Financial Risk Forecasting
Seminar 1

Jon Danielsson
London School of Economics

Version 4.0 August 2024

1 R, data and sample statistics

1.1 Introduction

The software we use in this course is R, which is open source and can be installed for free. There are several ways one can use R, and in this course we recommend RStudio. Please download R and RStudio, and familiarise yourself with the basic setup. This introduction to RStudio might be of help.

For this course, it is essential to register for a WRDS (Wharton Research Data Services) account. This is done by going to wrds-www.wharton.upenn.edu/register and completing the requested details. In the department field, enter the module code to ensure easy authorisation by the Library, which is usually done within 2 working days. Do NOT write Department of Finance in the department field as this will delay the account authorisation due to procedures that have to be followed as part of LSE's contracts with suppliers. Once authorised by the Library, your WRDS account will expire on 20 June 2025.

Some of the same data can also be found on the data chapter in the risk forecasting notebook, `stocks.csv`.

For more information about R and especially how it is used in forecasting risk, refer to the R risk forecasting notebook. You can find references to resources for learning R in the notebook.

1.2 The plan for this week

1. Familiarisation with R and RStudio
2. Learn basic R commands
3. Download, import, manipulate and save financial data
4. Create a simple plot
5. Create, customise and export plots

1.3 Basic R commands

We can enter some basic R comments into the `console`, and you see the output immediately printed on the console.

```
1+1
2^5
a=2
b=100
exp(a)+b/10
result=exp(a)+b
cat("the answer is:",log(a),"or",result,"\n")
```

Usually it is best to create a file for all the code you want to execute, perhaps `seminar1.R`. You can then execute the comments in this file either one by one or all at once.

Also, its best to keep the data files you download below in the same directory (folder).

R needs to be able to find your data, and for that you may need to set the `working directory`. You can do that with the RStudio menu item `Session/Set Working Directory` or the R command `setwd()`.

1.4 Downloading, importing and manipulating financial data

We will download data on a number of stocks and manipulate it. The database to use is provided by the Center for Research in Security Prices, and is usually known as CRSP. You will access it through WRDS.

1.5 Ticker, Company Name, PERMNO

There are different ways of identifying a company in CRSP, and we need to be careful with what we choose. It is very common to associate a stock with its ticker, but if the company has a merger, this might be subject to change. For example, if we consider JP Morgan (ticker: JPM), historically it has been officially registered with different names before some mergers and acquisitions happened (Chemical Banking Corp, Chase Manhattan Corp, etc), each which a different ticker, but it is essentially the same company, and by specifying the ticker JPM we would be losing years of financial data. For this reason, we work with the permanent company number, or PERMNO, which is immutable.

For the purpose of this course, we will mostly be working with *comma-separated values* , or *.csv* files and `RData` files.

See the file formats chapter in the risk forecasting notebook for more details on tickers and data formats and financial data chapter for details and issues in fianncial data.

1.6 Downloading the data

Once logged on to WRDS, select CRSP and go to "Stock / Security Files / Daily Stock File".

1. Choose 1 January 1990 to the latest date
2. Select PERMNO and 14593 10107 47896 70519 11850 43449 12060

PERMNO	Latest ticker	Latest Name
14593	AAPL	Apple
10107	MSFT	Microsoft
47896	JPM	JPMorgan Chase
70519	C	Citigroup
11850	XOM	Exxon
43449	MCD	McDonalds
12060	GE	General Electric

3. Select the following information:
   - From the identifying information: Company Name, Ticker
   - From the time series information: Price, Holding Period Return
   - From the distribution information: Cumulative Factor to Adjust Price
4. Use comma-delimited text (.csv file) and default date format (YYYYMMDDn8) for the output

Click "Submit Query" and them the output link at top.

Save the output into a file in the same directory you indeed to keep your work in as `crsp.csv`.

1.7 Variable description

All the details on the variables we download can be found in the *Variable Descripions* section of WRDS. It is important to distinguish between the type of returns we are using, whether we are including dividends or not. For example, the description of the time series variables we have downloaded is:

- PRC: "price is the closing price or the negative bid/ask average for a trading day. If the closing price is not available on any given trading day, the number in the price field has a negative sign to indicate that it is a bid/ask average and not an actual closing price"
- RET: "return is the change in the total value of an investment in a common stock over some period of time per dollar of initial investment. RET(I) is the return for a sale on day I. It is based on a purchase on the most recent time previous to I when the security had a valid price. Usually, this time is I - 1 [...]"

1.8 Importing data into R

Open RStudio and select the directory you chose in the last step as the `Working Directory`. Then.

```r
data <- read.csv('crsp.csv')
dim(data)
head(data)
head(data$RET)
head(data[,6])
names(data)
unique(data$PERMNO)
unique(data$TICKER)
unique(data$COMNAM)
```

1.9 One firm

```r
citi <- data[data$PERMNO == 70519,]
dim(citi)
head(citi)
tail(citi)
unique(citi$TICKER)
highest_citi <- max(citi$RET) * 100
paste0("Highest return for Citi: ", highest_citi, "%")
lowest_citi <- min(citi$RET) * 100
paste0("Lowest return for Citi: ", lowest_citi , "%")
```

1.10 Simple plots

R can easily build plots to visualise data. We use the `plot` command for this.

```r
plot(citi$PRC, type = "l", main = "Price of Citi")
```

Note, these are unadjusted prices.

This is quite ugly, and we discuss ways to improve the plot later. Or see the plot chapter in the risk forecasting notebook.

1.11 Recap

In this seminar we have covered:

- Downloading and working with R and RStudio
- Data types in R, basic operations, accessing elements
- For loops and if statements
- Downloading and importing data from CRSP into R
- Extracting columns from a data frame
- Finding the maximum value of a variable
- Making a simple plot

Some new functions used:

- `dim()`
- `exp()`
- `log()`
- `read.csv()`

- `head()`
- `names()`
- `unique()`
- `paste()` and `paste0()`
- `plot()`

1.12 Optional exercises

1. Download other stock data, see the data chapter in the risk forecasting notebook for some other sources, and repeat the analysis above.
2. Print the maximum and minimum return for all stocks.