Exercise Book

Mark Trede, Willi Mutschler and Rafael Kawka

Version: November 16, 2015

Overview of Exercises

1 A short introduction to R **1**

1.1 Starting and quitting . . . 1
1.2 Scripts . . . 1
1.3 Working directory . . . 1
1.4 Help and comments . . . 1
1.5 Packages . . . 1

2 Importing data into R **2**

2.1 Reading text files . . . 2
2.2 Reading excel files . . . 2
2.3 Other data formats . . . 2
2.4 Missing values and trimming . . . 2

3 Describing data in R **3**

3.1 Head and tail . . . 3
3.2 Attaching dataframes . . . 3
3.3 Simple plots . . . 3
3.4 Stock returns . . . 3

4 Graphics with R **4**

4.1 School data . . . 4
4.2 Index returns . . . 4

5 Programming with R **5**

5.1 Using functions . . . 5
5.2 Sequences and other vectors . . . 5
5.3 Random numbers . . . 5
5.4 Loops . . . 6
5.5 Functions . . . 6
5.6 Numerical optimization . . . 6

6 Probability theory **7**

6.1 Moments . . . 7

7 Multiple linear regression **8**

7.1 Student teacher ratio (I) 8

7.2 Capital asset pricing model 8

7.3 Student teacher ratio (II) 9

7.4 Omitted variable bias 9

7.5 Asymptotic normality 10

7.6 Pitfalls in the linear regression model (I) 10

7.7 Pitfalls in the linear regression model (II) 11

7.8 Pitfalls in the linear regression model (III) 11

8 Multivariate random variables **12**

8.1 Joint distributions 12

8.2 Gaussianity or else? 12

8.3 Gaussian and uncorrelated, but dependent 12

8.4 Delta method 13

9 Stochastic convergence and limit theorems **14**

9.1 Law of large numbers 14

9.2 Law of large numbers for the variance 14

9.3 Central limit theorem 14

9.4 Central limit theorem for dependent data 15

9.5 Limits of maxima (I) 15

9.6 Limits of maxima (II) 16

9.7 Limits of maxima (III) 16

10 Estimators and their properties **17**

10.1 Counter examples 17

11 Least Squares and Method of Moments **18**

11.1 Nonlinear least squares 18

11.2 Method of moments for the binomial distribution 18

11.3 Method of moments for the geometric distribution 18

11.4 Method of moments for the Gumbel distribution 18

11.5 Method of moments for the Pareto distribution 19

11.6 Method of moments for the uniform distribution 19

11.7 Method of moments for the linear regression model 19

12 Maximum likelihood estimation **20**

12.1 Extreme values 20
12.2 Parameters of the uniform distribution 20
12.3 Censored lognormal distribution 20
12.4 Exponential model 21
12.5 Tobit model 21
12.6 Probit model 22
12.7 Logit model 22
12.8 Heckman regression 23
12.9 Count data 23
12.10 Stochastic frontier analysis 24
12.11 ARCH models 24
12.12 Duration models 25
12.13 Ultra-high-frequency data 25
12.14 Spatial dependence 26

13 Instrumental variables **27**

13.1 The miracle of the instruments 27
13.2 Linear combinations of instruments 28
13.3 Compulsory School Attendance 28
13.4 A simple example 29
13.5 Money demand 29
13.6 Tests for the IV model 30

14 GMM **31**

14.1 The R package gmm 31
14.2 Nonlinear least squares estimation and GMM 31
14.3 Ordinary least squares estimation and GMM 32
14.4 Maximum likelihood estimation and GMM 32
14.5 Instrumental variables estimation and GMM 33
14.6 Moment conditions and moment existence 33
14.7 Standard CAPM 34
14.8 Consumption-based CAPM 34
14.9 Minimum distance estimation 35

15 Indirect inference **36**

15.1 AR(1) processes . . . 36

15.2 Filter models using the Kalman filter . . . 36

15.3 Estimation of the Cox-Ingersoll-Ross model . . . 37

15.4 Ornstein-Uhlenbeck process . . . 37

15.5 Time-aggregated observations . . . 38

16 Bootstrap **39**

16.1 Omitted variables bias does not go away . . . 39

16.2 Confidence intervals for the Gini index . . . 39

16.3 Confidence intervals for correlation coefficients . . . 39

16.4 The t-test . . . 40

16.5 The percentile-t-method . . . 41

16.6 Heavy tails and variance testing . . . 41

16.7 Time series . . . 42

16.8 Bootstrap test for the Zipf index of city size distributions . . . 42

1 A short introduction to R

1.1 Starting and quitting

Start R. The window you see is the "R Console" and we will call it the command window in the following. Inside the command window, compute $1+1$, $2-1$, $3/2$, $2*4$ and 2^{10}. Quit R using the command `q()`, without saving the workspace.

1.2 Scripts

Restart R. In the menu, choose "Datei", then "Neues Skript". A new window opens. Type the following four lines:

```
a <- 3
b <- 4
c <- a+b
print(c) pi Pi PI
```

Mark the lines and press Strg+R (or Ctrl+R). Save your script under any name (preferably with the extension `.R`) on your hard disk or USB flash drive. Quit R, restart, open the script, and execute it.

1.3 Working directory

In R, you can obtain the current working directory using the command `getwd()` ("get working directory"). This is the directory where R saves files, and where it looks for files to read.

1. Find out where your working directory is.
2. You can change the working directory using the command `setwd("x:/path")` where `x:` is the drive (e.g. hard disk) and `path` the complete path. Note that the path does not contain the backslash ("\") which is usually used in Windows, but the slash ("/"). Change your working directory as you like. Check if the change was successful.

1.4 Help and comments

1. A very important command in R is the question mark (`?`) followed by any name of a function. This way you can start the help function, giving you details about any R command. Read the help page for the command `mean`.
2. The hash sign (`#`) is the comment sign. Everything following the comment sign is ignored (until the end of line). Insert some comments into your script and re-execute it.

1.5 Packages

An advantage of R is the large number of packages available on CRAN. Packages increase the functionality of R. If your computer is connected to the internet, you can install new packages by choosing the menu items "Pakete", "Installiere Paket(e)...".

1. Install the package `xlsx`. Then activate the package using the command `library(xlsx)`. Typing `library(help=xlsx)` will give you more information about the new commands.
2. Install the package `AER`. We will need it for the next exercises.

2 Importing data into R

When learning a new computer language, the most basic standard problem is how to import data. In this exercise you will learn a number of ways to read datasets. You can type the commands either in the command window or, preferably, write and save a script and then execute it.

2.1 Reading text files

Download the file `bsp1.txt` from the course page and save it in the directory `c:/temp` (of course, you may use other directories). Change the working directory to `c:/temp`. Use the command

```
bsp1 <- read.csv("bsp1.txt")
```

to import data into object `bsp1` and type `print(bsp1)`, or simply `bsp1`, to see the dataset.

2.2 Reading excel files

Download the excel file `bsp2.xlsx` from the course page and save it. Reading excel files is rather uncomfortable in R.

1. Open the file from Excel and save it as `bsp2.csv`. In contrast to the English version, the German version of Excel does not write a decimal point, but a comma, and entries are not separated by commas, but semicolons.[1] If your data are saved in the German format, you can read the data using one of the two following commands

   ```
   bsp2 <- read.csv("bsp2.csv",dec=",",sep=";")
   bsp2 <- read.csv2("bsp2.csv").
   ```

   Import the dataset and have a look at it using `print(bsp2)`.

2. If you insist to read Excel files, the best way to do it is by means of the package `xlsx`. Activate the package using `library(xlsx)`. Read the help text of the command `read.xlsx`. Load the file `bsp2.xlsx` using the command `read.xlsx` and print the data.

2.3 Other data formats

1. There is a large number of packages to make foreign data formats readable in R. The most important package is `foreign`, which can be used to read SPSS and Stata files (but not Excel). Install and activate the package `foreign` and read the corresponding help with `library(help=foreign)`. Load `bsp3.dta` into the object `bsp3` and print it.

2. R also has got its own data format. You can save objects using the command `save` and then re-load them with `load`. Save the object `bsp3` in the file `bsp3.Rdata`, quit R, restart, and load the file `bsp3.Rdata`. Print `bsp3`.

3. Try `scandat <- scan()` and insert some data. Edit your data with `edit(scandat)`.

2.4 Missing values and trimming

`NA` stands for a missing value. NaN stands for Not a Number (example `0/0`). Missing values can produce errors in some functions and you should either remove them (trimming) or replace them with a 0. Create a vector `y <- c(1:3,NA,NA,4:2)` and (i) trim or (ii) replace them with 0.

[1] Please always check, if your Excel version uses the German or the English format.

3 Describing data in R

Imported datasets are usually stored as dataframe objects. A dataframe is almost the same as a matrix. Each row is an observation, and each column is a variable. Create a new script for the following exercises to be able to repeat the commands.

3.1 Head and tail

On the internet site of the course you will find the file `indices.csv`. It contains the daily index values of the two indices DAX and FTSE 100 from 8/9/2005 to 8/9/2010.[2] Load the data into R and save them as dataframe `indices`. Large dataframes cannot be printed nicely. A good way to learn about the structure of the dataframe is the command `head(indices)`. Try it (by the way, you can also use `tail(indices)`). If you are only interested in the variable names of the dataframe, just type `names(indices)`. For a thorough insight try also `str(indices)`, `class(indices)` and `attributes(indices)`.

3.2 Attaching dataframes

If you use the `attach` command, the columns of the dataframe are accessible by the column names as ordinary variables. Now you can directly access the two variables `dax` and `ftse`. Type `attach(indices)`. *Note: The help page for* `attach` *notes that attach can lead to confusion: The possibilities for creating errors when using attach are numerous.* Therefore we are going to avoid it and use the `$` sign to attach variables, i.e. `dax <- indices`$dax; ftse < -indices$`ftse`. Another way is to directly access the columns of the dataframe, i.e. `dax <- indices[,1]; ftse <- indices[,2]`. Save the DAX series into `dax` and the FTSE series into `ftse`.

3.3 Simple plots

Type `plot(dax)` to create a graph showing the time series of the DAX index. Create a new graph of the time series of the logarithm of the DAX index.

3.4 Stock returns

1. Save the number of observations into the variable `n`. Hint: The command `dim(x)` returns the number of rows and columns of `x` as a vector.
2. Generate a new variable containing the daily returns of the DAX index:

   `rdax <- log(dax[2:n]/dax[1:(n-1)])`

   and plot them. Define and plot the returns of the FTSE in a similar way (`rftse`).
3. Activate the package `MASS`. Use the command `truehist` to draw the histogram of the DAX returns.
4. For the DAX returns, compute the mean (`mean`), the standard deviation (`sd`), the variance (`var`), the median, the 1%-and the 99% quantiles (`quantile`), and the range (`range`).
5. Sometimes boxplots are a nice way to present a dataset. Type `boxplot(rdax,rftse)`.
6. Plot the DAX returns against the FTSE returns using `plot(rdax,rftse)`.
7. Compute the correlation between the DAX returns and the FTSE returns (`cor`).
8. Compute the correlation of the DAX returns with its lagged (by one day) return.

[2]Since working with calendar dates is a bit cumbersome in R, the information about the dates has been omitted.

4 Graphics with R

A strength of R is its flexible way to create graphics. The following exercises illustrate that. Please write scripts for these exercises.

4.1 School data

1. Download the dataset `caschool.csv` into the object `caschool`. This dataset is discussed in great detail in the textbook of Stock and Watson. The codebook (`caschool.pdf`) is downloadable from the internet site of this course. Draw a scatterplot of the variable `testscr` against `str`.

2. Re-create the same plot with nicer and more informative axis labels (the axes options in the `plot` command are `xlab` and `ylab`).

3. Re-create the plot again and add a title (using the `main` option of the `plot` command).

4. The `col`-option can be used to change the colors of the points or lines. Try it. A list of all available color names is `colors()`. One can even color different parts of the plot differently, but we omit that here.

5. The command `points` adds one or more points into an existing plot. Add the point (mean of `str`, mean of `testscore`) to your last plot in red color. If you want to change the point symbol, you can use the option `pch`, see also `?points`.

6. The `text` command inserts text into an existing plot. Label the red point with the text "mean". The easiest way to position the text is by means of the mouse. Use the command `locator`, e.g. as in `text(locator(1),"mean")`.

7. One can partition the window into an array of small windows. You can prepare a partition using `par(mfrow=c(n,m))` where $n \times m$ is the number of plots (n rows and m columns). Prepare a window for four scatterplots (2×2). Plot the scatterplots of `testscr` against (a) the teacher-student ratio `str`, (b) the percentage of English language learners `el_pct`, (c) the percentage qualifying for reduced price lunch `meal_pct`, (d) the percentage qualifying for income assistance `calw_pct`.

4.2 Index returns

1. Download the dataset `indices.csv`. Generate a new variable with starting value 100 that represents the relative time series of the DAX index. Plot the time series of this normalized DAX index using the command `plot` with the options `type="l"` (for "line") and `col="blue"`.

2. Add the normalized FTSE index to the last plot. Use the command `lines` with the color option `col="red"`.

3. Use the `legend` command to add a legend explaining the meaning of the two colored lines. You may use the command `locator` to find a suitable position for the legend.

5 Programming with R

5.1 Using functions

1. Functions are called by its name followed by the arguments in parentheses. The syntax is

   `function(Argument 1=arg1, Argument 2=arg2, ...)`

   You can specify arguments either by regarding the order they need to be called (`log(10,10)`) or by specifying the argument itself (`log(10,base=10)`).

   You can call several functions at a time using ";". Whenever you encounter the symbol + instead of >, you have forgotten to close parentheses. Just type `")"` or hit ESC.

   Try:

   `sqrt(2); sin(pi); exp(1); log(10); log(10,10);log(10,base=10);`
   `sqrt(2` (without closing the bracket!)

2. The concatenation function c() creates vectors. You can pick the i-th item of a vector using square brackets. Try:

   ```
   simpsons <- c("Homer","Marge","Bart","Lisa","Maggie")
   x <- c(1,2,3,4,5,6,7,8,9,10)
   x <- c(1:10)
   length(simpsons); sum(x); mean(x)
   simpsons[3]
   ```

3. Consider the vector `x <- 0:10`. Use the function `sum()` to calculate the sum of all values that are smaller than 5, i.e. 0+1+2+3+4 = 10.

5.2 Sequences and other vectors

In R, sequences are generated by the `seq`-command. An abbreviated form for integers is `from:to`. To generate a vector with repeated elements use the command `rep`.

1. Generate the vectors $x = (1, 2, \dots, 100)$ and $y = (2, 4, 6, \dots, 1000)$.
2. Generate an equi-spaced grid from -4 to 4 with 500 grid points.
3. Generate a vector of $n = 100$ missing values (`NA`).
4. Generate the vector $x = (0, 1, 2, 0, 1, 2, \dots, 0, 1, 2)$ of length 300.
5. Generate the vector $x = (0, \dots, 0, 1, 0, \dots, 0)$ of length 100 with the 1 at position 40.

5.3 Random numbers

There are random number generators for a large number of distributions. The general syntax is

`rNAME(n,parameters)`

where `NAME` is an abbreviation of the distribution name (e.g. `norm`, `lnorm`, `binom`, etc.), `n` is the number of values to be drawn, and `parameters` are the parameter(s) of the distribution.

1. Activate the `MASS` package. Generate a vector `x` of $n = 10000$ random numbers drawn from the standard normal distribution and plot the histogram.

2. Generate a vector `r` of $n = 500$ random numbers drawn from the t-distribution with 3 degrees of freedom (see `?rt`). Execute `plot(r)`.

3. Cumulate the vector `r` using the command `cumsum`. Plot the cumulated series.

5.4 Loops

In general, one should try to avoid loops in R as they often slow down the computations considerably. In this course, we will ignore this advice for didactical reasons. The type of loop that is used most often, is the `for`-loop. Unfortunately, the help function does not work for the loop commands, please type `?Control` to read the help text. The syntax of the `for`-loop is

```
for( [var] in [sequence]) { [commands] }
```

where `[var]` is an index variable and `[sequence]` is a vector of values to be assigned to the index variable. In our applications, we often need to store the results computed within the loop in a result vector. In this case, it is advisable to initiate an empty vector before the loop starts:

```
Z <- rep(NA,100)
for(i in 1:100) { [compute something with result x]; Z[i] <- x }
```

1. Generate a vector `r` of $n = 500$ random numbers drawn from the t-distribution with 3 degrees of freedom. Use a `for`-loop to compute the moving average of `r` within a window of length 21.
2. Write a program using a `for`-loop over $r = 1, \ldots, 10000$ to perform the following steps for every r: Generate a sample of size $n = 100$ from the lognormal distribution $LN(0, 1)$. Find the maximum and store it. After the loop is performed, plot the histogram of the maxima.

5.5 Functions

Functions are very powerful in R. Their general syntax is

```
f <- function(arg1,arg2,...)  { [commands to compute output var]; return(var)}
```

where the arguments can be scalars, vectors, matrices etc. For example, the following function computes and returns $x^2 + 2y^2$.

```
fexmpl <- function(x,y) { z <- x^2+2*y^2; return(z)}
```

Once the function has been defined it can be used like any other internal R function.

1. Define a function $f(x) = x^2 + \sin(x)$ where x can either be a scalar or a vector. Define a grid of length 500 on the interval $[-3, 3]$ and plot the function.
2. Define a function that computes the empirical raw moment of order p for a sample $x_1, \ldots, x_n$, i.e. $m_p = \frac{1}{n} \sum_{i=1}^{n} x_i^p$.

5.6 Numerical optimization

There are two commands for numerical optimization: `optimize` for univariate optimization and `optim` for multivariate optimization.

1. Numerically find the minimum of the function $f(x) = x^2 + \sin(x)$. *Hint: It lies between -1 and 0.*
2. Numerically find the minimum of the function $f(x, y) = x^2 + \sin(x) + y^2 - 2\cos(y)$. First get a view of the function using the following commands

```
f <- function(x,y) x^2+sin(x)+y^2-2*cos(y)
x <- seq(-5,5,by=.2);y <- seq(-5,5,by=.2);z <- outer(x,y,f)
persp(x,y,z,phi=-45,theta=45,col="yellow",shade=.65 ,ticktype="detailed")
```

You can try to edit phi and theta to get a better view.

6 Probability theory

6.1 Moments

1. Show that the moments of the standard normal distribution $N(0,1)$ are $\mu_r = 0$ for odd orders r, and $\mu_r = \prod_{i=1}^{r/2} (2i-1)$ for even orders r.

2. Let $X \sim N(\mu, \sigma^2)$ and $Y = \exp(X)$. Derive the expectation of Y.

3. The distribution function of the Pareto distribution with parameters $K > 0$ and $\alpha > 0$ is

$$F_X(x) = 1 - \left(\frac{K}{x}\right)^{\alpha}.$$

   where $x \geq K$. Derive the density f_X and the moment of order $p < \alpha$. Do moments of order $p \geq \alpha$ exist?

7 Multiple linear regression

Linear models are estimated in R by the command `lm`. This command has an unusual syntax and returns a rather complex object (called `lm` object). Be prepared: it takes some time to get used to that. We start with the simple linear regression model that is used in the textbook by Stock and Watson. The codebook `caschool.pdf` for the dataset can be downloaded from the course site.

7.1 Student teacher ratio (I)

1. Load the dataset `caschool.csv` into the object `caschool` and make `testscr` as well as `str` accessible. Perform the following commands:
   ```
   regr <- lm(testscr~str)
   print(regr)
   ```
   Create the scatterplot of `testscr` against `str` and then type `abline(regr)`. The color (`col`), the line type (`lty`), and the line width (`lwd`) can easily be changed by the options of the plot command. Try it.
2. The student teacher ratio `str` in the school district Antelope is 19.33 an. Predict the variable `testscore` for the district Antelope using the `predict` command. To do so, type `predict(regr,newdata=data.frame(str=19.33))`. Add the predicted value to the plot (in blue color).
3. Among other things, `lm` objects also contain the residuals of the regression. You can extract them using the function `residuals` with the `lm`-object as argument. Compute the sum of the residuals.
4. Create a plot showing the residuals. Add the horizontal axis using the command `abline(h=0)` (the `h` is for horizontal).
5. Plot the residuals against the variable `str`.
6. Load the `AER` package. Execute the commands `print(summary(regr))` and `print(coeftest(regr,vcov=vcovHC))`. Interpret the outputs.
7. Test the hypothesis $H_0 : \beta = -1$. Write down each step of the test procedure. *Hint: You can also make use of `linearHypothesis` function of the car package.*
8. Give a 95% confidence interval for β.

7.2 Capital asset pricing model

Load the dataset `capm.csv` and make the variables accessible. The variable `rdai` contains the daily returns (in %) of Daimler from 9/9/2009 to 8/9/2010, the variable `rdax` contains the DAX returns. The CAPM implies that the intercept of the simple linear regression

$$r_{DAI,t} = \alpha + \beta r_{DAX,t} + u_t$$

is zero.

1. Estimate the model and test the null hypothesis $H_0 : \alpha = 0$.
2. The coefficient β is a measure of the systemic risk. Give a 95% confidence interval for β.

7.3 Student teacher ratio (II)

The `lm`-command is also used to perform multiple linear regressions. The syntax is close to the simple linear models. Put the endogenous variable to the left of the tilde. On the right of the tilde you list the exogenous variables, separated by plus signs. It looks like this: `lm(y~x1+x2+x3)`.

1. Load the dataset `caschool.csv` into the object `caschool` and make `testscr`, `str`, `el_pct` and `expn_stu` accessible. Perform the following commands:
   ```
   regr <- lm(testscr~str+el_pct)
   print(regr)
   ```
   Explain the output.
2. Regress `testscr` on `str`, assign the residuals of the regression into the variable `r1` and plot them. Now regress `testscr` on `str`, `el_pct` and `expn_stu`, put the residuals into the variable `r2` and add them to the plot. Compute the sum of squared residuals for both regressions.
3. Consider the regression of `testscr` on `str`, `el_pct` and `expn_stu`. Using the `predict`-command, predict the value of `testscr` for a school district with an average class size (`str`) of 25 students, a percentage of English learners (`el_pct`) of 60% and an average expenditures per student (`expn_stu`) of 4000$. How would the result change if the average class size was reduced to 17?
4. Reconsider the regression of `testscr` on `str`, `el_pct` and `expn_stu`. Let `regr` be the object containing the regression results. Execute the commands `print(summary(regr))` and `print(coeftest(regr,vcov=vcovHC))`. Interpret the output.
5. Test the null hypothesis that the coefficients on `str` and `expn_stu` both equal 0 and the coefficient on `el_pct` equals -0.7. *Hint: Use the linearHypothesis function of the car package.*

7.4 Omitted variable bias

Load the dataset `omitted.csv` into the object `omitted` and make it accessible. There are five variables: `y`, `x1`, `x2`, `x3` and `x4`. The sample has been generated in R; the sample size is $n = 500$. The true regression surface is

$$Y = 1 + 2X_1 + 3X_2 + 4X_3 + 5X_4.$$

The exogenous variable X_1 is uncorrelated with X_2, X_3, X_4. The variables X_2 and X_3 are positively correlated, so are X_3 and X_4. The variables X_2 and X_4 are uncorrelated.

1. Estimate the intercept and the slope coefficients for X_1, X_2, X_3, X_4 from the dataset (the estimates should be close to the true values 1,2,3,4,5).
2. Estimate a regression of Y on X_2, X_3 and X_4. Explain why the estimates are still close to the true values.
3. Estimate a regression of Y on X_1, X_2 and X_3. Which coefficients are still estimated accurately? And why?

7.5 Asymptotic normality

1. Consider the multiple linear regression model $y = X\beta + u$. In R, generate the matrix X by executing the following commands:

   ```
   library(MASS)
   set.seed(123)
   X <- cbind(1,mvrnorm(n=100,c(5,10),matrix(c(1,0.9,0.9,1),2,2)))
   ```

   The true coefficient vector is
   $$\beta = \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}$$
   and the error terms are i.i.d. uniformly distributed on the interval $[-1, 1]$. Hence, the assumption of normally distributed error terms is violated.

2. Write an R program that generates $R = 10000$ random samples of size $n = 100$ each (the easiest way to do so is to use a `for` loop). Generate an empty vector `V <- rep(NA,10000)`. For each sample $i = 1, \ldots, R$, compute the OLS estimate $\hat{\beta}$ of β and store the second component of $\hat{\beta}$ in the i-th element of the vector `V`.

3. Plot the histogram of `V`.

4. Compute the mean m and standard deviation s of `V` and add the density of $N(m, s)$ to the plot. *Hint: You can use* `curve(dnorm(x,mean=m,sd=s),add=T)` *to add the Gaussian density with mean m and std. deviation s to the plot.*

5. Move the command that generates X into the loop (without the seed command). Now there is a new, random X for each sample. Is the normal approximation still valid?

6. Try if the approximation is worse for sample size $n = 10$ (you will have to shorten X in this case).

7.6 Pitfalls in the linear regression model (I)

A simple linear regression is very easily performed by any statistical program. However, there are many mistakes and misinterpretations that can be made. A critical inspection of your regression results is crucial. Three of the more common mistakes are illustrated in the following.

Load the dataset `gehaelter.csv` into the object `gehaelter` and make the variables accessible. The dataset contains observations on 100 graduates about their length of study (`dauer`), their initial salary (`gehalt`) and their major (`fach`, 1=chemistry, 2=economics).[3]

1. Draw the scatterplot of salary against length of study.

2. Perform a linear regression of salary on length of study and add the estimated regression line to the scatterplot. What is the effect of the length of study on the salary?

3. Repeat 1. and 2. separately for chemistry and economics graduates. What is the effect of the length of study on salary in each group?

4. Re-draw the scatterplot with economists colored in blue and chemists colored in red.

[3] The data are fictional and have been generated by a computer algorithm.

7.7 Pitfalls in the linear regression model (II)

Load the dataset `storch.csv`. It contains observations on the stork population (eyries) in Lower Saxony and the number of births in Germany from 1958 to 2004.

1. Plot the scatterplot of the number of births against the number of storks and perform a linear regression. What is the effect of the number of storks on the number of births?
2. Repeat the exercise with the number of out-of-wedlock births.

7.8 Pitfalls in the linear regression model (III)

Load the `indices.csv` (this dataset has been used before, see exercise 3.1). Execute the following commands:

1. ```
   n <- dim(indices)[1]
   kdax <- dax[6:n]
   kftselag <- ftse[1:(n-5)]
   ```
   There are two new variables: the DAX index `kdax` and the FTSE-100 index lagged by five trading days (`kftselag`).
2. Regress the DAX index `kdax` on the lagged FTSE index `kftselag`. Interpret the estimated coefficients.
3. Test the null hypothesis that the DAX index does not depend on the lagged FTSE index (significance level 0.05).
   ```

8 Multivariate random variables

The package `MASS` includes a command to generate i.i.d. draws from the multivariate normal distribution. Type `library(MASS)` to activate it.

8.1 Joint distributions

Consider the bivariate density

$$f(x,y) = 40 \cdot (x-0.5)^2 \cdot y^3 \cdot (3-2x-y)$$

for $(x,y) \in [0,1] \times [0,1]$ and $f(x,y) = 0$ else.

1. Show that $f(x,y)$ is really a density function.
2. Derive the marginal densities $f_X(x)$ and $f_Y(y)$ and plot them.
3. Derive the conditional density of X given $Y = y$ and plot it for $y = 0.01$ and $y = 0.95$.
4. Are X and Y independent?

8.2 Gaussianity or else?

Load the dataset `gaussian.csv` into the object `gaussian`. Each column of the dataframe `gaussian` is a variable (V1, V2, V3, V4).

1. Split the screen into 2×2 (see exercise 7). Plot the histogram for each variable and add the density of the standard normal distribution to each histogram. Are the variables normally distributed?
2. Compute the correlation matrix. Are the variables correlated?
3. Plot the 4×4 matrix of scatterplots (use the command `pairs`). Are the variables independent?
4. Compute the sum $Y = V1+V2+V3+V4$ and plot the histogram of Y. Is the sum normally distributed?

8.3 Gaussian and uncorrelated, but dependent

Let $X \sim N(0,1)$ and define

$$Y = U \cdot X$$

where

$$U = \begin{cases} -1 & \text{with probability } 0.5 \\ 1 & \text{with probability } 0.5 \end{cases}$$

1. Determine the distribution of Y.
2. Derive the covariance between X and Y.
3. Generate a random sample of size $n = 1000$ from $(X,Y)'$ and show the scatterplot.
4. For the sample, compute the sum $X + Y$ and plot its histogram.

8.4 Delta method

A very important linear transformation is a first order Taylor approximation. First, we consider the univariate case. Let $X \sim N(\mu, \sigma^2)$. Define $Y = f(X)$ where f is differentiable (at least at μ).

1. Write down the first order Taylor approximation of f around μ. Note that Y becomes a linear transformation of X.
2. Determine the approximate distribution of Y.
3. Under what conditions do you expect the approximation to be accurate?
4. Now turn to the multivariate case and let $X \sim N(\mu, \Sigma)$ be a random vector of length K. Define $Y = f(X)$ where f is a scalar valued differentiable function.[4] Denote the gradient of f as D_f. Write down the first order Taylor approximation of f around μ.
5. Determine the approximate distribution of Y.

[4] Of course, one could also consider vector-valued functions.

9 Stochastic convergence and limit theorems

9.1 Law of large numbers

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of arbitrarily distributed random variables with finite variance σ^2. Define the sequence of random variables

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

1. Write an R program to illustrate the law of large numbers.
2. Now suppose that the sequence $X_1, X_2, \ldots$ is an $AR(1)$ process:

$$(X_i - \mu) = \rho\,(X_{i-1} - \mu) + \varepsilon_i$$

where $\varepsilon_i \sim iid(0, \sigma_\varepsilon^2)$ is not necessarily normally distributed and $|\rho| < 1$. Show that the law of large numbers still holds despite the intertemporal dependence.

9.2 Law of large numbers for the variance

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of arbitrarily distributed random variables with mean μ, variance σ^2, and finite kurtosis, i.e. $E(X_i^4) < \infty$. Define the sequence of random variables

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2, \quad \text{where } \bar{X} = n^{-1}\sum_{i=1}^{n} X_i.$$

1. Write an R program to illustrate that $S_n^2 \to \sigma^2$ in probability.
2. Now draw the samples from a t-distribution with 3 degrees of freedom, i.e. $X_i \stackrel{iid}{\sim} t_3$. The kurtosis of the t_3-distribution is infinite. Use your R program to show that S_n^2 does no longer converge to σ^2 in probability.

9.3 Central limit theorem

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of arbitrarily distributed random variables with mean μ and finite variance σ^2. Define the sequences of random variables

$$Y_n = \sum_{i=1}^{n} X_i, \qquad Z_n = \sqrt{n}\frac{\left(\frac{1}{n}Y_n\right) - \mu}{\sigma}.$$

1. Write an R program to illustrate the central limit theorem.
2. Show that the central limit theorem still holds if we replace the standard deviation σ in the denominator of Z_n by the estimated standard deviation

$$S_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}.$$

3. Now let $X_1, X_2, \ldots$ be an i.i.d. sequence of t-distributed random variables with 1.5 degrees of freedom. Show that the convergence in distribution breaks down.

9.4 Central limit theorem for dependent data

Suppose that the sequence $X_1, X_2, \ldots$ is an $AR(1)$ process, i.e.

$$(X_i - \mu) = \rho\,(X_{i-1} - \mu) + \varepsilon_i$$

where $\varepsilon_i \sim iid(0, \sigma_\varepsilon^2)$ is not necessarily normally distributed and $|\rho| < 1$.

1. Show that X_i has mean equal to μ and finite variance equal to $\sigma_\varepsilon^2/(1-\rho^2)$.
2. To derive the asymptotic distribution of the mean, do the following steps:
   (a) Derive the asymptotic distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i$
   (b) Show that
   $$\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i = \sqrt{n}\left[(1-\rho)\left(\frac{1}{n}Y_n - \mu\right) + \rho\left(\frac{X_n - X_0}{n}\right)\right]$$
   with $Y_n = \sum_{i=1}^n X_i$.
   (c) Show that
   $$plim\left[\frac{\rho}{1-\rho}\left(\frac{X_n - X_0}{\sqrt{n}}\right)\right] = 0$$
   Hint: Use Tchebychev's Inequality.
   (d) Put your results of (a),(b) and (c) together and derive the asymptotic distribution of the sample mean. That is, show that
   $$Z_n = \sqrt{n}\frac{\left(\frac{1}{n}Y_n\right) - \mu}{\sigma} \xrightarrow{d} U \sim N(0,1)$$
   for $\sigma = \sqrt{\sigma_\varepsilon^2/(1-\rho)^2}$.
3. Write an R program to demonstrate the central limit theorem for the AR(1) process.

9.5 Limits of maxima (I)

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of standard normally distributed random variables. Define the random variable

$$M_n = \max_{i=1,\ldots,n} X_i$$

and its normalized version $R_n = (M_n - d_n)/c_n$ where

$$\begin{aligned} d_n &= \sqrt{2\ln n} - \frac{\ln(4\pi) + \ln\ln n}{2\sqrt{(2\ln n)}} \\ c_n &= (2\ln n)^{-1/2}. \end{aligned}$$

1. Write an R program to illustrate that R_n converges in distribution.
2. The limit distribution of R_n is the Gumbel distribution. Add the Gumbel density $\exp\left(-x - e^{-x}\right)$ to a histogram of R_n.

9.6 Limits of maxima (II)

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of t-distributed random variables with 1.5 degrees of freedom. Define the random variables

$$M_n = \max_{i=1,\ldots,n} X_i$$

and its normalized version $R_n = M_n/c_n$ with

$$c_n = F_{t_{1.5}}^{-1}\left(1 - \frac{1}{n}\right)$$

where $F_{t_{1.5}}^{-1}$ is the quantile function of the $t_{1.5}$-distribution (see the R command `qt`).

1. Write an R program to illustrate that R_n converges in distribution.
2. The limit distribution of R_n is the Frechet distribution (with tail index 1.5). Add the Frechet density $1.5x^{-2.5}\exp\left(-x^{-1.5}\right)$ to a histogram of R_n.

9.7 Limits of maxima (III)

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables uniformly distributed on the interval $[0, 1]$. Define the random variables

$$M_n = \max_{i=1,\ldots,n} X_i$$

and its normalized version $R_n = (M_n - d_n)/c_n$ where

$$\begin{aligned} d_n &= 1 \\ c_n &= \frac{1}{n}. \end{aligned}$$

1. Write an R program to illustrate that R_n converges in distribution.
2. The limit distribution of R_n is the Weibull distribution. Add the Weibull density $\exp(x)$ to a histogram of R_n.

10 Estimators and their properties

10.1 Counter examples

Let $X_1, X_2, \ldots$ be a sample from some random variable X with $E(X) = \mu$ and $Var(X) = 1$. While the variance is known, we would like to estimate the expectation μ.

1. Give an example of an estimator that is unbiased but inconsistent.
2. Give an example of an estimator that is biased but consistent.
3. Give an example of an estimator that is asymptotically biased but consistent.

11 Least Squares and Method of Moments

11.1 Nonlinear least squares

1. Consider the exponential model
$$y_i = \exp(\alpha + \beta x_i) + u_i$$
where $u_i \sim N(0, \sigma^2)$. Since the error term is additive one cannot simply take logarithms to make the model linear. Load the dataset `expgrowth.csv` from the course site and estimate the parameters α and β by minimizing
$$\sum_{i=1}^{n} (y_i - \exp(a + bx_i))^2$$
numerically with respect to a and b.

2. Consider the following example from Davidson and MacKinnon (2004),
$$y_t = \beta_1 + \beta_2 x_{t1} + \frac{1}{\beta_2} x_{t2} + u_t.$$
Assume that $u_t \sim N(0, 1)$. Load the dataset `DMacK1.csv` and estimate the parameters β_1 and β_2.

11.2 Method of moments for the binomial distribution

Consider the binomial distribution $Binom(n, \theta)$ with parameters $n > 0$ and $0 < \theta < 1$. The expectation and variance of $X \sim Binom(n, \theta)$ are
$$\begin{aligned} E(X) &= n\theta \\ Var(X) &= n\theta(1-\theta). \end{aligned}$$
Derive the moment estimators of n and θ. Ignore the restriction $n \in \mathbb{N}$.

11.3 Method of moments for the geometric distribution

Consider the geometric distribution with parameter λ. The expectation of $X \sim Geom(\lambda)$ is $E(X) = 1/\lambda$.

1. Give a moment estimator of λ.
2. Explain why the moment estimator is biased.
3. Explain why the moment estimator is consistent.

11.4 Method of moments for the Gumbel distribution

Consider the Gumbel distribution (also called extreme value distribution) with parameters $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$. The expectation and variance of $X \sim Gumbel(\alpha, \beta)$ are
$$\begin{aligned} E(X) &= \alpha + 0.5772 \cdot \beta \\ Var(X) &= \frac{1}{6}\beta^2\pi^2. \end{aligned}$$
Derive the moment estimators.

11.5 Method of moments for the Pareto distribution

The Pareto distribution has two parameters, $K \geq x > 0$ and $\alpha > 0$ and density $f_X(x) = \alpha K^\alpha x^{-\alpha-1}$. The expectation and variance of $X \sim Pareto(K, \alpha)$ are

$$\begin{aligned} E(X) &= \frac{\alpha K}{\alpha - 1} \\ Var(X) &= \frac{\alpha K^2}{(\alpha-2)(\alpha-1)^2}. \end{aligned}$$

1. Derive the moment estimators. What happens if $\alpha < 2$?
2. Write an R program to simulate the distribution of the moment estimator of $\alpha = 5$ (with $K = 1$ fixed). Generate $R = 10000$ samples $X_1, \ldots, X_n$ of size $n = 100$ each. What happens if you increase the sample size to $n = 1000$? What happens if you consider an $\alpha < 2$?

11.6 Method of moments for the uniform distribution

1. Consider the uniform distribution with parameters a and b (where $b > a$). The expectation and variance of $X \sim unif(a, b)$ are

   $$\begin{aligned} E(X) &= \frac{a+b}{2} \\ Var(X) &= \frac{(b-a)^2}{12}. \end{aligned}$$

   Derive the moment estimators.
2. Write an R program to simulate the distribution of the moment estimators of $a = 0$ and $b = 1$. Generate $R = 10000$ samples $X_1, \ldots, X_n$ of size $n = 40$ each. Are the estimators approximately normally distributed? Check if the moment estimator of a is always smaller than (or equal to) the minimum in the sample.

11.7 Method of moments for the linear regression model

Consider the linear regression model under standard assumptions

$$y = X\beta + u.$$

Left-multiply the model equation by X' and take expectations. Show that the method of moment estimator of β is identical to the OLS estimator.

12 Maximum likelihood estimation

12.1 Extreme values

Let $X \sim Pareto(K, \alpha)$ where the parameter $K \geq x > 0$ is known but the tail parameter α is unknown. The density function of Pareto distribution is

$$f_X(x) = \alpha K^{\alpha} x^{-\alpha-1}.$$

1. Derive the maximum likelihood estimator of α.
2. The Pareto distribution is an excellent approximation of large daily stock return losses (of, say, more than 2%). Load the dataset `daxreturns.csv`. It contains the daily DAX returns (in %) from 16/7/2001 to 13/7/2011 (without holidays). Multiply all DAX returns by (-1) in order to make losses positive, delete all losses that are smaller than 2%, and estimate the tail parameter α for the remaining observations.
3. Plot the likelihood of the observations as a function of α.

12.2 Parameters of the uniform distribution

Consider the uniform distribution on the interval $[a, b]$ with density

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{else.} \end{cases}$$

1. Derive the maximum likelihood estimators of a and b.
2. Write an R program to generate $R = 10000$ sample of size $n = 100$ each. For each sample compute and store the maximum likelihood estimates $\hat{a}$ and $\hat{b}$. Plot their histograms.

12.3 Censored lognormal distribution

Let $X \sim LN(\mu, \sigma^2)$ and let $X_1, \ldots, X_n$ be a sample drawn from X. The X_i are not observable. Instead one can only observe

$$Y_i = \begin{cases} X_i & \text{if } X_i < c \\ c & \text{if } X_i \geq c \end{cases}$$

where c is a known constant. The likelihood of $Y_1, \ldots, Y_n$ is the product of all densities $f_X(y_i)$, for observations with $Y_i < c$, times the product of all probabilities that $Y_i = c$ for observations with $Y_i = c$.

1. Write an R function that computes the likelihood of μ and σ^2 given the observations $Y_1, \ldots, Y_n$ (and given c).
2. Load the dataset `censoredln.csv`.
3. Numerically maximize the likelihood function. The censoring value is $c = 12$.
4. Compute the asymptotic covariance matrix of $\hat{\mu}$ and $\hat{\sigma}^2$.

12.4 Exponential model

Consider the exponential model

$$y_i = \exp(\alpha + \beta x_i) + u_i$$

and load the dataset `expgrowth.csv` from the course site.

1. Assume that the error terms are i.i.d. and $u_i \sim N(0, \sigma^2)$. Write an R function that calculates the log-likelihood of α, β and σ^2.
2. Numerically find the maximum likelihood estimates of α, β and σ^2. Compare your results with exercise 11.1.
3. Compute the asymptotic covariance matrix of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$.
4. Assume that the error terms are i.i.d. with known density

$$f_{u_i}(u) = \frac{1}{2} \exp(-|u|).$$

   Numerically find the estimates of α and β.

12.5 Tobit model

The Tobit model is a linear regression model where observations are censored from below at zero. A latent (unobservable) variable y_t^* is assumed to depend linearly on a vector x_t of exogenous variables,

$$y_t^* = x_t' \beta + u_t$$

where $u_t \sim N(0, \sigma^2)$. The observations are

$$y_t = \begin{cases} y_t^* & \text{if } y_t^* > 0 \\ 0 & \text{else.} \end{cases}$$

1. Given the vector of exogenous variables x_t, derive the probability that $y_t = 0$.
2. The likelihood of $y_1, \ldots, y_T$ is the product of all densities $f_{y_t}(y_t)$, for observations with $y_t > 0$, times the product of all probabilities that $y_t = 0$ for observations with $y_t = 0$. Derive the log-likelihood.
3. Load the dataset `tobitbsp.csv`. The dataset contains the observed endogenous variable `y` and three exogenous variables `x1`, `x2`, `x3` (where `x1` is just a vector of ones). The data are simulated but have similar means, crossproducts etc. as the data in "Estimation of Relationships for Limited Dependent Variables" by James Tobin, *Econometrica*, 26 (1958) 24-36.[5] Numerically compute the maximum likelihood estimates $\hat{\beta}$ and $\hat{\sigma}^2$.
4. Estimate an OLS regression without taking into account the censoring at zero. Compare the OLS estimates with the Tobit estimates.
5. Compute the standard errors of $\hat{\beta}$ and $\hat{\sigma}^2$.

[5]This was the first article to use Tobit estimation (although the name was coined later); it can be downloaded from the course site.

12.6 Probit model

Suppose the endogenous variable y_t can only take two values,

$$y_t = \begin{cases} 1 & \text{with probability } p_t \\ 0 & \text{with probability } 1 - p_t. \end{cases}$$

We would like to model the probability p_t as a function of a vector of exogenous variables x_t. In particular, assume that the probability of $y_t = 1$ equals the value of the cdf of $N(0,1)$ at $x_t'\beta$:

$$p_t = \Phi\left(x_t'\beta\right) = \int_{-\infty}^{x_t'\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz.$$

1. Derive the log-likelihood of β. The distribution function of $N(0,1)$ is `pnorm` in R.
2. Load the dataset `mroz.csv`. The file contains the data used in the article "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions" by Thomas Mroz, *Econometrica*, 55 (1987) 765-799.[6]

   Use `inlf` ("**in l**abour **f**orce") as endogenous variable and `nwifeinc`, `educ`, `exper`, `exper2`, `age`, `kidslt6`, and `kidsge6` as exogenous variables. Add a vector of constants (ones). Numerically compute the maximum likelihood probit estimate $\hat{\beta}$.
3. Interpret the parameter estimates.
4. Predict the probability of $inlf = 1$ for a woman with the following covariates

   nwifeinc = 30, educ = 14, exper = 10, age = 44, kidslt6 = 0, kidsge6 = 3.
5. Calculate the standard errors of $\hat{\beta}$. Is $\hat{\beta}_{educ}$ significantly different from zero?
6. Suppose that the true distribution of the disturbances is not $N(0,1)$ but a uniform distribution on the interval $[-1,1]$. Write a simulation program to show that the maximum likelihood estimator is no longer consistent under this kind of misspecification.

12.7 Logit model

Suppose the endogenous variable y_t can only take two values,

$$y_t = \begin{cases} 1 & \text{with probability } p_t \\ 0 & \text{with probability } 1 - p_t. \end{cases}$$

We would like to model the probability p_t as a function of a vector of exogenous variables x_t. In particular, assume that the probability of $y_t = 1$ equals the value of the logistic function at $x_t'\beta$:

$$p_t = \Lambda\left(x_t'\beta\right) = \frac{\exp\left(x_t'\beta\right)}{1 + \exp\left(x_t'\beta\right)}.$$

1. Derive the log-likelihood of β. In R, the distribution function Λ of the logistic distribution is computed by `plogis`.
2. Redo the application of exercise 12.6.2. Numerically compute the maximum likelihood logit estimate $\hat{\beta}$.
3. Predict the probability of $y = 1$ for the values of the exogenous variables given in exercise 12.6.4.
4. Calculate the standard errors of $\hat{\beta}$. Is $\hat{\beta}_{educ}$ significantly different from zero (at significance level 0.05)?

[6] The article can be downloaded from the course site. The data are available on the internet site of Jeffrey Wooldridge, Econometric Analysis of Cross Section and Panel Data, 2nd ed., 2010. A short description of the dataset can be found on the course site.

12.8 Heckman regression

If the sample is not selected randomly standard OLS methods cease to be consistent. Consistent estimators are, however, still possible. We consider the following simple sample selection model (see Davidson and MacKinnon, 2004, p. 486),

$$\begin{bmatrix} y_t^* \\ z_t^* \end{bmatrix} = \begin{bmatrix} X_t\beta \\ W_t\gamma \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}, \qquad \begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix} \right)$$

where X_t and W_t are vectors of exogenous variables, β and γ are unknown parameter vectors, σ is the standard deviation of u_t and ρ is the correlation between u_t and v_t. Both y_t^* and z_t^* are latent (unobservable). Actually observed are

$$\begin{array}{l} y_t = y_t^* \\ z_t = 1 \end{array} \quad \text{if } z_t^* > 0$$

and

$$\begin{array}{l} y_t \text{ unobserved} \\ z_t = 0 \end{array} \quad \text{if } z_t^* \leq 0.$$

1. Derive the log-likelihood of $(\beta, \gamma, \rho, \sigma)'$. Hint: If y_t is not observed, its contribution to the log-likelihood is $\ln P(z_t = 0)$, else it is $\ln P(z_t = 1) f(y_t^* | z_t = 1)$.
2. Load the dataset `mroz.csv` (see exercise 12.6). Use `hours` as endogenous variable y_t and `inlf` as selection variable z_t. Define the vectors
$$W_t = \begin{bmatrix} \text{kidslt6} \\ \text{kidsge6} \\ \text{age} \\ \text{educ} \end{bmatrix}, \quad X_t = \begin{bmatrix} \text{wage} \\ \text{age} \\ \text{age}^2 \\ \text{educ} \end{bmatrix}$$
and numerically compute the maximum likelihood estimates $\hat{\beta}, \hat{\gamma}, \hat{\rho}$ and $\hat{\sigma}$.
3. Calculate the standard errors for all parameters.

12.9 Count data

The standard multiple linear regression model is not working properly if the endogenous variable y_i takes on only small integer values. In this case one should use "count data" regression methods. We consider a fictional application of the simplest count data regression model – the Poisson regression model. Let y_i denote the number of goals scored by soccer player i (e.g. during a championship). Assume that y_i has a Poisson distribution with probability function

$$P(y_i = k) = \frac{e^{-\mu_i}\mu_i^k}{k!}.$$

The parameter μ_i of the Poisson distribution depends on exogenous variables such that,

$$\mu_i = \exp(X_i'\beta).$$

The vector of exogenous variables X_i includes a constant of unity, position (1=striker, 0=else), age, age^2, training time (in hours per week), fixed salary, and goal bonus (both in 1000 Euro).

1. Load the artificial dataset `players.csv`. It contains information about 300 players.
2. Write an R program to estimate the vector of coefficients β by maximum likelihood.
3. Compute $\hat{\beta}$ and its standard errors.
4. What is the probability that a striker aged 25 scores more than 3 goals if he is training 15 hours per week, has a fixed salary of 700,000 Euro and receives no bonuses.

12.10 Stochastic frontier analysis

See Greene, 2008, section 16.9.5a. Consider the Cobb-Douglas production function

$$y = Ax_1^{\alpha} x_2^{\beta}.$$

By definition, the production function returns the maximal output for given inputs, and actual production cannot be larger than y. Due to inefficiencies, actual production could be modeled (in logs) as

$$\ln y = \ln A + \alpha \ln x_1 + \beta \ln x_2 - u$$

where u is a non-negative random variable. Since other disturbances (e.g. measurement errors) can enter the production function, it is more common to add another, symmetrically distributed, disturbance term,

$$\ln y = \ln A + \alpha \ln x_1 + \beta \ln x_2 - u + v.$$

Assume that $u \sim Exp\,(\lambda)$ and $v \sim N(0, \sigma^2)$ are independent.

1. Show that the density function of $\varepsilon = v - u$ is

$$f_{\varepsilon}\,(x) = \frac{\lambda}{2} \exp\left(\lambda x + \frac{1}{2}\lambda^2\sigma^2\right) \Phi\left(\frac{-x}{\sigma} - \lambda\sigma\right)$$

   where Φ is the cdf of $N(0, 1)$. Hint: $f_{v-u}(x) = \int_0^{\infty} f_v\,(u + x)\, f_u\,(u)\, du$.

2. Write an R program to estimate the parameters $A, \alpha, \beta, \lambda$ and σ by maximum likelihood.

3. Load the dataset `sfa.csv`. This dataset is an abbreviated version of table F7.2 of Greene, 2008. The original data appeared in Zellner and Revankar, "Generalized Production Functions", *Review of Economic Studies*, 36 (1969) 241-250. Reported is the value added in the transportation equipment manufacturing industries of 25 US states and capital and Labour inputs. Compute the ML estimates and their standard errors.

4. Tabulate the estimated inefficiencies for the 25 states.

12.11 ARCH models

Models with autoregressive conditional heteroscedasticity have many applications in empirical finance. We only consider the simple case of an $ARCH(1)$-process. Let X_t denote the stock return in period t. Suppose

$$X_t = \sigma_t \varepsilon_t$$

with $\varepsilon_t \sim N(0, 1)$ and

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2.$$

1. Factorize the joint density function of $X_1, \ldots, X_T$.

2. Ignore the marginal density of X_1 and write an R function to compute the log-likelihood of $X_2, \ldots, X_T$.

3. Load the (artificial) dataset `arch1bsp.csv` and estimate ω and α by maximizing the log-likelihood numerically.

4. Compute the covariance matrix of $\hat{\omega}, \hat{\alpha}$.

12.12 Duration models

There is a huge number of duration models, but we only consider a particularly easy case, see Davidson and MacKinnon, 2004, pp. 490ff. Suppose that how long a state endures is measured by a non-negative random variable T with density function $f(t)$ and cdf $F(t)$. Define the survival function $S(t) = 1 - F(t)$ and the hazard function

$$h(t) = \frac{f(t)}{S(t)}.$$

The hazard function can be interpreted as the probability that the state ends in the next instant, given it has not ended yet.

1. Let T have the cdf $F(t;\theta,\alpha) = 1 - \exp\left(-(\theta t)^{\alpha}\right)$ with parameters θ and α. Derive the density $f(t)$, the survival function $S(t)$ and the hazard function $h(t)$.
2. Assume that n completed (independent) durations t_1,..,t_n have been observed. Derive the log-likelihood function. Use $f(t) = h(t)S(t)$ to split the log-likelihood function into two sums.
3. Suppose the parameter θ depends on some exogenous vector X_i in the following way,

$$\theta_i = \exp(X_i'\beta).$$

   Rewrite the log-likelihood accordingly.
4. If some spells are incomplete (i.e. they have not ended yet) the log-likelihood can be adapted easily by simply dropping their contributions to the hazard part of the log-likelihood.
5. Load the artificial dataset `spells.csv`. The first variable is the duration, the other three variables are exogenous (one is the intercept). Spells with duration 0.5 are incomplete. Estimate the parameters $\beta_1, \beta_2, \beta_3$, and α and their standard errors by maximum likelihood.

12.13 Ultra-high-frequency data

A model of the duration between individual transactions on stock exchanges has been suggested by Engle and Russell, "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data", *Econometrica*, 66 (1998) 1127-1162. The article can be downloaded (password protected pdf) from the internet site of this course. Let X_i denote the duration between transaction $i-1$ and transaction i. The model assumes that

$$X_i \sim \psi_i \varepsilon_i$$

where ε_i is i.i.d. standard exponentially distributed with density function e^{-x}. The scale parameter depends on previous observations in a way similar to ARCH models

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j X_{i-j}.$$

For simplicity, we set $p = 1$.

1. Factorize the joint density function of $X_1, \ldots, X_T$.
2. Ignore the marginal density of X_1 and write an R function to compute the log-likelihood of $X_2, \ldots, X_T$.
3. Load the (artificial) dataset `acd1bsp.csv` and estimate ω and α_1 by maximizing the log-likelihood numerically.
4. Compute the covariance matrix of $\hat{\omega}, \hat{\alpha}_1$.

12.14 Spatial dependence

Observations may not only be dependent over time, but also over space. For instance, real estate prices can be influenced by prices in neighboring regions. A simple case of spatial dependence is the spatial autoregressive model,

$$y = \rho W y + \alpha + \delta z + u \tag{1}$$

where y is an $(n \times 1)$-vector of endogenous variables, W is a symmetric $(n \times n)$-weight matrix, Z is an $(n \times 1)$-vector of a (single) exogenous variable, $u \sim N\left(0, \sigma^2 I\right)$ is an $(n \times 1)$-vector of disturbances. The unknown parameters of the model are α, δ, ρ, and σ, the spatial autocorrelation is driven by the parameter ρ. The weight matrix W can be specified in a number of ways. Often, element W_{ij} simply indicates if regions i and j are direct neighbors, $W_{ij} > 0$, or not, $W_{ij} = 0$. If m_i is the number of direct neighbors of i, then $W_{ij} = 1/m_i$, such that $\sum_j W_{ij} = 1$. Since the model (1) cannot be estimated consistently by OLS, we perform a maximum likelihood estimation of the parameters.

1. Solve (1) for y and derive its multivariate normal distribution (ie. its expectation vector and its covariance matrix).

2. Use the multivariate distribution of y to show that the log-likelihood function is

$$-\frac{n}{2} \ln\left(2\pi\sigma^2\right) + \ln\left(\det\left(I_n - \rho W\right)\right) - \frac{\left(y - \rho W y - \alpha - \delta z\right)' \left(y - \rho W y - \alpha - \delta z\right)}{2\sigma^2}$$

   Hints: If $X \sim N(\mu, \Sigma)$ is multivariate normal with K dimensions, then its density at $x = (x_1, \ldots, x_K)'$ is $f_X(x) = (2\pi)^{-K/2} \left[\det(\boldsymbol{\Sigma})\right]^{-1/2} \cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mu)\right)$. The following results for determinants (of suitable matrices) may also help: $\det(AB) = \det A \det B$, $\det(aA) = a^n \det A$, where A is $n \times n$ and a is a real scalar, and $\det(A^{-1}) = \det(A)^{-1}$.

3. Load the datasets `spatialdata.csv` and `neighbourhood.csv`. The first one contains data on house prices (column 1) and disposable household income (column 2) in the 413 German "Kreise"; the second one is the 413×413 neighborhood matrix.

4. Calculate the normalized neighborhood matrix W such that each row of W sums to unity.

5. Write an R program to compute the log-likelihood function and estimate the parameters α, δ, ρ and σ of the model.

6. Compute the standard errors for $\hat{\alpha}$, $\hat{\delta}$, $\hat{\rho}$, and $\hat{\sigma}$. Test if there is significant spatial autocorrelation.

13 Instrumental variables

13.1 The miracle of the instruments

Since instruments are elements of information sets, one can construct an arbitrary number of additional instruments by (nonlinear) transformations of instruments. The following example shows that creating instruments "out of nothing" is possible but does not work very well in practice. Consider the following simple linear model,

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

for $t = 1, \ldots, T$. The error term u_t is correlated with both exogenous variables x_{1t}, x_{2t} but uncorrelated with an instrument variable w_t,

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ u_t \\ w_t \end{pmatrix} \sim N\left(\begin{bmatrix} 5 \\ 5 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 & 0.3 & 0.5 & 0.7 \\ 0.3 & 1 & 0.5 & 0.7 \\ 0.5 & 0.5 & 1 & 0 \\ 0.7 & 0.7 & 0 & 1 \end{bmatrix}\right).$$

1. Activate the packages `MASS` and `AER`. Generate a sample of size $n = 1000$ from the multivariate normal distribution using the `mvrnorm` command of the `MASS` package. Show that the command `ivreg` (of the `AER` package) does not work as there is only one instrument but two endogenous regressors.

2. Write a program that performs the following steps.

- Create an empty matrix Z with $R = 1000$ rows and 3 columns.
- Start a `for`-loop over $r = 1, \ldots, R$.
- Inside the loop, generate a sample of size $n = 1000$ from the multivariate normal distribution using the `mvrnorm` command of the `MASS` package.
- Use the columns for x_1, x_2 and u to compute the values of the endogenous variable

$$y_t = 1 + 2x_{1t} + 3x_{2t} + u_t.$$

- Use the column for w to create *two* instruments w_1 and w_2,

$$\begin{aligned} w_{1t} &= w_t^2 \\ w_{2t} &= w_t^3. \end{aligned}$$

- Use the command `ivreg` of the `AER` package to compute the IV estimation. Save the coefficient estimates $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$ in row r of the matrix Z.
- End the loop.
- Compute the median of the three estimates $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$.
- Compute the standard errors of the estimates.
- Split the screen using the command `par(mfrow=c(3,1))` and plot the three histograms.

13.2 Linear combinations of instruments

This exercise is close to exercise 8.2 of Davidson and MacKinnon (2004). Consider the simple IV estimator $\hat{\beta}_{IV}$, computed first with an $T \times K$ matrix W of instruments, and then with another $T \times K$ matrix WJ, where J is a $K \times K$ nonsingular matrix. Show that the two estimators coincide. Hence, if the model is just identified, linear combinations of the K instruments have no effect.

13.3 Compulsory School Attendance

This exercise is a replication of some parts of the article "Does Compulsory School Attendance Affect Schooling and Earnings?" by Angrist and Krueger, *Quarterly Journal of Economics* 106 (1991) 979-1014.

1. Load the Stata dataset `AngristKrueger1991Data.dta`. For persons born between 1930Q1 and 1939Q4, plot the years of education against the year of birth (see Figure I in the article). Do the same for persons born between 1940Q1 and 1949Q4 (see Figure II).
2. For 1930Q1 until 1949Q4, plot the mean log weekly earnings against the year of birth (see Figure V).
3. From now on, we only consider persons born between 1920Q1 and 1929Q4. Drop all other observations.[7] Regress the log weekly earnings on the years of education and a set of nine dummies for the year of birth[8] using the OLS command `lm` (see column (1) in Table IV of the article).
4. Compute age as the difference 1970 minus date-of-birth, e.g. a person born in 1925Q3 has age $1970 - 1925.75 = 44.25$. Add age and age-squared to the OLS regression (see column (3) in Table IV).
5. Activate the `AER` package. The `ivreg` command can be used for instrumental variables estimation; its syntax is close to the syntax of the `lm` command, see `?ivreg`.

   The instrumental variables used by Angrist and Krueger are the year of birth, and the year of birth interacting with the quarter of birth. To avoid multicollinearity, one quarter per year has to be dropped from the list of instruments. To economize on time and computer resources, define the instrument variable as a `factor`.[9]

   Estimate an IV regression of log weekly wage on education and year dummies using the instruments of Angrist and Krueger (see column (2) in Table IV).
6. Add age and age-squared to the IV regression (see column (4) in Table IV).

[7] If you have `attach`ed the dataframe, please first delete all variables from your workspace by `rm(list=ls())`. Then re-load the dataset.

[8] The easiest way to deal with the dummy variable is as follows: Create a new variable in the following way: `Dyear <- factor(yob)`. If this variable is included as a regressor in the `lm` command, R will automatically generate the necessary dummy variables.

[9] Suppose the date of birth (`dob`) is given as 1920, 1920.25, 1920.5, 1920.75, ... Then execute the following commands to create a `factor` of instruments:

```
Dq <- dob
Dq[Dq-floor(Dq)==0.75] <- 0
Dq <- factor(Dq)
```

The `factor Dq` can now be used as an instrument, representing all required dummy instruments.

13.4 A simple example

This "simple example" is close to exercise 8.10 of Davidson and MacKinnon (2004). Consider the model

$$\begin{aligned} y_t &= \beta_0 x_t + \sigma_u u_t \\ x_t &= \pi_0 w_t + \sigma_v v_t \end{aligned}$$

with

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

and $t = 1, \ldots, T$. Write an R program to generate at least $R = 1000$ samples for x and y with sample size $T = 10$ using the parameters $\sigma_u = \sigma_v = 1$, $\pi_0 = 1$, $\beta_0 = 0$, and $\rho = 0.5$. For the exogenous instrument $w = (w_1, \ldots, w_T)'$, use independent drawings from the standard normal distribution, and then rescale w so that $w'w$ is equal to T.

For each simulated sample, compute the simple IV estimator (if you use the `ivreg` command of the `AER` package, make sure to drop the intercept by including "`-1`"as a regressor). Then draw the empirical distribution function[10] of the realizations of the estimator on the same plot as the cdf of the normal distribution with mean zero and variance $\sigma_u^2/(T\pi_0^2)$.

In addition, for each simulated sample, compute the OLS estimator, and plot the empirical distribution function of the realizations of this estimator on the same axes as the empirical distribution function of the realizations of the IV estimator.

Redo the exercise for sample size $T = 100$, and – if your computer is fast enough – also for $T = 1000$.

13.5 Money demand

Load the dataset `money.csv`. The data is taken from the web site of Davidson and MacKinnon (2004). The file contains seasonally adjusted quarterly data for the logarithm of real money supply (m_t), real GDP (y_t), and the 3-month treasury bill rate (r_t) for Canada.

1. This is exercise 8.25 of Davidson and MacKinnon (2004). Estimate the model

$$m_t = \beta_1 + \beta_2 r_t + \beta_3 y_t + \beta_4 m_{t-1} + \beta_5 m_{t-2} + u_t$$

   by OLS for the period 1968:1 to 1998:4. Then perform a Durbin-Wu-Hausman test for the hypothesis that the interest rate, r_t, can be treated as exogenous, using r_{t-1} and r_{t-2} as additional instruments.

2. This is exercise 8.26 of Davidson and MacKinnon (2004). Estimate the model by generalized instrumental variables, treating r_t as endogenous and using r_{t-1} and r_{t-2} as additional instruments. Are the estimates much different from the OLS ones?

3. For the IV estimation, perform a test of over-identifying restrictions.

[10] The easiest way to do so is to use the R function `ecdf`, e.g. `plot(ecdf(...))`.

13.6 Tests for the IV model

Load the dataset `fertility.csv` and its description (`fertility.pdf`) from the internet site of the course. The dataset is provided on the internet site of the textbook by Stock and Watson. It is a subset of the data used by Angrist and Evans, "Children and their parents' labor supply: Evidence from exogenous variation in family size", *American Economic Review*, 88 (1998) 450-77. Since some variables included in Angrist and Evans are missing, we cannot reproduce their results exactly.

1. The variable `morekids` indicates if there are more than two children. The variable `samesex` indicates if the first two children are both boys or both girls. Compute the fraction of families that had another child if the first two children were of the same sex, and the fraction if the first two children were of different sex (see Table 3, married women, 1980 data, lower half of the table, in Angrist and Evans).
2. We would like to estimate the causal effect of `morekids` on the number of weeks worked by the mother. Perform an OLS regression of `weeksm1` on `morekids` plus all other variables except `samesex`. Explain why OLS is inappropriate for estimating the causal effect.
3. Explain why the variable `samesex` is a valid instrument for the regression of `weeksm1` on `morekids`.
4. Perform an IV regression of `weeksm1` on `morekids` using `samesex` as instrument.
5. Perform an asymptotic t-test of the null hypothesis that the coefficients of `hispan` and `othrace` are equal. Hint: The estimated covariance matrix of $\hat{\beta}_{IV}$ can be computed by `a$sigma^2*a$cov` where `a` is the object returned by the command `ivreg`.
6. Perform a Wald test of the null hypothesis that the three coefficients of `boy1st`, `boy2nd`, and `hispan` are all equal to zero.

14 GMM

14.1 The R package gmm

Install and activate the R package `gmm`.

1. Read (at least) section 2 of the R vignette "Computing Generalized Empirical Likelihood and Generalized Method of Moments with R" (`gmm_with_R.pdf`) which can be found on the internet site of the course or in the documentation of the package.

2. Explain the relationship between the elementary zero functions f and the functions g used in the `gmm` package.

3. This is the example given in section 3.1 of the R vignette. Suppose you want to estimate the parameters μ and σ of a normal distribution X by GMM using the three moment conditions

$$\begin{aligned} E(X) &= \mu \\ E((X-\mu)^2) &= \sigma^2 \\ E(X^3) &= \mu\left(\mu^2+3\sigma^2\right). \end{aligned}$$

   Write an R function with arguments $\theta = (\mu, \sigma)$ and data X that computes and returns the moment conditions g.

4. Set the random number seed, `set.seed(123)`. Generate $n = 100$ random numbers from the normal distribution $N(4, 2^2)$. Using the starting values $(\mu_0, \sigma_0) = (0, 0)$ run the `gmm` command and save the estimation results in the object `res`. Print `summary(res)` and interpret the output.

14.2 Nonlinear least squares estimation and GMM

Nonlinear least squares estimation is a special case of GMM. Consider the nonlinear regression model

$$y_t = x_t(\beta) + u_t$$

where $x_t(\beta)$ is nonlinear function of the parameters and the data. Assume that the u_t are i.i.d. with $E(u_t) = 0$ and $Var(u_t) = \sigma^2$ and independent of the x_t.

1. Formulate the general model and its least squares estimation in the GMM framework.

2. As a special case, consider the exponential model (see exercise 11.1),

$$y_t = \exp\left(\alpha + \beta x_t\right) + u_t$$

   where $u_t \sim N(0, \sigma^2)$. Load the dataset `expgrowth.csv` from the course site and estimate the parameters α and β and their standard errors by GMM using the command `gmm`. Compare your results with the maximum likelihood estimates computed in exercise 11.1.

14.3 Ordinary least squares estimation and GMM

Ordinary least squares estimation is a special case of GMM. This exercise compares the two uses of the `gmm` package, i.e. explicitly taking into account linearity or not. Consider the linear regression model

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + u_t$$

for $t = 1, \dots, n$. Assume that the standard assumptions are satisfied.

1. Load the dataset `olsgmm.csv`. It contains $n = 100$ (artificial) observations (y_1, x_{11}, x_{21}),..., (y_n, x_{1n}, x_{2n}). Attach the dataframe to make the three variables directly accessible.

2. Estimate the linear regression model using the ordinary least squares command `lm`.

3. Estimate the linear regression model explicitly taking into account linearity. Save the results and compare it to the OLS estimator.

4. Program a function `g1` with two arguments. The first argument is the vector of deep parameters $\theta = (\alpha, \beta_1, \beta_2)$. The second argument is the data matrix. The function `g1` should return the moment conditions as a matrix,

$$\begin{bmatrix} f_{11} & f_{21} & f_{31} \\ f_{12} & f_{22} & f_{32} \\ \vdots & \vdots & \vdots \\ f_{1n} & f_{2n} & f_{3n} \end{bmatrix}$$

where

$$\begin{aligned} f_{1t} &= (y_t - \alpha - \beta_1 x_{1t} - \beta_2 x_{2t}) \\ f_{2t} &= (y_t - \alpha - \beta_1 x_{1t} - \beta_2 x_{2t}) x_{1t} \\ f_{3t} &= (y_t - \alpha - \beta_1 x_{1t} - \beta_2 x_{2t}) x_{2t} \end{aligned}$$

for $t = 1, \dots, n$.

5. Now re-estimate the linear regression model using the `gmm` syntax for general nonlinear models. Set the starting values to (0,0,0) (these are the true values used for simulating the data). Compare the result with the result obtained when taking into account linearity.

14.4 Maximum likelihood estimation and GMM

Maximum likelihood estimation is a special case of GMM. Let $X_1, \dots, X_n$ be a random sample from the random variable X. We know the distributional family of X (e.g. normal distribution) but we do not know the parameters. Denote the density function by f and the parameters by θ.

1. Show that the maximum likelihood estimation can be formulated in the GMM framework. Hint: Use the score function.

2. As a special case, consider the censored lognormal distribution (see exercise 12.3). Let $X \sim LN(\mu, \sigma^2)$ and let $X_1, \dots, X_n$ be an unobserved sample from X. The observations are

$$Y_i = \begin{cases} X_i & \text{if } X_i < c \\ c & \text{if } X_i \geq c \end{cases}$$

where $c = 12$ is a known constant. Load the dataset `censoredln.csv` and estimate the parameters μ and σ and their standard errors by GMM using the command `gmm`. Compare your results with the maximum likelihood estimates computed in exercise 12.3.

14.5 Instrumental variables estimation and GMM

Instrumental variable estimation is a special case of GMM. Consider the linear regression model

$$y = X\beta + u$$

with $u \sim N(0, \sigma^2 I)$. The error term and the regressor matrix X may be correlated but there is a set of instrumental variables W such that $E(u_t|W_t) = 0$.

1. Formulate the general model and the IV estimation in the GMM framework.
2. As a special case, consider the money demand model of exercise 13.5,

   $$m_t = \beta_1 + \beta_2 r_t + \beta_3 y_t + \beta_4 m_{t-1} + \beta_5 m_{t-2} + u_t$$

   with the logarithm of real money supply (m_t), real GDP (y_t), and the 3-month treasury bill rate (r_t) for Canada. Load the dataset `money.csv` and estimate the parameters $\beta_1, \ldots, \beta_5$ by GMM using r_{t-1} and r_{t-2} as instruments for the endogenous regressor r_t.

14.6 Moment conditions and moment existence

Consider the simple linear model without an intercept

$$y_t = \beta x_t + u_t.$$

Assume that x_t has a t-distribution with 3 degrees of freedom and variance 1. The unit variance can be attained by dividing the t-distribution by $\sqrt{3}$, i.e. `rt(n,df=3)/sqrt(3)`. The error terms u_t are independent of x_t; they have a t_3-distribution with variance σ^2. Set $\beta = 0.9$ and $\sigma^2 = 1$.

1. Generate a sample $(x_1, y_1), \ldots, (x_n, y_n)$ of size $n = 100$.
2. Compute the GMM estimates $\hat{\beta}$ and $\hat{\sigma}$ using the moment conditions

   $$\begin{aligned} g_{t1} &= y_t - \beta x_t \\ g_{t2} &= (y_t - \beta x_t)\, x_t \\ g_{t3} &= (y_t - \beta x_t)^2 - \sigma^2. \end{aligned}$$

3. Within a loop $r = 1, \ldots, R$, repeat steps 1. and 2. a large number of times and plot the histogram of $\hat{\sigma}$. Is the distribution of $\hat{\sigma}$ well approximated by a normal distribution?
4. Check if the weighting scheme (`wmatrix="ident"` or `wmatrix="optimal"`) influences the distribution of $\hat{\sigma}$.
5. Change the distribution of x_t and u_t from the t_3-distribution with variance 1 to the standard normal distribution.
6. If your computer is fast enough (or you are willing to wait longer), increase the sample size n and redo this exercise.

14.7 Standard CAPM

In their overview article about GMM applications in finance,[11] Jagannathan et al. (2002) consider the stochastic discount factor representation of the standard capital asset pricing model,

$$\begin{aligned} E\left(m_t R_{it}\right) &= 1 \\ m_t &= \theta_1 + \theta_2 R_{mt} \end{aligned}$$

where $R_{it} = 1 + r_{it}$ is the gross return (and r_{it} the return) of asset i and R_{mt} the market portfolio gross return.

1. Rewrite the standard CAPM such that it fits into the GMM framework, i.e. formulate the moment conditions.
2. Type `data(Finance)` to load the dataset that is included in the `gmm` package. Read `?Finance` to learn about the data structure.
3. Estimate the standard CAPM for the first five companies (i.e. `WMK`, `UIS`, `ORB`, `MAT`, `ABAX`) using the variable `rm` as the (net) market return.
4. Use the function `specTest` to test the overidentifying restrictions.

14.8 Consumption-based CAPM

In their overview article about GMM applications in finance,[12] Jagannathan et al. (2002) consider the moment equations

$$E\left(\left(\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R_{i,t+1} - 1\right) z_t\right) = 0$$

for $i = 1, \ldots, N$. Here, c_t is consumption in period t, $R_{i,t}$ is the gross return of asset i from $t-1$ to t, z_t is a vector of variables known at time t, the parameter β is the time-preference parameter, and the parameter γ is the coefficient of relative risk aversion in the utility function $u(c) = c^{1-\gamma}/(1-\gamma)$.

1. Load the datasets `consumptiondata.csv` and `dax30ann.csv`. The consumption dataset contains information about aggregate consumption levels in current prices from 1970 to 1991 (West Germany) and 1991 to 2010 (Germany). We will only consider variable `V7` (see `LangeReihenKonsum2011Q3.pdf`, page 8). The second dataset contains the start-of-year levels of the DAX30 performance index from 1969 to 2011.
2. Compute the consumption growth rates for West Germany from 1971 to 1991 and for Germany from 1992 to 2010 and concatenate them.
3. Let $z_t = (1, c_t/c_{t-1}, R_{DAX,t})$. Set up the model in the GMM framework and estimate β and γ using the `gmm` package.

[11] Jagannathan, R., Skoulakis, G. and Wang, Z. (2002), Generalized Method of Moments: Applications in Finance, *Journal of Business and Economic Statistics*, 20: 470-481. The password protected article is downloadable from the course site.

[12] Jagannathan, R., Skoulakis, G. and Wang, Z. (2002), Generalized Method of Moments: Applications in Finance, *Journal of Business and Economic Statistics*, 20: 470-481. The password protected article is downloadable from the course site.

14.9 Minimum distance estimation

Consider the following, highly simplified, model of earnings dynamics (F. Guvenen, "An empirical investigation of labor income processes", *Review of Economic Dynamics*, 12 (2009) 58-79),

$$\begin{aligned} y_t^i &= \beta_i t + u_t^i \\ u_t^i &= \rho u_{t-1}^i + \eta_t^i \end{aligned}$$

where y_t^i is log-earnings of person i with t periods of Labour market experience, β_i is an individual specific random effect with variance σ_β^2, ρ is the persistence parameter, and η_t^i are i.i.d. innovations with variance σ_η^2. It can be shown that for $h \geq 2$ the covariance between Δy_t^i and Δy_{t+h}^i is

$$Cov\left(\Delta y_t^i, \Delta y_{t+h}^i\right) = \sigma_\beta^2 - \left[\rho^{h-1}\left(\frac{1-\rho}{1+\rho}\right)\sigma_\eta^2\right]. \tag{2}$$

1. The theoretical $(H+1) \times (H+1)$-covariance matrix of $\left[\Delta y_t^i, \Delta y_{t+1}^i, \ldots, \Delta y_{t+H}^i\right]$ depends on the three unknown parameters σ_β^2, ρ and σ_η^2. The GMM approach to estimation requires that the differences between the elements of the theoretical covariance matrix[13] and its empirical counterparts should be minimized with respect to the parameter vector $\theta = (\sigma_\beta^2, \rho, \sigma_\eta^2)$. Set $H = 10$ and write an R program that can estimate the parameters by GMM using the command `gmm`. Note that the first order covariances must not enter the estimation since (2) is only valid for $h \geq 2$.

2. Load the artificial dataset `logearnings.csv`. The rows are individuals $i = 1, \ldots, N$, the column are the periods $t = 1, \ldots, T$ with $N = 2000$ and $T = 15$. Compute the parameter estimates and their standard errors.

3. Perform a test of the overidentifying restrictions.

[13]Due to the symmetry of the covariance matrix, the elements below (or above) the diagonal are omitted.

15 Indirect inference

15.1 AR(1) processes

The seemingly simple autoregressive process

$$x_t = \rho x_{t-1} + \varepsilon_t$$

with $\varepsilon_t \sim N(0, \sigma^2)$ is sometimes surprisingly hard to estimate. In the following, always use

```
x <- filter(rnorm(n),rho,method="r",init=rnorm(1))
```

to generate a path of length n.

1. Simulate the distribution of the estimator $\hat{\rho}$ for $\rho = 0.8$ and $n = 100$. Use the command `ar` with options `order=1` and `aic=F` to estimate ρ (if you like, try different estimation methods, e.g. `ols` or `mle`).
2. Simulate the distribution of $\hat{\rho}$ for the unit root process with $\rho = 1$.
3. Simulate the distribution of $\hat{\rho}$ for the explosive process with $\rho = 1.01$.
4. Write an R program to estimate the $AR(1)$ parameter ρ by indirect inference. The auxiliary model is, of course, itself an $AR(1)$ process. The number of auxiliary paths should be $H = 10$.
5. Determine the distribution of the indirect inference estimator $\hat{\rho}$ by simulation for values of $\rho = 0.8, 1, 1.01$.

15.2 Filter models using the Kalman filter

Consider the univariate dynamic linear model

$$\begin{aligned} y_t &= \theta_t + v_t, \qquad v_t \sim N(0, V) \\ \theta_t &= \theta_{t-1} + w_t, \qquad w_t \sim N(0, W) \\ \theta_0 &\sim N(m_0, C_0) \end{aligned}$$

where only y_t is observed, but we are interested in the unobservable state variable θ_t. This model is sometimes called random walk plus noise. In R, the package `dlm` provides commands to deal with dynamic linear models. The notation in this exercise is adapted to the `dlm` package.

1. Install and activate the package `dlm`.
2. Define a `dlm`-object `mod <- dlm(FF=1,GG=1,V=9,W=1,m0=0,C0=100)`. This object represents the random walk plus noise model with known parameters V and W (and m_0 and C_0).
3. Load the dataset `rwnoise.csv` and plot the variable `y`.
4. Add the Kalman filtered estimated state variable $\hat{\theta}_t$ to the plot. The Kalman filtered series can very easily be computed by the command `dlmFilter(y,mod)$m[-1]`.
5. In general, the model parameters V and W are unknown. Estimate V and W by indirect inference. The auxiliary parameters are the $MA(1)$ parameter and the error term variance of an $ARIMA(0,1,1)$ model[14] (hence, the model is exactly identified).

[14] To estimate an $ARIMA(p,d,q)$ model in R, use the command `a <- arima(x,order=c(p,d,q))`. The coefficients can be extracted by `a$coef` and error term variance is `a$sigma2`.

15.3 Estimation of the Cox-Ingersoll-Ross model

Cox, Ingersoll and Ross, "A Theory of the Term Structure of Interest Rates", *Econometrica* 53 (1985) 385-407, suggest a continuous-time model for the short-term interest rate. The stochastic process is described by the stochastic differential equation

$$dX_t = (\theta_1 - \theta_2 X_t)\, dt + \theta_3 \sqrt{X_t} dW_t \tag{3}$$

where W_t is a standard Wiener process and $X_0 > 0$.

1. Activate the R package `sde`. Generate and plot a single path on the time interval $[0, 200]$ of an Cox-Ingersoll-Ross process with parameters $\theta_1 = 0.03$, $\theta_2 = 0.5$, and $\theta_3 = 0.08$ and starting value $X_0 = 0.06$ using the command `sde.sim`. Set the number of steps to $N = 200$.
2. Continuous-time models are sometimes estimated by discretizing them in a crude way. The discretized version of (3) is, of course,

   $$X_t = X_{t-1} + (\theta_1 - \theta_2 X_{t-1}) + \theta_3 \sqrt{X_{t-1}} \varepsilon_t \tag{4}$$

   with $\varepsilon_t \sim N(0, 1)$ and starting value $X_0 = \theta_1/\theta_2$. Find estimators for the parameters $\theta_1, \theta_2, \theta_3$ in (4) that can be computed fast (e.g. least squares estimators).
3. Load the dataset `cirpath.csv`. The process is not observed continuously. The dataset only contains observations of X_t at discrete time points $t = 1, \ldots, 200$. Estimate the parameters θ_1, θ_2, and θ_3 by indirect inference with the auxiliary model (4). Assume that the (unobserved) starting value is $X_0 = \theta_1/\theta_2$.

15.4 Ornstein-Uhlenbeck process

Consider the continuous-time stochastic process described by the stochastic differential equation

$$dX_t = \lambda\, (\mu - X_t)\, dt + \sigma dW_t \tag{5}$$

where W_t is a standard Wiener process, $\lambda > 0$ is a parameter of the strength of mean-reversion, μ is the long-run mean, and $\sigma > 0$ is a volatility parameter.

1. Install and activate the R package `sde`. It provides commands to simulate paths of stochastic processes described by stochastic differential equations. Generate and plot a single path on the time interval $[0, 100]$ of an Ornstein-Uhlenbeck process with parameters $\lambda = 0.9$, $\mu = 0$, and $\sigma = 1$ and starting value $X_0 = 2$ using the command `sde.sim`. Note that the parametrization of the `sde` command differs from (5) with $\theta_1 = \lambda\mu$, $\theta_2 = \lambda$, and $\theta_3 = \sigma$. Set the number of steps to $N = 100$.
2. Continuous-time models are sometimes estimated by discretizing them in a rough way. The discretized version of (5) is, of course, $X_t - X_{t-1} = \lambda\, (\mu - X_{t-1}) + \sigma\varepsilon_t$ with $\varepsilon_t \sim N(0, 1)$ and starting value $X_0 = \mu$. Rewriting gives

   $$X_t = \lambda\mu + (1 - \lambda)\, X_{t-1} + u_t$$

   with $u_t \sim N(0, \sigma^2)$. This exercise shows that simply estimating the discrete model can be severely misleading!

   For this create an empty vector `Z` of length $R = 1000$. Write a loop over $r = 1, \ldots, R$ performing the following steps for each replication.

- Generate a path of the Ornstein-Uhlenbeck process `x` given in exercise 1.
- Fit an $AR(1)$ process to the path using the command `ar(x,order=1,aic=F)` or, alternatively, the command `arima(x,order=c(1,0,0))`. Both commands estimate (4), but only `arima` reports the estimated intercept.
- Save the AR coefficient in `Z[r]`. The AR coefficient is the estimate of $1 - \lambda$.
- After the loop, plot the histogram of `Z`. Comment on the distribution of $1 - \hat{\lambda}$.

3. Load the dataset `oupath.csv`. The process is not observed continuously. The dataset only contains observations of X_t at discrete time points $t = 1, \ldots, 100$. Estimate the parameters λ, μ, and σ by indirect inference with the auxiliary model (4). Assume that the (unobserved) starting value is $X_0 = \mu$.

15.5 Time-aggregated observations

Consider the geometric Brownian motion described by the stochastic differential equation

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

where W_t is a standard Wiener process, μ is the drift parameter and $\sigma > 0$ is the volatility parameter, and the starting value is $X_0 = 100$. Suppose the process X_t is not observed continuously. The only observations are the time-aggregates

$$Y_t = \int_{t-1}^{t} X_t dt \tag{6}$$

for $t = 1, \ldots, T$. The Y_t could be interpreted as average stock prices over time intervals $[t-1, t]$, that are relevant for Asian option pricing. This exercise explores how to estimate μ and σ from observations $Y_1, \ldots, Y_T$ by indirect inference.

1. Install and activate the R package `sde`. It provides commands to simulate paths of Brownian motion (`BM`) and geometric Brownian motions (`GBM`). Generate and plot a single path of the geometric Brownian motion X_t with $\mu = 0.00025$ and $\sigma = 0.015$ (these values are more or less realistic for daily stock returns), and starting values $X_0 = 1$, on the time interval $[0, T]$ with $T = 30$. Let the number of steps be $N = 3000$, i.e. 100 steps per period.
2. The time integrals in (6) can be approximated by the sums
$$Y_t \approx \sum_i X_i \cdot \Delta$$
where the sum is over all i between $t - 1$ and t, and $\Delta = T/N = 0.01$ is the interval length. Create an empty matrix `Z` of dimensions $R \times 4$ with $R = 1000$. Write a loop over $r = 1, \ldots, R$ performing the following steps for each replication.
   - Generate a path of the geometric Brownian motion X_t given in exercise 1.
   - Calculate the four integrals Y_1, Y_2, Y_{15}, Y_{30} and save them in row r of the matrix `Z`.
   - After the loop, calculate the means for each column and the variance-covariance matrix of `Z` (using the command `cov`).
3. Load the dataset `timeaggr.csv`. It contains 30 time-aggregated observations $Y_1, \ldots, Y_{30}$. Estimate μ and σ by indirect inference. As auxiliary model use an $ARIMA(1, 0, 1)$ process with intercept. Also include the error term variance of the $ARIMA$ model in your estimation.[15] Assume that the starting value is $X_0 = 1$.

[15] To estimate an $ARIMA(p, d, q)$ model in R, use the command `a <- arima(x,order=c(p,d,q))`. The coefficients can be extracted by `a$coef` and error term variance is `a$sigma2`.

16 Bootstrap

16.1 Omitted variables bias does not go away

This exercise shows that the bootstrap does not help to eliminate omitted variable bias. Reconsider exercise 7.4.

1. Load the dataset `omitted.csv` and estimate the model without the relevant variable X_4 by OLS.
2. Bootstrap the bias and standard error of the coefficients of X_2 and X_3. Set the number of bootstrap replications to $B = 5000$.

16.2 Confidence intervals for the Gini index

Install and activate the package `ineq`. It provides functions for inequality measures and concentration measures as well as Lorenz curves.

1. Load the dataset `earnings.csv`. It contains earnings of 11648 individuals. Compute the Gini coefficient of earnings.
2. Bootstrap the standard error of the Gini coefficient.
3. Compute the bootstrap 0.95-confidence interval for the Gini coefficient using the percentile method.

16.3 Confidence intervals for correlation coefficients

The distribution of the empirical correlation coefficient

$$\hat{\rho} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}$$

is rather complicated except for some special cases. Of course, being based on moments, $\hat{\rho}$ is asymptotically normally distributed (if the relevant moments exist). However, in small samples, confidence intervals for ρ are not trivial.

1. Install and activate the package `copula`. It provides functions and commands to deal with copulas. Generate a single sample of size $n = 1000$ from (X, Y) by executing

   ```
   x <- qexp(rcopula(gumbelCopula(1.3),1000)).
   ```

   Row i of the $(n \times 2)$-matrix `x` is the pair (X_i, Y_i). Plot the sample and compute the correlation coefficient.
2. Simulate the distribution of $\hat{\rho}$ for sample size $n = 50$ and show that the distribution is not normal.
3. Draw a single sample of size $n = 50$. Compute the bootstrap 0.95-confidence interval for ρ using the percentile method with $B = 1000$ bootstrap replications. Check if the true value (about 0.43) is covered by the interval.

16.4 The t-test

This exercise shows that the ordinary t-test is a special case of the parametric bootstrap. Consider the simple linear regression model

$$y_t = \alpha + \beta x_t + u_t, \quad u_t \sim N(0, \sigma^2)$$

with $\alpha = 1$, $\beta = 0$ and $\sigma = 2$. Load the dataset `ttestboot.csv`. It contains the exogenous variable x and the endogenous variable y. The number of observations is $n = 9$. We want to test $H_0 : \beta = \beta_0 = 0$ against $H_1 : \beta \neq 0$.

1. Compute the ordinary OLS test statistic
$$\frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})} \tag{7}$$
and the p-value of the test (both are printed by `summary` of the `lm` object).

2. The parametric bootstrap makes use of the fact that the distribution of u_t is known apart from the variance σ^2. Resamples can be generated by drawing new error terms from the normal distribution $N(0, \hat{\sigma}^2)$ where $\hat{\sigma}^2$ is an unbiased estimate of σ^2. Compute the estimates $\hat{\alpha}$, $\hat{\beta}$ and
$$\hat{\sigma}^2 = \frac{1}{7} \sum_{t=1}^{9} \hat{u}_t^2.$$
Use the function `residuals` to extract the residuals from the `lm` object, and the function `coefficients` to extract the parameter estimates $\hat{\alpha}$ and $\hat{\beta}$.

3. Prepare an empty vector `Z` of length $R = 5000$. Write a loop over $r = 1, \dots, R$ performing the following steps:
   - Draw a random sample $u_1^*, \dots, u_9^*$ from $N(0, \hat{\sigma}^2)$.
   - Compute
   $$y_t^* = \hat{\alpha} + \hat{\beta} x_t + u_t^*, \quad t = 1, \dots, 9.$$
   - For the resample $(x_1, y_1^*), \dots, (x_9, y_9^*)$, calculate the bootstrap test statistic
   $$\frac{\hat{\beta}_r^* - \hat{\beta}}{SE(\hat{\beta}_r^*)}$$
   and save it as `Z[r]`.

4. Plot the histogram of `Z` and add the density function of the t-distribution with `7` degrees of freedom.

5. Calculate the proportion of `abs(Z)` that is larger than the absolute value of the original test statistic (7). Compare your result with the p-value computed above.

16.5 The percentile-t-method

In the lecture, we considered the distribution of $\hat{\theta} - \theta$ and its bootstrap approximation $\theta^* - \hat{\theta}$ to determine confidence intervals. An asymptotically more efficient method is to consider the distribution of

$$\frac{\hat{\theta} - \theta}{SE(\hat{\theta})}$$

and its bootstrap approximation

$$\frac{\theta^* - \hat{\theta}}{SE(\theta^*)}.$$

Since these quantities look like t-statistics, this method is often called the percentile-t-method.

1. Start with
$$P\left(c_1 \leq \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \leq c_2\right) = 1 - \alpha$$
and derive an expression for the $(1-\alpha)$-confidence interval.

2. Start with
$$P\left(c_1 \leq \frac{\theta^* - \hat{\theta}}{SE(\theta^*)} \leq c_2\right) = 1 - \alpha$$
and derive an expression for the bootstrap $(1-\alpha)$-confidence interval.

3. Explain the algorithm to compute the bootstrap confidence interval.

4. Reconsider exercise 13.5. Write a program that computes the bootstrap 0.95-percentile-t-confidence interval for the coefficient of the interest rate r_t estimated with instruments r_{t-1} and r_{t-2} (as done in 13.5.2).

16.6 Heavy tails and variance testing

Let X and Y be independent random variables both having a t-distribution with 3 degrees of freedom (a plausible model for returns). Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample of size $n = 200$.

1. Use simulations to show that the F-test of the hypotheses
$$\begin{aligned} H_0 &: Var(X) = Var(Y) \\ H_1 &: Var(X) \neq Var(Y) \end{aligned}$$
rejects the null hypothesis far too often at significance level $\alpha = 0.05$. Hint: Tests for equal variances can be performed by the R command `var.test(x,y)$p.value`.

2. Implement a nonparametric Wald-type bootstrap test for equality of variances. Use the ordinary F-statistic as test statistic (`var.test(x,y)$statistic`).

3. Implement a nonparametric LM-type bootstrap test for equality of variances. Use the ordinary F-statistic as test statistic (`var.test(x,y)$statistic`).

4. If your time and computer power allow, do a Monte-Carlo simulation to show that the bootstrap tests keeps the prescribed level α more closely than the ordinary F-test, i.e. that the proportion of rejections of the true null hypothesis is closer to 5% of the replications. Note, however, that the rejection probability is still substantially too high for both variants of the bootstrap tests.

16.7 Time series

This exercise shows in a simple setting how to bootstrap time series. In general, this approach works well if there is a parametric time series model based on an underlying white noise process (e.g. GARCH, ARIMA, VAR). Consider the simple $AR(1)$ model with intercept

$$(x_t - \mu) = \rho (x_{t-1} - \mu) + \varepsilon_t \tag{8}$$

with $\varepsilon_t \sim N(0, \sigma^2)$.

1. Load the dataset `ar1bsp.csv`. Estimate the parameters μ and ρ using the `arima` command. The standard errors that are reported are asymptotically valid.
2. Show that bootstrapping the standard errors of $\hat{\mu}$ and $\hat{\rho}$ does not work under the ordinary bootstrap resampling scheme, i.e. drawing $x_1^*, \ldots, x_T^*$ from $x_1, \ldots, x_T$ with replacement.
3. Program the following time series bootstrap approach. Estimate the model (8) for the original sample $x_1, \ldots, x_T$ and save the parameter estimates $\hat{\mu}$ and $\hat{\rho}$ and the residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T$. Initialize an empty $(R \times 2)$-matrix `Z` for, say, $R = 1000$. For $r = 1, \ldots, R$:

   - Draw a resample $\varepsilon_1^*, \ldots, \varepsilon_T^*$ from $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T$.
   - Set $x_1^* = x_1$ and compute $x_t^* = \hat{\mu} + \hat{\rho}\left(x_{t-1}^* - \hat{\mu}\right) + \varepsilon_t^*$ for $t = 2, \ldots, T$.
   - Estimate (8) for $x_1^*, \ldots, x_T^*$ and save the estimates μ^* and ρ^* in row r of `Z`.

   Compute the standard errors for both columns of `Z` and compare them to exercise 1.

16.8 Bootstrap test for the Zipf index of city size distributions

It is well known that the population size distribution of large cities can be approximated by the Zipf distribution which is a special case of the Pareto distribution with tail index $\alpha = 1$. Suppose, $X_1 \geq X_2 \geq \ldots \geq X_n$ is a descendingly ordered sample of city sizes. In regional economics, the tail index α is often estimated from the regression

$$\ln(i) = c - \alpha \ln X_i + u_i$$

where i is the rank of the city and X_i its size, the intercept parameter c is of no interest. Since the sample is ordered, the observations are no longer independent and the optimality properties of OLS vanish. In particular, the ordinary t-test does not work correctly anymore. In this exercise, ordered samples from the Zipf distribution (i.e. from the Pareto distribution with true tail index $\alpha = 1$) are generated by the command

```
x <- sort(exp(rexp(n)),decreasing=TRUE)
```

where `n` is the sample size.

1. Simulate and plot the distribution of $\hat{\alpha}$ for sample size $n = 20$. The regression can be performed by `obj <- lm(log(1:n)~log(x))`. The estimates can then be extracted by the function `coefficients(obj)`.
2. An important hypothesis is $H_0 : \alpha = 1$ against $H_1 : \alpha \neq 1$. Simulate and plot the distribution of the test statistic
$$T = \frac{\hat{\alpha} - 1}{SE(\hat{\alpha})}$$
and show that it is not t_{n-2}-distributed even though H_0 is true.
3. Explain why the simulations done in 1. and 2. can be used to find the critical values of a parametric LM-type bootstrap test of $H_0 : \alpha = 1$.