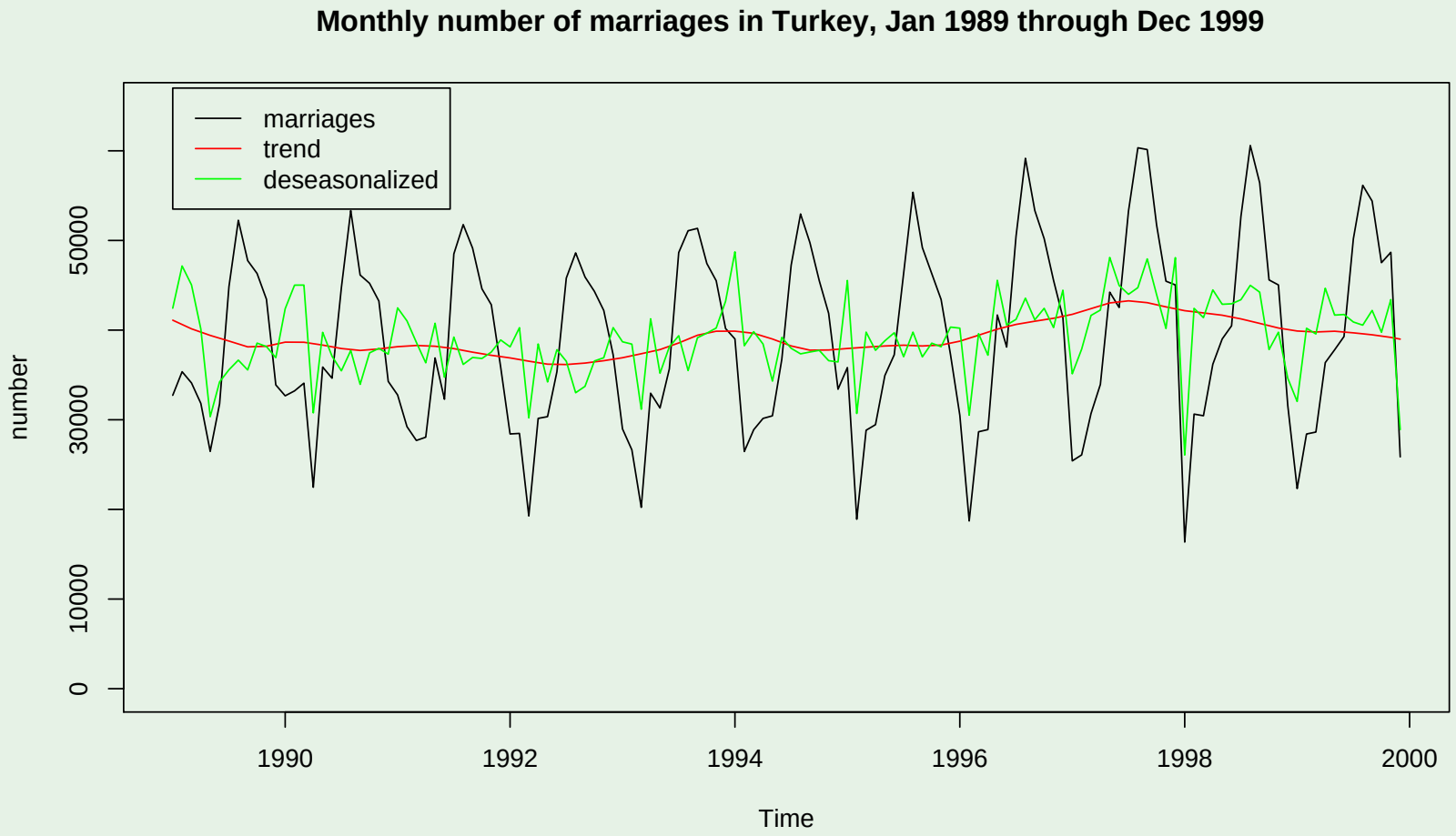
3. Marriages in Turkey



3. Marriages in Turkey: stl analysis



3. Marriages in Turkey (ctd.)



4. Stochastic Models Behind the Observations

- A stochastic model is a mathematical model which describes a chance setup.
- The paradigm of inductive statistics is:

> Regard the observed data
> as the outcome of a chance setup.

This means:

- We have to identify a suitable stochastic model.
- We have to "learn" about its parameters.

5. Time Series Analysis: A GARCH Example

The simplest GARCH model is:

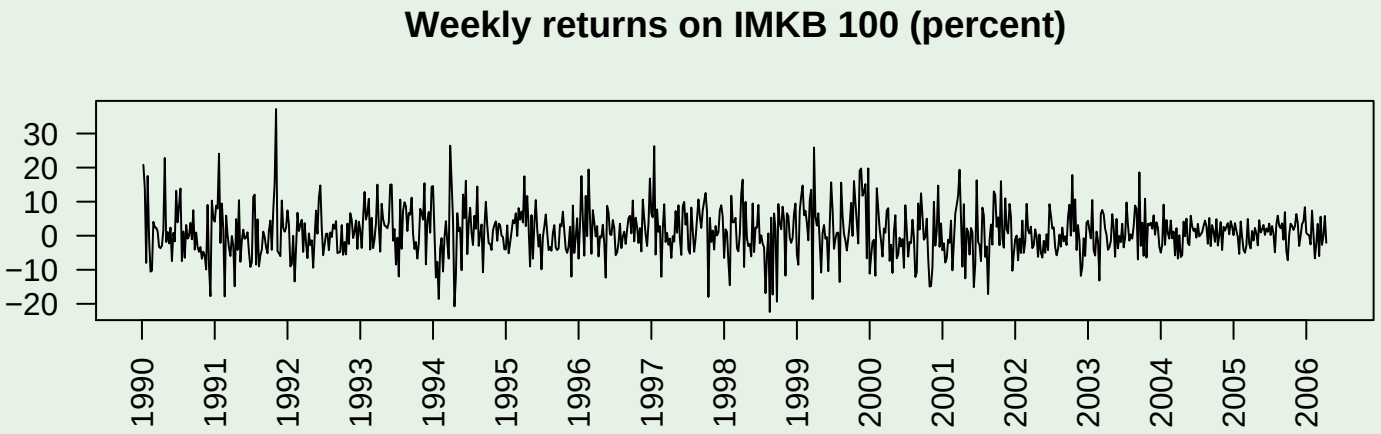
$$\epsilon_t = \nu_t \cdot \sqrt{h_t}, \quad h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta h_{t-1}$$

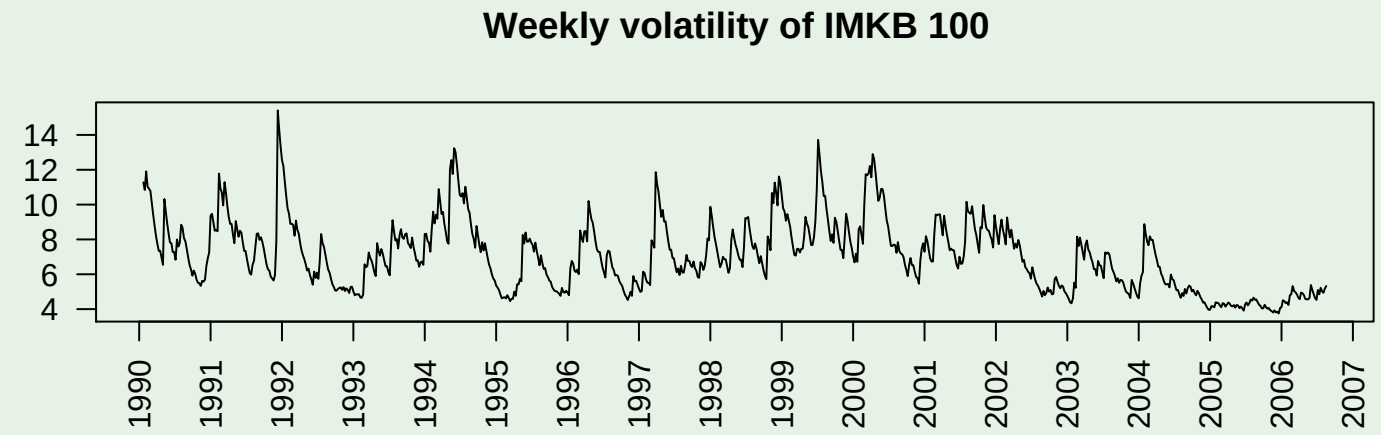
(Engle 1982; Bollerslev 1986)

For a series of weekly returns on İMKB 100:

$$\begin{array}{rcccccc} h_t & = & 2.713 & + & 0.137 & \epsilon_{t-1}^2 & + & 0.819 & h_{t-1} \\ & & (0.955) & & (0.020) & & & (0.024) & \end{array}$$

5. A GARCH Example (ctd.)



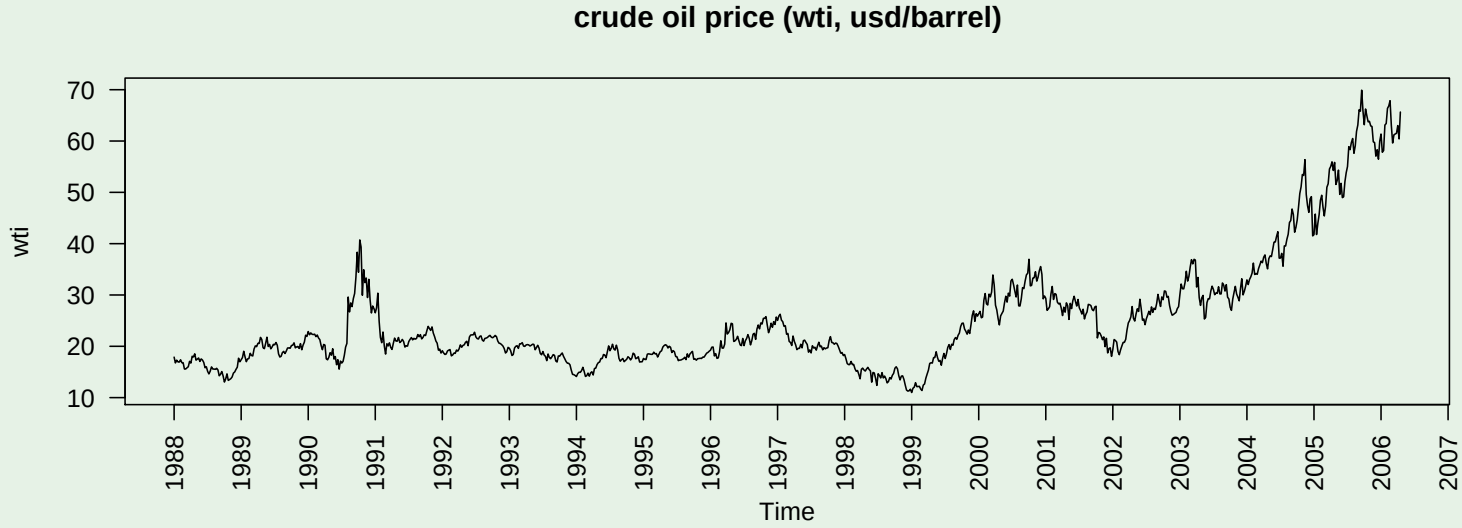


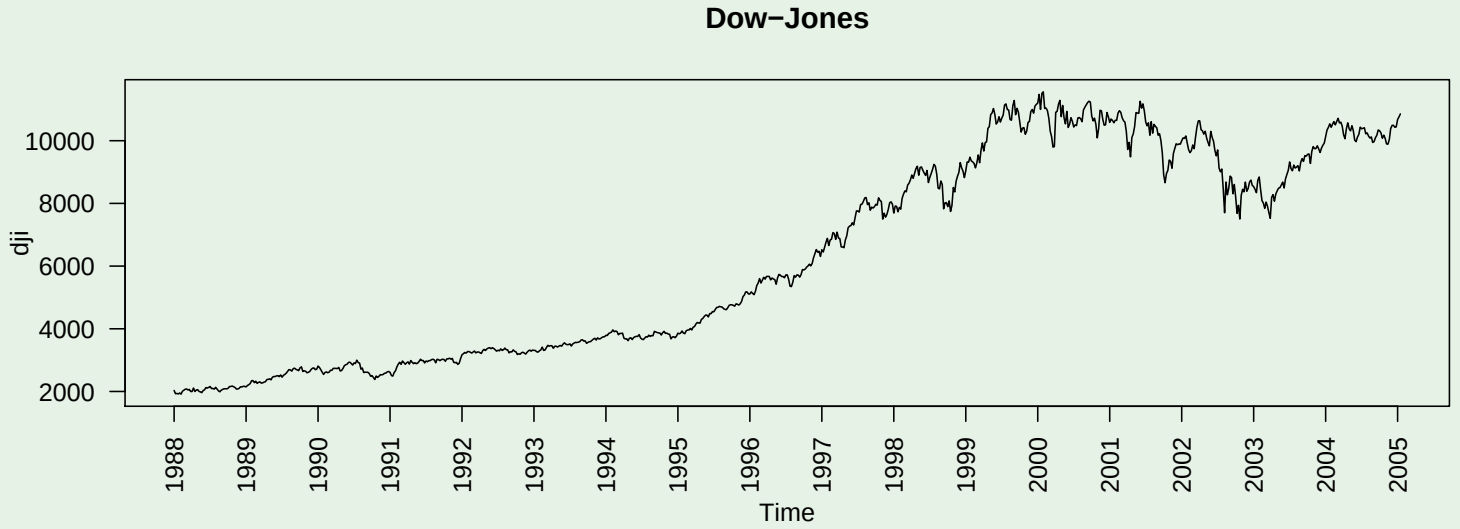
6. Modeling Volatility Spillovers

Example: Volatility spillovers between crude oil and the stock market.

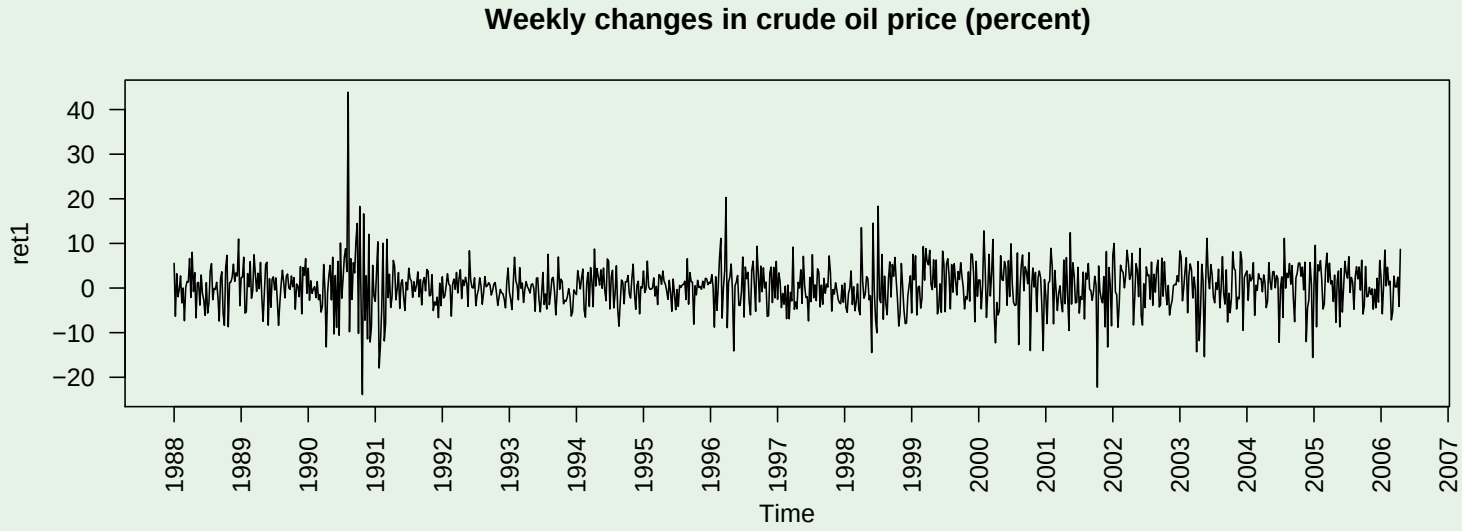
- Wild fluctuations in the price of crude oil will somehow affect the stock market.
- Often we observe periods of high variability in $both$ time series.
- Sophisticated models are needed to investigate this effect.

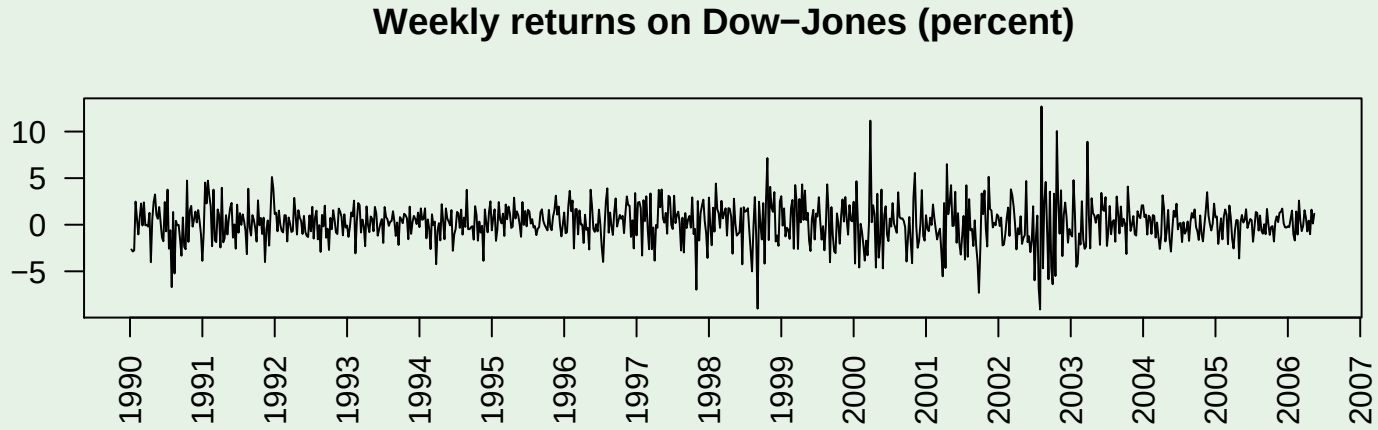
6. Modeling Volatility Spillovers (ctd.)





6. Modeling Volatility Spillovers (ctd.)





6. Modeling Volatility Spillovers (ctd.)

To investigate volatility spillovers, we need some kind of a bivariate GARCH model:

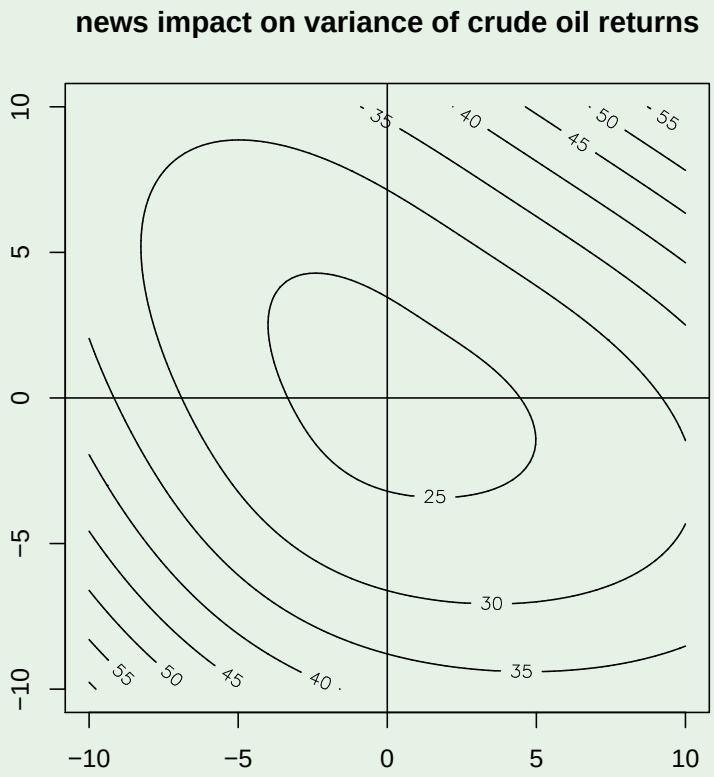
$$\boldsymbol{\epsilon_t} = \mathtt{H}_t^{1/2} \cdot \boldsymbol{\nu_t}$$

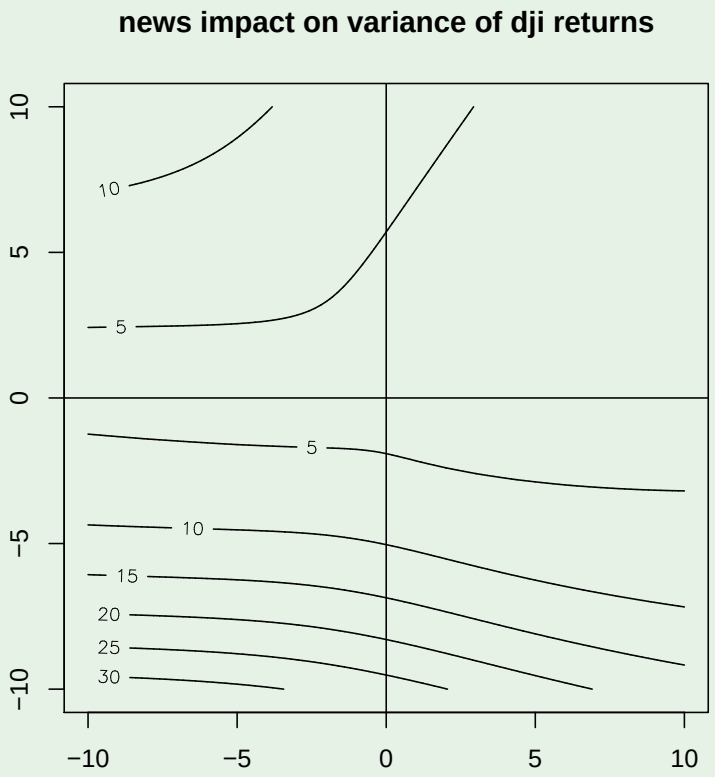
The conditional covariance matrix is defined as

$$\begin{aligned} \mathtt{H}_t \quad = \quad & \mathtt{C}'\mathtt{C} \quad + \quad \mathtt{A}'\boldsymbol{\epsilon_{t-1}}\boldsymbol{\epsilon}'_{\boldsymbol{t-1}}\mathtt{A} \; + \; \mathtt{B}'\mathtt{H}_{t-1}\mathtt{B} \\ & + \boldsymbol{S_w}(\boldsymbol{\epsilon_{t-1}}) \cdot \boldsymbol{\Gamma}'\boldsymbol{\epsilon_{t-1}}\boldsymbol{\epsilon}'_{\boldsymbol{t-1}}\boldsymbol{\Gamma} \end{aligned}$$

This is implemented in the R-package mgarchBEKK (Harald Schmidbauer & Vehbi Sinan Tunalıoğlu)

6. Modeling Volatility Spillovers (ctd.)





7. The R Project and Linux

- R harmonizes very well with LaTeX.
- The R project and further applications. . .