# Autoregressive processes I.

Beáta Stehlíková

# Estimating parameters of an AR model in R

# Estimating parameters: `astsa` library

Library `astsa` (*applied statistical time series analysis*) and its function `sarima` - also for more complicated models:

```
library(astsa)
```

# Example from the lecture - modelling spread

Example from the lecture - modelling spread, i.e., the difference between long term and short term interest rate.

Read the data from the website:

```
rs <- read.table("RSQ.txt")       # short term rate
r20 <- read.table("R20Q.txt")    # 20Y rate
```

Create the variable `spread`:

```
spread <- r20 - rs
```

# Example from the lecture - modelling spread

We have a vector `spread` - we **we transform it to a time-series object**, which includes the information about the structure of the data

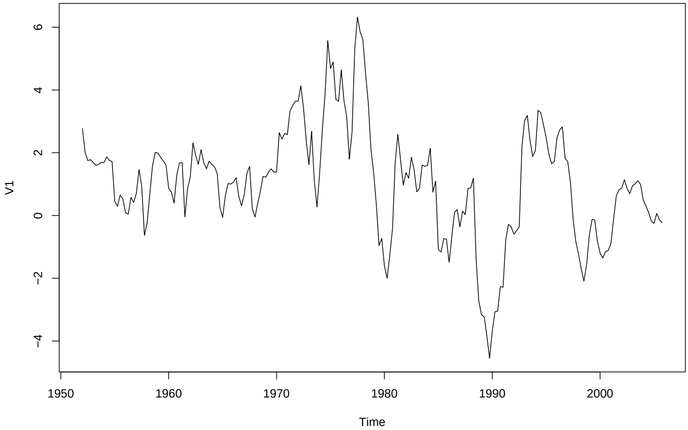
We use the function `ts`:

```
spread <- ts(spread, frequency=4, start=c(1952,1))
```

Plot its evolution (the x-axis now has the correct time):

```
plot(spread)
```

# Example from the lecture - modelling spread



# Example from the lecture - modelling spread

We can plot the sample autocorrelation function:

Lecture: It looks like an ACF of an AR(1) process, but it is not a good model for these data. We repeat this - we estimate the model and check residuals.

How to use the function `sarima`:

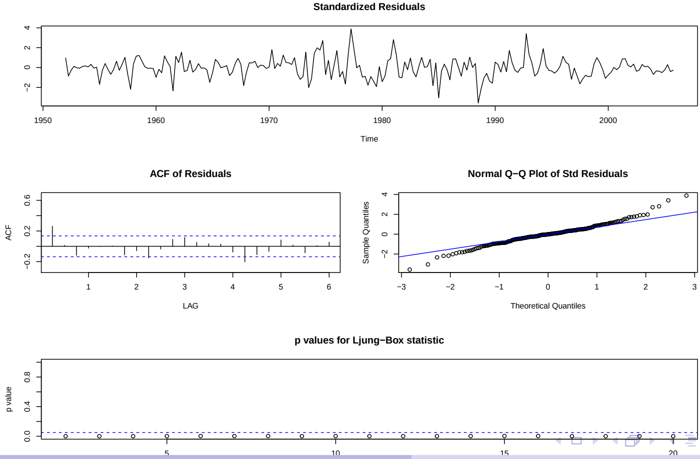
```
sarima(data,p,0,0)  # AR(p) model pre data
sarima(data,p,0,0, details=FALSE)
```

We get:

- estimates of parameters, standard errors, information criteria
- graphically: residuals, their sample ACF, p-values of Ljung-Box statistics

# Example from the lecture - modelling spread

```
sarima(spread,1,0,0, details=FALSE)
```



# Example from the lecture - modelling spread

**Exercises:**

- Why are these residuals not good?
- Estimate AR(2) model for the variable `spread`.
- Are the residuals good?

We will study stationarity on the lecture and verify the stationarity of this model.

# Predictions

Function `sarima.for`:

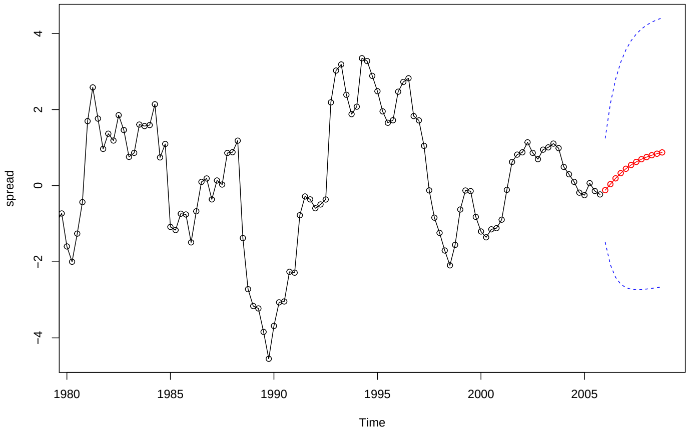
```
# prediction from AR(p) model for n periods ahead:
sarima.for(data,n,p,0,0)
```

# Example from the lecture - modelling spread

Prediction from AR(2) model for the next three years (so 12 quarters):

```
sarima.for(spread,12,2,0,0)
```

# Example from the lecture - modelling spread



# World Bank data and ggplot plot

# Libraries

`WDI` - World Development Indicators, access to data directly from R
`ggplot` - nice graphs

Install and load the libraries.

```
library(WDI)
library(ggplot2)
```

# Searching data in WDI

For example:

```
WDIsearch('gdp')

WDIsearch('gdp.*capita')
# for those know it: also other regular expressions
```

If there is a lot of results, we might want to display some of the first ones:

```
WDIsearch('gdp.*capita')[1:5,]
```

# Searching data in WDI

```
##       indicator
## [1,] "GDPPCKD"
## [2,] "GDPPCKN"
## [3,] "NV.AGR.PCAP.KD.ZG"
## [4,] "NY.GDP.PCAP.CD"
## [5,] "NY.GDP.PCAP.KD"
##       name
## [1,] "GDP per Capita, constant US$, millions"
## [2,] "Real GDP per Capita (real local currency units, va
## [3,] "Real agricultural GDP per capita growth rate (%)"
## [4,] "GDP per capita (current US$)"
## [5,] "GDP per capita (constant 2000 US$)"
```

# Reading the data

```
data <- WDI(indicator='NY.GNP.PCAP.CD',
            country=c('CA','US','FR', 'DE'),
            start=1975)
```

- we find `indicator` using `WDIsearch`
- `country` in the `iso2c` format
- default `start` is 2005, but for many datasets we have also longer series (we see the beginning of the loaded data)

# Reading the data

```
head(data)
```

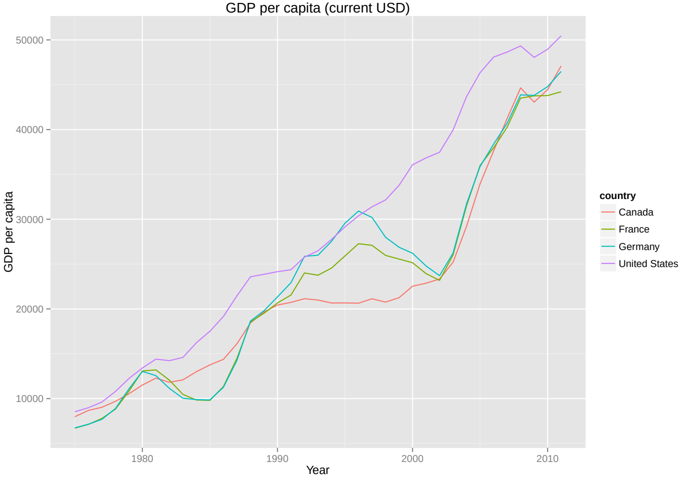
```
##   iso2c country NY.GNP.PCAP.CD year
## 1    CA  Canada          47090 2011
## 2    CA  Canada          44450 2010
## 3    CA  Canada          43060 2009
## 4    CA  Canada          44650 2008
## 5    CA  Canada          41230 2007
## 6    CA  Canada          37610 2006
```

# Graphs using ggplot

```
ggplot(data, aes(year, NY.GNP.PCAP.CD, color=country))
    + geom_line()
    + xlab('Year') + ylab('GDP per capita')
    + labs(title = 'GDP per capita (current USD)')
```

- data is a *data frame*,where we have the time series
- year, NY.GNP.PCAP.CD - from the data frame data we take year to x-axis and NY.GNP.PCAP.CD to y-axis
- color=country - graphs will be distinguished by colours (therefore color) based on the variable country

# Graphs using ggplot



# Exercise: Modelling logarithm of GDP and the growth rate

# Data

```
# GDP per capita (constant 2000 US$)
data <- WDI(indicator='NY.GNP.PCAP.KD',
            country=c('US'),
            start=1965)

# we sort the data (increasing time)
data <-  data[order(data$year),]

# our variable for modelling
log.y  <-  log(data$NY.GNP.PCAP.KD)
# we add time structure
log.y <- ts(log.y, start=1965, frequency=1)
```

# Modelling

We plot the data `plot(log.y)` and we see that the data are not stationary (increasing trend). Therefore we are going to model the logarithms (that actually the growth rate of the GDP).

In R:

```
sarima(x,p,1,0) # AR(p) model for the differences of x
```

Often, AR(1) model is suitable for modelling the growth rate of the GDP.

Estimate this model for our data. Check its stationarity and residuals.

Repeat for another country.