Notes on Risk Management

Anton Gerunov, Ph.D.

Notes on Risk Management

Author: Anton Antonov Gerunov

Publisher: Sofia University "St. Kliment Ohridski"

Faculty of Economics and Business Administration

ISBN: 978-954-9399-45-5

Reviewers:

- Prof. Dr. George Mengov
- Prof. Dr. Stefan Petranov

Contents

Preface

Notes on Risk Management is a handbook which aims to outline key theoretical insights about quantitative risk management and demonstrate their applications in a modern software environment. For this purpose, we use the R language for statistical computing which allows us to illustrate theory with abundant empirical examples, based on real-life data. The handbook is structured in 14 lectures, covering both traditional risk management topics (expectations, risk metrics, VaR-type models, etc.), as well as innovative approaches for risk modeling that leverage machine learning methods.

The contents are as follows:

- **Lecture one** is an introduction to risk management which defines key terms, roles, and strategies in this process.
- **Lecture two** outlines risk management based on qualitative data and its transformation into quantitative data for the purpose of more precise modeling.
- **Lecture three** presents the R language for statistical computing and shows its risk management functionalities.
- **Lectures four and five** outline key characteristics of financial markets and demonstrate the empirical correlation between risk and expected return in the context of efficient market ideas and Modern portfolio theory.
- **Lecture six** reviews critically foundational models for managing market risk – the Value at Risk (VaR) model, and the Expected Shortfall, ES (also Expected Tail Loss, ETL) models. We also show their marked sensitivity to input parameters.
- **Lecture seven** is an overview of different statistical distributions that can aid modeling.
- **Lecture eight** introduces Monte Carlo methods as risk management tools and illustrates them through financial options valuation. We show how analytic solutions for option pricing (the Black-Scholes equation) are practically identical with simulation results.
- **Lecture nine** underscores the importance of operational risk and proposes a Monte Carlo-based approach for managing it.
- **Lecture ten** introduces credit risk as an instance of statistical classification problem and illustrates traditional solutions (logistic regression) and machine learning approaches (naïve Bayes classifier, neural network, random forest models, etc.)
- **Lecture eleven** focuses on modeling and forecasting time series, thus demonstrating the rapidly growing risk as the forecast horizon expands.
- **Lecture twelve** comprises a critical evaluation of the model risk which stems from using quantitative models.

- **Lecture thirteen** presents key insights from behavioral economics regarding individual risk perception and interprets quantitatively alternative forms of the utility function. **Lecture fourteen** concludes.

This material can be useful for both advanced graduate students and practicing professionals in the field of risk management.

Lecture One: A Gentle Introduction to Risk

Contemporary organizations operate in a risky context. Unexpected events can lead to major losses in terms of market share, money, or reputation. Planning for those contingencies and creating mitigation tactics is one the key processes a modern business needs to undertake. The course on Risk Management focuses on formally defining what risk is, and presenting different approaches to modeling it.

We will look at risk from a more quantitative perspective which is usual for financial organizations and large corporations but only for the benefit of clarity of exposition. The main principles espoused go well beyond the large organization and the financial markets and can easily be applied in a lot a different contexts - ranging from the small company to personal decisions.

The field of Risk Management (RM) is broad and growing, with a large diversity of positions and roles. The unifying theme happens to be trying to predict the (almost) unpredictable and to minimize losses or maximize benefits. Such an ambitious goal calls for an effective blend between theoretical knowledge and practical skill. This need is also reflected in the structure of the Risk Management course - it combines key theoretical insights with hands-on tasks and problems to solve using state-of-the-art software applications.

The ultimate goal is to present the breadth of the discipline, outline the most poignant debates, focus on some subtleties and present some of the key instruments of the RM toolbox. The reader should turn into an intelligent yet critical user of RM methods and concepts. There are naturally many books that provide a more comprehensive treatment of risk - the reader is especially directed to the work of Crouhy et al. (2006) who provide an excellent and very accessible overview of Risk Management.

Definition of Risk

Casual conversations of risk underpin our everyday understanding of the concept. In this sense, risk is often associated with unfavorable developments (expected or unexpected) that lead to a loss. Risk is therefore associated with downside effects.

This is not the case in the field of economics and finance. Here risk is defined as the **deviation from an expected outcome** which can either be positive, leading to profit (risk on the upside) or negative, leading to loss (risk on the downside). In fact we can have an even more nuanced view on risk and uncertainty by resorting to Frank Knight's (1921) distinction - he delineates four different cases, as follows:

- **Certainty** - where no possibility of deviation is present, and therefore all events happen with a probability of p = 1. This is often the case with natural laws - if an object on Earth is dropped it is supposed to fall to the ground, accelerating at a rate of g.
- **Risk** - this is a situation where deviations from the expectations are possible and those

are well-defined. The agent knows all the possible outcomes and can attach probabilities to them. It is therefore a situation where the mathematical expectation *E[x]* can be defined and its respective probabilities p_i are clear. For example when a fair coin is thrown, the outcomes can be either heads or tails, and each of them occurs with the probability of 0.5.

- **Uncertainty** - in this case the agents know the possible outcomes of a given situation but they cannot easily attach probabilities to them. For example, when building a nuclear power plant, one can assume that it either operates safely, there are minor accidents, or there is a large accident. While these outcomes are easy to spell, it is unclear as to what their exact probabilities are. In such cases the agents tend to attach approximate probabilities or formally model the system in order to obtain estimates of its behavior.
- **Ambiguity** - under ambiguity agents can defined neither the outcomes of a given situation, nor the probabilities associated with them. This is by far the most difficult situation of decision making and hardly lends itself to formal modeling. Some scholars compare situations of low-level tactical warfare to this case.

To ease modeling and interpretation, we will take consider situations as instances of *risk* even though this will sometimes not be fully accurate. In cases of uncertainty and ambiguity, we will try to define outcomes and approximate probabilities and use standard tools for modeling and managing risk. In the case of certainty modeling is, of course, trivial.

Types of Risks

There are many situations in which risks occur. The understanding and set of risk management tools are easily transferable across those but it may still be of use to define a typology of the different types of risk. This provides for a clearer analytic framework and can be crucial to business and executive communication. Here, the classification of Crouhy et al. (2005) is followed. We can distinguish between eight major types of risk in business:

- **Market risk** - this is the risk that financial market volatility will have a negative effect on the value of an asset. Depending on the source of risk we can define equity price risk, interest rate risk, Forex risk, and commodity price risk.
- **Credit risk** - the risk associated with the change in the credit rating of a counterparty which can change the value of own assets.
- **Liquidity risk** - this category essentially comprises two large sets of risks. The first one refers to a company being illiquid - the risk of not having enough cash flow to cover its operations. The second one refers to the market being illiquid and thus the company is unable to sell a desired asset (at the desired price, or at all).
- **Operational risk** - the risk that stems from poor organization of labor, faulty controls, inadequate systems, management failure, and human error. An issue of particular interest in the group of operational risk is also fraud.

- **Legal and regulatory risk** - this stems from the changes in law and the legal environment that can affect the value of assets, or positions. Changes in tax laws directly affect cash flow, and more indirectly - operations. So is the case with other laws and regulations, with some of them, like nationalizations, providing extreme examples.
- **Business risk** - these refer to the classic risk of economic and business transactions - the uncertainty of forecasting demand, or the precise pricing, or costs. Traditionally, they have received large attention in the literature but the palette or risks that a modern manager should consider is much broader.
- **Strategic risk** - this refers to the risk, associated with making a significant investment or undertaking decisive action on the strategic level that can be associated with a potentially very large profit or loss. In essence, this is the risk stemming from strategic planning at the company level (new products, new markets, new business models, etc.)
- **Reputation risk** - this is the potential damage that can be done to firms when their activity is seen as unethical, unacceptable, or fraudulent. Since a large part of business operations is built on trust and history, potential reputational damages can be very significant. Likewise, maintaining good reputation can be monetized.

The list here provides a broad overview of the spectrum of risks that a modern organization should be cognizant of. Depending on the business model, history, and mode or operations, some of the risks will be more prominent, while others - unimportant. However, the risk manager is well advised keep at least topline track of most of them as changing economic circumstances may dramatically change their importance for the company.

The Risk Management Process

Once a risk (or a risk group) is identified, the next stage would be to enter a risk management process in order to ensure that upsides can be captured by the organization, and downsides can be avoided. The process of risk management is therefore an exercise into maximizing gains and minimizing losses. From a behavioral standpoint, however, the human brain is rigged in such a way to feel losses more painfully than rejoice at similar gains (asymmetric utility curve). This has also permeated the cultures of many organizations and therefore they tend to view risk management as an exercise predominantly in minimizing downside risks.

A primer of the process of risk management can be given in the following steps: identification, measurement, finding tools for mitigation, devising strategy, and evaluation and learning.

Risk Identification refers to the important phase of defining pertinent risks for the organization, again given the objective environmental constraints. Here, we need to particularly focus on the business model and see what is most crucial.

Risk measument refers to the formal process of estimating the impact of the defined risks on current and future operations and cash flow. This can be done either formally using some model (like the CAPM and VaR that we look into) or informally using experience and expert

evaluation (like the Risk Matrix), or a combination of the two. Even if no data is available for statistical inference, one should not shy away from using expert knowledge to formalize risk - even a meticulous management of opinion can be beneficial to organizations. This is so since oftentimes managers and employees tend to be oblivious of downsides and reminders help. Apart from measurement or risks, one should define their probabilities of happening and thus their impact on the organization. Thus large risks with large probabilities will tend to be top priority, while small risks with small probabilities - not.

Finding mitigation tools is connected to defining a set of strategies or actions that the company can use in case the identified risk occurs. Those mitigation tools also need to be evaluated in terms of effectiveness and impact and the best ones - selected, retained, and used if needed.

Devising Risk Mitigation Strategy refers to the conscious decision on what risks to undertake, to manage and what mitigation tools to use. It may be perfectly sensible for a company to choose to ignore a risk or give it low priority. In a world of conflicting priorities and scarce resources, small-impact events should hardly be considered. There are four distinct strategies for managing risk:

- *Avoid* - include the attempts of organizations to avoid risk by changing their operations or their financial positions. This is deemed appropriate if the risk's expected impact is relatively large with respect to the size of organization and it is not worth taking.
- *Transfer* - if the risk has to be undertaken for regulatory, business, or strategic reasons, risk managers may attempt to transfer it to another party that can bear it better. For example a building or a transaction can be insured, and a financial market position can be hedged.
- *Mitigate* - the company may choose not to transfer the risk but to undertake action to mitigate it, and decrease its downside. For example, if an insurance company fears for a large proportion of fraudulent claims, it can improve its fraud detection systems.
- *Keep* - a last option is to just keep the risk and undertake no action. If the risk has low potential downside, it's unimportant, or there are no sensible options to avoid, transfer, or mitigate, the organization will have to live with it.

It is important to note that the key task of the risk management exercise is not to avoid risk - it is to **take the optimal amount of risk**. A basic observation is that higher risk is connected to a higher expected return, thus avoiding risk jeopardizes profit. The balance between the two is crucial in order to ensure organizational effectiveness, efficiency, and excellence.

The Risk Management Professional

There are now many places in which risk management professionals have found employment. Those range from commercial and investment banks, funds, large (and often medium) corporations, international organizations, and the public sector. Their role is key in

informing senior stakeholder and decision-makers on the risks the organization is potentially facing and **supporting a risk-adjusted decision**.

Challenges are numerous, and they largely reside in the culture and preparedness of organizations to take advantage of risk management analyses. Oftentimes, executives feel that caution is unnecessary (and often detrimental to business results and bonuses), while experts question the quality of models and conclusions. Paraphrasing Samuelson's metaphor, it is sometimes the risk manager that advises taking away the punch when the party is at its peak. This role is crucial to the success of modern organizations, and has large potential value to add but in some cases its value needs to be continuously proven to senior management.

Project 1

Using Google Trends, investigate the popularity of the different types of risk over the past five years. Correlate the popularity data with at least three suitable metrics of economic or social development (GDP growth, inflation, unemployment, debt, investment, etc). Comment on the results.

Lecture Two: Quality Qualitative Evaluations of Risk

Risk Management is an important activity for any organization as it enables it to minimize the downsides of its actions and optimize on the economic opportunities it has. Many organizations, however, do not have a designated Risk Management Department and this function is taken up by other professionals. There are also instances in which little, if any, data that can help on quantitative estimates of risk is collected. While those situations are most difficult to implement RM paradigms and initiatives, it is precisely here that most value can be unlocked.

We will take issue with the activity of managing risk using qualitative data and approaches and apply it to the practice of project management. Two concrete tools - the Risk Matrix, and the Risk log are presented to aid the practitioner. Those ideas are very pragmatically used in the Project Management Body of Knowledge and we follow the approach outlined there (PMI, 2013).

Practical Process of Managing Risk

Project managers have to actively manage the risks in their projects. It is naturally most preferable to focus on prevention (proactive) rather than dealing with the risk (reactive), whenever possible. The ability to forecast and plan for a risk helps not merely in prevention but also in active management and successful communication with stakeholders.

In projects there is a clear trade-off between likelihood and impact of risks as time progresses towards completion. As the project nears its end, the likelihood of the risk decreases but the cost of repair increases rapidly due to investment already made.

Risk management proceeds along the common generic model with six clearly identifiable steps:

- **Risk Planning** - the process begin with devising a plan on how to conduct risk management, which outlines general principles, responsibilities, timelines, resources. This stage, together with the next one should proceed in a very intensive communication with stakeholders.
- **Identify Risks** - a usual part of the process, which helps the project manager to outline risks. General tools include document review, expert meetings, simple modeling. A review of methods that can be useful follows in the next section.
- **Qualitative Analysis** - allow for estimating the impact of risks and their probability. As no hard data is available those are usually evaluated on an ordinal scale. For example, impact may be categorized as “low”, “medium”, “high” or “very high”, and the probability can be categorized as “improbable”, “not very likely”, “likely” or “very likely”.
- **Semi-quantitative Analysis** - essentially the third stage converts expert intuition into data that can be fed into the risk management process. For this purpose ordinal

evaluations can be converted into numeric ones. For example "improbable" can be recoded as p = 0.1, "not very likely" as p = 0.3, "likely" as 0.6 and "very likely" as p = 0.9. Alternative scales are possible depending on the needs of the project, but it is still advised that organizations are consistent in the usage. We should make an important note here - the numerical values attached to ordinal categories do not have a precise quantitative meaning, as they are approximations at best and should be interpreted with the necessary care. Still, this phase allows the analyst to calculate expected values and prioritize risks.

- **Risk Response Plan** - after the full picture is obtained, the expected monetary value of risks should be used to craft response strategies. In the case of negative risks they are avoidance, transfer or mitigation; for positive risks one can exploit, enhance, or share. At any rate, the organization may just decide to accept and keep the risk without further ado.
- **Monitor and Control** - this stage includes implementing risk response plans, tracking identified risks, monitoring residual risks, identifying new risks, and evaluating risk process effectiveness (risk audit). This stage presents the active practice of risk management throughout the project life cycle and should end with a set of lessons learnt that can improve RM practice over to the next project.

Risk Management with No Data

As we have seen, calculating probabilities and estimates of expectations is easier when data on the events is actually present. Under conditions of data availability the analyst can utilize empirical probabilities which should approximate the true ones as N grows large, and can further estimate expectations by calculating averages, weighted averages or by forecasting. However, if no data is present this can hardly be done, and the risk management professional needs to collect information, codify it, quantify it, and then use it as a basis for decision-making. This process is particularly challenging as it has to take recourse to a large set of potential stakeholders with their possibly diverging opinions and their vested political interests in the outcome of the RM process - e.g. it may well be the case that business development professionals are more optimistic about deadlines that developers, as they are driven by different incentives. Thus the RM professional should both serve as a leader and as an arbiter in the process of collecting data, and thus the responsibilities include:

- *Deciding on what data to collect* - definition of what data is needed for the RM process. Data should only be collected if its utility is larger than its cost. No superfluous data should be collected as this both increases costs and dilutes focus.
- *Identifying sources* - there are different ways to collect data, and often more than one source can provide it. Focusing on quality, reliability, and objectivity of the source is key.
- *Collecting objective data without priming the result* - as the RM analyst collects data (s)he might want to help other experts understanding the problem and situation at

hand but without priming them as to what answers they should give. This is particularly important in the case of close relationship with the respondents as they will try to be helpful and agreeable, to the detriment of the project.

- *Adjudicating between conflicting views* - if no hard quantitative data is available, there is plenty of room for disagreement. The RM analyst needs to adjudicate between views in case that no agreement is possible, and find ways to align more closely the estimates of different stakeholders.
- *Cleaning and processing data* - now that data is collected it needs to be processed and prepared into a machine-readable (most often quantitative) form that can be fed into the RM process.
- *Reaching consensus on data* - the RM analyst also needs to communicate data and try to reach broad consensus and agreement on the meaning of the numbers and their correctness. Oftentimes this will have to involve the project sponsor due to issues of seniority.

The methods for collecting data along those steps are presented in more detail in the next section.

Collecting Qualitative data

If no structured hard data is available for the purposes of RM, we can use a variety of sources to collect qualitative data and then process it. We enumerate the most important ones, as follows:

- **Documentation Reviews** – including project charter, contracts, and planning documentation, which can help identify project bottlenecks and potential risks.Those involved in risk identification might look at this documentation, as well as lessons learned, articles, and other documents, to collect further data.
- **Brainstorming** - a common approach whereby important stakeholders and experts gather to generate a list of ideas, as wide as possible, with no fear of criticism of any kind. Those ideas are then prioritized and processed to reach a list of risks.
- **Delphi technique** - a technique whereby experts participate anonymously, and a facilitator tries to glean information by using a questionnaire. The process goes on iteratively in a few rounds until consensus emerges. This method is useful as it helps in the reduction of bias and prevents some of the participants from influencing others (as can be the case during brainstorm sessions).
- **Interviews** - interviewing experts, stakeholders, experienced project managers (PMs) and RMs can yield invaluable insights into the RM process and help tap organizational knowledge for the needs of estimating risk. One should beware as this method is sensitive to difference in opinion and considerations of company politics, and may involve the RM adjudicating between views.

- **Root cause analysis** - this is method which helps identifying new risks by reorganizing the already identified risks by their root cause. Such grouping can yield more valuable information.
- **Checklist analysis** - the RM professional looks at checklists of previous similar projects containing accumulated historical information and investigates problems that arose in past experience.
- **Assumption analysis** - the analyst seeks to identify risk from inaccuracy, instability, inconsistency, or incompleteness of assumptions that can have negative repercussions on the project.
- **SWOT analysis** – this is a classic tool that delineates Strengths, Weaknesses, Opportunities, Threats (hence the name) and can help identify risks by looking at weaknesses, opportunities, and threats. Depending on the scope and goal of project, focus may be on the internal side (S and W), or the external side (O and T).
- **Influence diagrams** - show the causal influences among project variables, the timing or time ordering of events, and the relationships among other important features and their outcomes. This helps outline the ecosystem of risks and the interrelations of variables.
- **Cause and Effect Diagrams and Process Flows** - this process involves creating process maps and finding precise causal links between the variables. This is pretty close to influence diagrams, but with greater degree of precision, and confidence about the causal links. Again, this helps in the identification of risk.

These methods can be used to gather different information, but the RM professional should focus on three main things - identification of risk, measuring its potential impact, and estimating its probability. The table below summarizes which of these can be gleaned using the different methods we presented.

Method	Identify Risk	Impact	Probability
Documentation Reviews	Yes	No	No
Brainstorming	Yes	Yes	Yes
Delphi Technique	Yes	Yes	Yes
Interviews	Yes	Yes	Yes
Root Cause Analysis	Yes	No	No
Checklists Analysis	Yes	No	No
Assumptions Analysis	Yes	No	No
SWOT Analysis	Yes	No	Yes
Influence Diagrams	Yes	No	No
Cause and Effect Diagrams	Yes	No	No

Risk Matrix and Risk Log

Once data is collected and entered into processable form, the RM analyst can proceed into quantifying it and using it for formal modeling of project risks. This usually entails encoding verbal descriptions into numerical values. Those values are of course approximations but are still needed for the analysis. Here we present the PM Body of Knowledge (2014) methodology of the conversion,

Probabilistic statements can be converted into values by using the following scale, presented in the Probability Matrix. Essentially we obtain a discrete ordinal distribution by doing that instead of the true continuous distribution that is characteristic of the real world, but this approximation is still useful.

Probability Matrix

Scale	Rare	Unlikely	Possible	Likely	Almost certain
Probability	0.1	0.3	0.5	0.7	0.9

After probability is quantified, the next step is to quantity impact. For this to happen, we need to know what impact is characterized as high or low, and how it differs across key project dimensions. We present the impact matrix across four key dimensions - cost, time, scope, and quality. For example, a high cost impact entails increase of projected costs by 20-40%, while high time impact reflects a time increase by 10-20%. Similar difference can be seen also across other key variable. This discrepancy is due to the differential in relative importance of different dimensions and will also reflect the stakeholder's tolerance to problems in different areas. Still, the table presents some rules of thumb that can be useful to the practitioner.

Impact Matrix

Impact	Very Low / 0.05	Low / 0.1	Moderate / 0.2	High / 0.4	Very High / 0.8
Cost Increase	Insignificant	$< 10\%$	10-20%	20-40%	$> 40\%$
Time Increase	Insignificant	$< 5\%$	5-10%	10-20%	$> 20\%$
Scope Decrease	Barely noticeable	Minor	Major	Unacceptable	Useless
Quality Degradation	Barely noticeable	Minor	Major	Unacceptable	Useless

After quantifying probability and impact, we can calculate the expected impact. A useful tool for doing that is the Risk Matrix. Its columns represent different impacts, while its row correspond to probabilities. By placing risks in the appropriate cell, it allows for a quick typology of all the risk our project (or organization) faces and their relative importance.

Here we denote with "OK" risks with low expected impact that should not be a priority in the RM process, with "+" possibly important risks that the project (or risk) manager should monitor, and with "!" risks with high potential for large downside that should be managed

actively in order to ensure successful project completion. In that sense the Risk Matrix is a prioritization tool which allows for a rational focus on what risks are most pertinent to the project, and allows for their more effective control.

Risk Matrix

Impact	Insignificant	Minor	Moderate	Major	Catastrophic
Almost Certain	+	+	!	!	!
Likely	OK	+	+	+	!
Possible	OK	+	+	+	!
Unlikely	OK	OK	OK	+	+
Rare	OK	OK	OK	OK	+

Having an approximation to probabilities, one can venture to estimate the Expected monetary value (EMV) of those risks. This is done by finding the monetary impact of a given risk, M_i, and then multiplying it by its probability, thus reaching the following formula:

$$EMV_i = p_i * M_i$$

By summing all EMVs across risks, the RM practitioner receives an estimate of the total amount of money that the project or organization is likely to lose from risks on the downside and gain from the risks on the upside. Such a buffer should be budgeted and accounted for. Additionally, RM professionals can also venture to calculate EMVs across different groups of risks (e.g. time, or scope). For a given group of risks, j, the EMV is merely the sum of all expected monetary flows associated with risks from this group:

$$EMV_j = \sum_{i=1}^{n} p_i * M_i$$

Such calculations can be used to manage more effectively, but also to communicate across the organization and to focus attention on areas with highest EMVs.

Concluding Comments

Risk management in the event of limited or no data can be particularly challenging as information needs to be collected and presented in a quantitative form. Once values for (monetary) impact are calculated, and probabilities are pinned down in numbers, there is the large and sometimes irresistible temptation to leverage more sophisticated quantitative algorithms.

While it is in principle possible to do modeling, to undertake sensitivity analyses or even to utilize elaborate methods such as Monte Carlo simulations, we should only do that with extreme caution. Neither impact estimates nor probability estimates are true metrics

generated by an underlying process - they are merely approximations evinced by possibly subjective experts. Even if the numbers look close to what the data-generating process would bring about, they are only values on an ordinal scale. At best they preserve the correct ordering, and relative distance; at worse those values are meaningless and misleading.

The RM practitioner is thus well advised to focus on collecting high-quality data instead of relying too heavily on sophisticated statistical model that utilize flimsy numbers.

Project 2

Think about a project that your company or an organization you are aware of is about to commence. Create its complete Risk Matrix by enumerating risks, measuring their impact and probability, and devising mitigation strategies.

Lecture Three: The R Language as a Tool for Risk Management

This section of the lecture notes aims to provide a quick overview of basic functionality of the R software for statistical computing (R Core Team, 2014). It is intended to help the analyst understand the basic structure and logic of the language and prepare him or her for productive work with it.

The topics included range from an introduction into R through reading data and manipulating data, visualizing it, and conducting statistical analyses. The text is structured in a hands-on way and can be both read in its entirety or referred to on the go.

Introduction

The R language is a popular language for statistical programming, hailing from work on the S language in the late twentieth century. Since then, R is supported by a large community and continuously developed by a core team of programmers and statisticians. It is popular among academics and researchers and is recently gaining popularity in data-intensive businesses as well.

It is particularly notable for the following:

- **Scalable** - it scales well and can do analysis on large data sets - something that traditional tools either cannot, or were slow to adapt
- **Up-to-date** - it includes the most novel statistical methods and approaches, that are often not contained in proprietary software packages
- **Customizable** - it gives the user complete control over the processes
- **Diverse** - it provides numerous alternative ways to do particular analysis and visualizations, letting users choose their preferred one
- **Popular** - it is widely used and supported by a large and enthusiastic community that can be a source of help and inspiration
- **Open-source and freeware** - it is open-source which allows constant development by non-members of the core team and is also free of charge, making it accessible to large audiences.

Of course, this has to be taken against the steep learning curve and the sometimes demanding system requirements (especially for RAM). Overall, the R language is a versatile tool that can provide competitive advantage to those who master it and use it.

Reading-In Data

The first part of an analysis is to actually read data in, that can be worked with. A standard way is to have a source of data in some form and load it. If it is a table, one can use the `read.table` command. The program will try to find the specified name in the command in the home directory and load it. If it is not in the home directory, the user will have to specify the full path.

```
read.table(file = "data.csv")
```

This command gives a lot of flexibility as it has further options that can specify all details needed - the header, the separator, column names, special symbols, etc. The best way to understand a command is to write it over on the command line with a `?` in front of it. This will open the help file, giving more details about it, and concrete examples of usage and syntax.

```
?read.table
```

A simpler version of this is a command to read .csv files, which has less options and is easier to use.

```
read.csv("data.csv")
```

If you only write this, the data will be read but not stored in the R environment. You have to assign it to an object. The way to do this is to use the assignment operator `<-`.

```
data <- read.csv("data.csv")
```

In this way we create an object named `data` and assign it to contain the data that was read into the file.

Another way to load data is to just use one of the data sets that comes in R itself and perform analysis and visualization on it. To see what data is available type:

```
data()
```

For example we can load the data on closing prices of major European Stock market indices over the period 1991-1998. This is done with the following command:

```
data(EuStockMarkets)
```

The data appears in the global environment and is ready to work with.

Summarizing Data

A common first step is to look at a few numerical and visual summaries of data in order to better understand its form, structure, and some information. To look at the first lines of the data set you can do the following with the `head` command.

```
head(EuStockMarkets)
```

```
##          DAX    SMI    CAC   FTSE
## [1,] 1628.75 1678.1 1772.8 2443.6
## [2,] 1613.63 1688.5 1750.5 2460.2
## [3,] 1606.51 1678.6 1718.0 2448.2
## [4,] 1621.04 1684.1 1708.1 2470.4
## [5,] 1618.16 1686.6 1723.1 2484.7
## [6,] 1610.61 1671.6 1714.3 2466.8
```

And to look at the last lines, you can use the `tail` command:

```
tail(EuStockMarkets)
```

```
##             DAX    SMI    CAC   FTSE
## [1855,] 5598.32 7952.9 4041.9 5680.4
## [1856,] 5460.43 7721.3 3939.5 5587.6
## [1857,] 5285.78 7447.9 3846.0 5432.8
## [1858,] 5386.94 7607.5 3945.7 5462.2
## [1859,] 5355.03 7552.6 3951.7 5399.5
## [1860,] 5473.72 7676.3 3995.0 5455.0
```

General descriptive statistics can be obtained via the `summary` command :

```
summary(EuStockMarkets)
```

```
##       DAX             SMI             CAC             FTSE
##  Min.   :1402   Min.   :1587   Min.   :1611   Min.   :2281
##  1st Qu.:1744   1st Qu.:2166   1st Qu.:1875   1st Qu.:2843
##  Median :2141   Median :2796   Median :1992   Median :3247
##  Mean   :2531   Mean   :3376   Mean   :2228   Mean   :3566
##  3rd Qu.:2722   3rd Qu.:3812   3rd Qu.:2274   3rd Qu.:3994
##  Max.   :6186   Max.   :8412   Max.   :4388   Max.   :6179
```

Data Selection

A common task is to select only a subset of data to manipulate it. For example, we may be interested in the mean or the standard deviation of only one of the four indices. A common way to select (or subset) data is to use `[x, y]` after the name of the data. Here `x` correspond to the column number, and `y` to the row number. If we want to select all columns or rows, we just put `,` instead of a number.

For example, to select the first column of the four market indices, we can type `EuStockMarkets[,1]`. If we want to select the first row of data (observation), we can type:

```
EuStockMarkets[1,]
```

```
##      DAX      SMI      CAC     FTSE
## 1628.75 1678.10 1772.80 2443.60
```

Finally if we want to select the first observation of the first column, then it is:

```
EuStockMarkets[1,1]
```

```
##      DAX
## 1628.75
```

If we want to select just a column (variable), like the first index, the can use the "$" sign and the name of this variable. Note that this is not applicable for objects formatted as time series. Once a subset of data is selected, it can be assigned to an object in the R environment. Here we assign the index FTSE to the `FTSE` object:

```
FTSE <- EuStockMarkets[,4]
```

We can either use the object and apply functions to it, or use the selection straight away. Here we calculate the mean (`mean`) and the standard deviation (`sd`) of the index:

```
mean(FTSE)
```

```
## [1] 3565.643
```

```
sd(FTSE)
```

```
## [1] 976.7155
```

Instead of the object we can also use the selection:

```
mean(EuStockMarkets[,4])
```

```
## [1] 3565.643
```

```
sd(EuStockMarkets[,4])
```

```
## [1] 976.7155
```

Data Classes

R supports many data classes that describe different types of data and that require different analytic methods and have different visualization needs. Fortunately, a lot of functions in R are generic, i.e. they check the type of data and find the most appropriate method for this data. Sometimes, however, it is useful to know the class of the data object one works with, and sometimes it is imperative to be able to change it.

A few common types of data are:

- **Numeric** - quantitative values that can be processed through different analytics and plotted. Example: asset prices.

- **Character** - a string of letter, usually a name, or some textual information. Example: company names.
- **Factor** - an ordinal or nominal value that distinguishes between data categories. Example: company sector.
- **Time Series** - numeric data that has a temporal dimension to it. Example: asset returns on given days.

To understand what is the class of a given object or subset, we can use the `class` command, and to get an idea of how the object is structured, we can use the `str` command“”

```
class(EuStockMarkets[,1])
```

```
## [1] "ts"
```

```
str(EuStockMarkets[,1])
```

```
##  Time-Series [1:1860] from 1991 to 1999: 1629 1614 1607 1621 1618 ...
```

Sometimes specific analytic or plotting methods require a different class of data that the analyst may have. In this case data has to be coerced into the needed class. This is done by using the `as.XX` command where XX refers to the new class of data. For example if we want to coerce the time series object EuStockMarkets[,1] into numeric class and assign it to object `DAX` we do the following:

```
DAX <- as.numeric(EuStockMarkets[,1])
class(DAX)
```

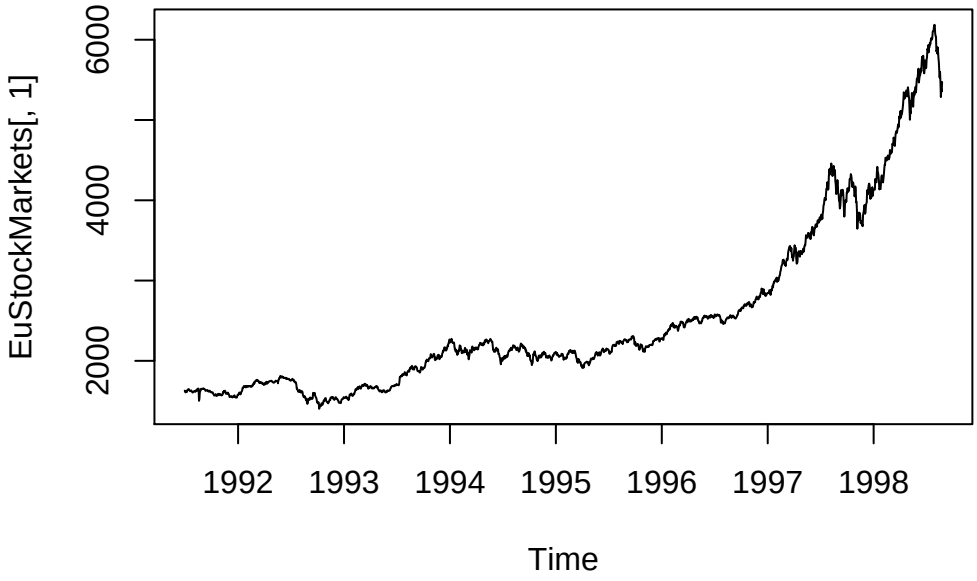
```
## [1] "numeric"
```

One should be careful what data class is required by a certain method and be careful that the data class at hand and the required one are consistent. Lack of this can lead to either errors or incorrect results.

Plotting Data

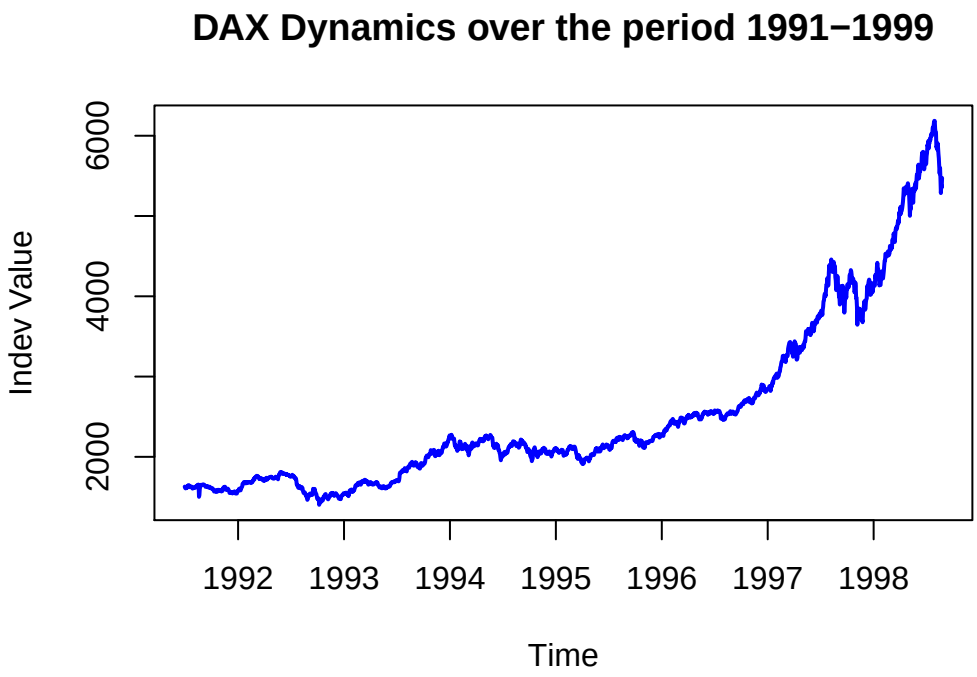
A key feature of any statistical language is its plotting facilities. Visualization of data helps better understanding, enable the discovery of patterns and trends in data, and finally makes for much better and clearer communication of results. The R language has many alternative plotting facilities. The most basic one is the `plot` command. It is a generic command which will produce a different default type of plot depending on the data at hand. We can use it to plot the first index of the `EuStockMarkets` data:

```
plot(EuStockMarkets[,1])
```



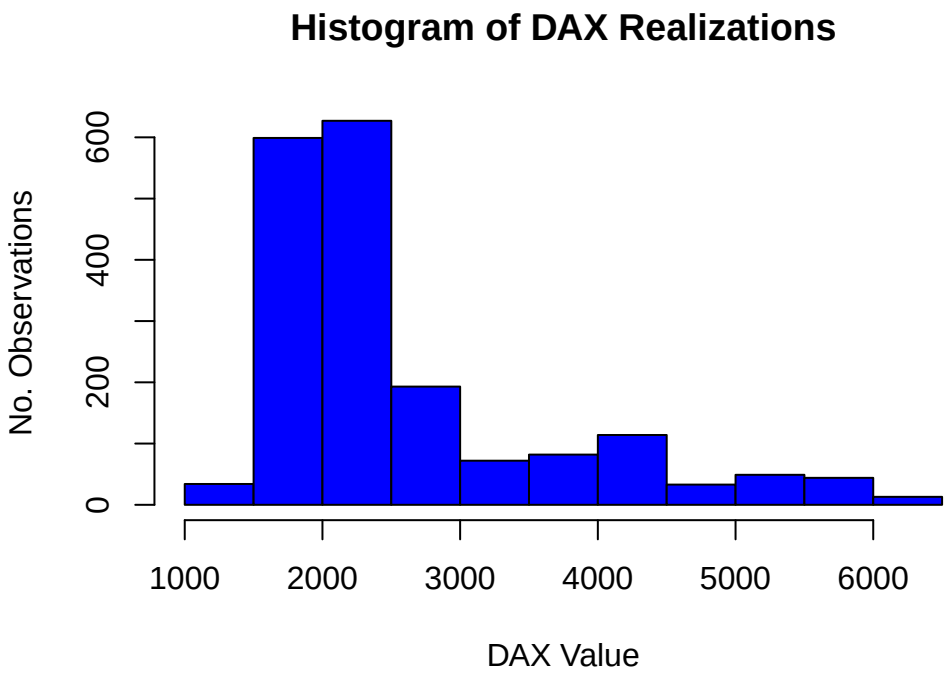
We can customize color with the option `col`, change the line width with the option `lwd` and the two axis names with the options `xlab` and `ylab`. The main title is set with the option `main`, as follows:

```
plot(EuStockMarkets[,1], col="blue", lwd=2, ylab="Indev Value",
     main = "DAX Dynamics over the period 1991-1999")
```



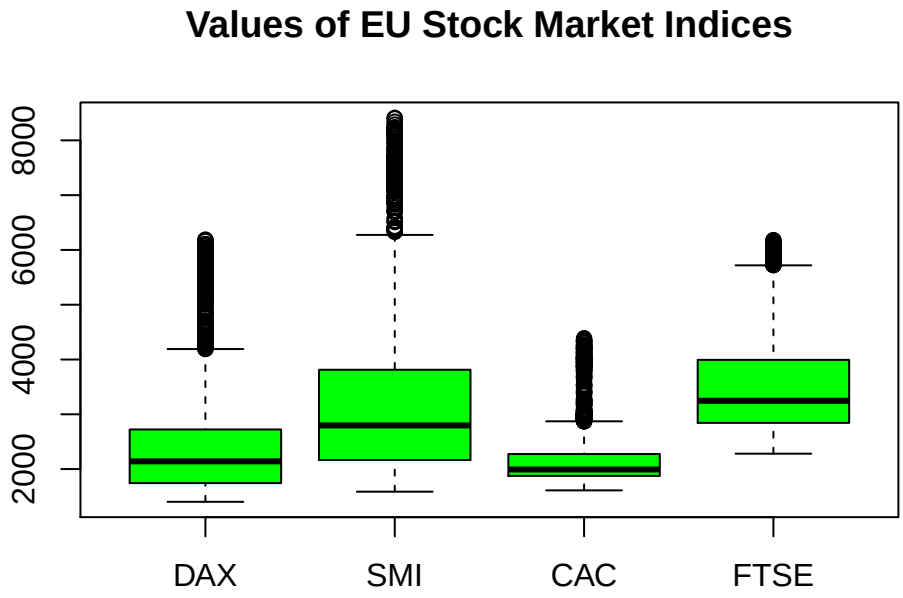
There are many other options that can be explored with the `?plot` command. Another important graph would be the histogram. We make a histogram (`hist`) of the DAX values as follows:

```
hist(EuStockMarkets[,1], col="blue", ylab="No. Observations",
     xlab = "DAX Value", main = "Histogram of DAX Realizations")
```



A useful type of graphic is the boxplot. We can see the comparative values of different groups of observations using it. Here we use the `boxplot` command to compare the four indices:

```
boxplot(EuStockMarkets, col="green",
        main="Values of EU Stock Market Indices")
```

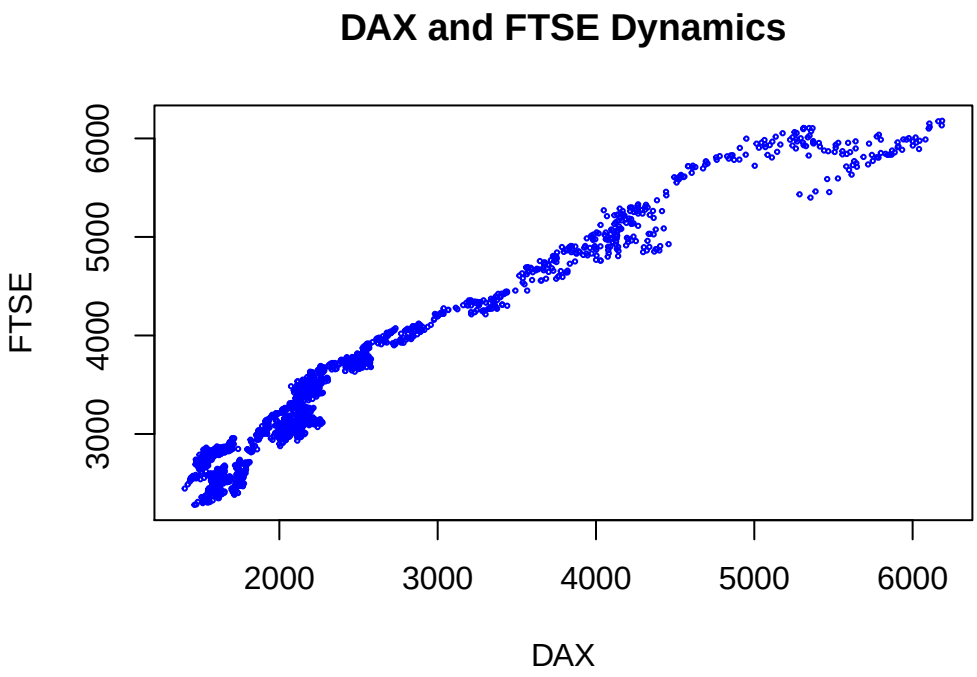


Finally, we can see how related are two pairs of indices using a scatter plot. For this purpose we convert time series data into numeric data and assign them to two object - `DAX` and `FTSE`:

```
DAX <- as.numeric(EuStockMarkets[,1])
FTSE <- as.numeric(EuStockMarkets[,4])
```

Now we use the scatter plot. Given this types of data it will be automatically generated by the `plot` command, but we can also explicitly set using the option `type` in the command:

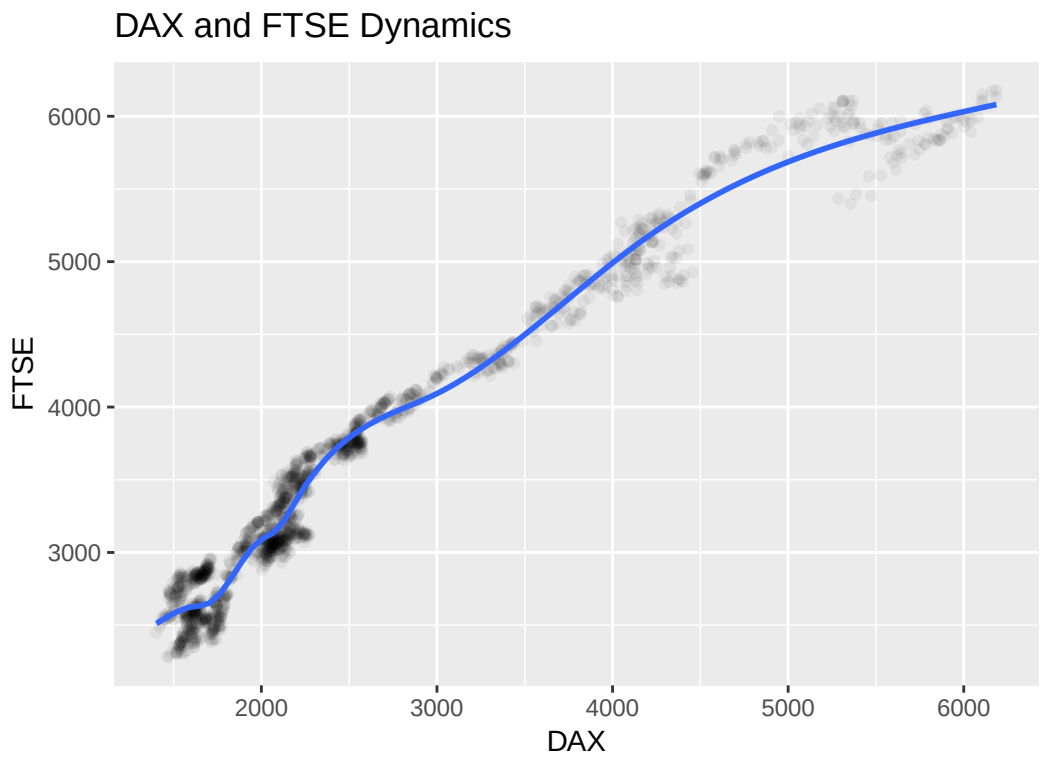
```
plot(DAX, FTSE, col="blue", main="DAX and FTSE Dynamics", cex=0.3)
```



Now that we have plotted them against each other we can see that they are move very closely together, thus implying positive and relatively strong correlation between the two. All the plots generated by the commands in R can be saved to disk using the `dev.copy()` command, followed by the `dev.off()` one.

The are numerous alternative graphic systems in R which also produce more visually appealing graphs. The two leading contenders are Lattice graphs and the packages `ggplot2`. They are also a bit harder to learn and tend to be somewhat less generic that their more basic counterparts in the R language. However, a smooth transition to `ggplot2` may be facilitated by the command `qplot` which is easy to use and customize. We demonstrate the same scatter plot with it, and easily add a line of best bet (with `smoother` option):

```
library(ggplot2)
qplot(DAX, FTSE, main="DAX and FTSE Dynamics", geom=c("point","smooth"),
      alpha=I(0.05))
```



Basic Analytics

The R language supports a wide range of statistical analyses that can provide insight into data and help in the practical risk management process. Apart from the rich functionality that comes with the base packages there are constantly new tools and techniques that are being developed by the community.

Here we provide an overview of two basic but very common statistical operations. Once the logic is apparent, it can easily be transferred to other operations. It is often of interest to quantitatively measure the correlation between variables in order to see the potential for risk hedging. The basic command `cor` can help:

```
cor(DAX, FTSE)
```

```
## [1] 0.9751778
```

The high correlation between DAX and FTSE formalizes their close connection that was already apparent in plotting. Linear regression is commonly used to observe the effect of one independent variable (or many variables) over a dependent one. The command used for it is 'lm' and R automatically prints output:

```
lm(DAX~FTSE)
```

```
##
## Call:
```

```
## lm(formula = DAX ~ FTSE)
##
## Coefficients:
## (Intercept)         FTSE
##   -1331.237        1.083
```

Alternatively, the analyst may choose to store the linear regression into and object and then preview a summary of that. Apart from the convenience of storing results, the `summary` command on the `lm` object will provide more information about the regression itself.

```
lm <- lm(DAX~FTSE)
summary(lm)
```

```
##
## Call:
## lm(formula = DAX ~ FTSE)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -408.43 -172.53  -45.71  137.68  989.96
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.331e+03  2.109e+01  -63.12   <2e-16 ***
## FTSE         1.083e+00  5.705e-03  189.84   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 240.3 on 1858 degrees of freedom
## Multiple R-squared:  0.951,  Adjusted R-squared:  0.9509
## F-statistic: 3.604e+04 on 1 and 1858 DF,  p-value: < 2.2e-16
```

It may be useful to extract the regression output for some purposes (e.g. if the coefficients are betas of interest). This is done by selecting the `coef` part of the regression summary and writing it to a disk (e.g. by using the `write.csv` command). This will give you most of the lm output (including coefficients, constant, errors, t-statistics, and p-values) in simple tabular from for further reuse or reporting.

```
write.csv(summary(lm)$coef, "lmresults.csv")
```

A lot of analytic commands follow this structure and logic but the user is recommended to also have a look at their help pages using the `?` command in order to gain greater insight and ease of use.

Advanced Analytics and Expansion

Apart from the basic commands, the R statistical languages is characterized by a dazzling variety of options, tools, and methods for covering the complete spectrum of analytic needs - from the traditional classification and regression problems to the more novel ones such as social network analysis.

Novel methods are usually found in packages - add-ons which are installed and loaded into the basic R version. A lot of packages are hosted on CRAN where they are listed only after passing a review for safety and operability. Every such package has a vignette which describes the new commands it adds to R. Many of them come with specific examples.

Packages are installed via the `install.packages()` command. One of the packages that is very well-suited to financial analysis and risk management is the *Performance Analytics* package. We install it, and then load it into R.

```
install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)
```

After this operation is done, it can be used. More information can be obtained on the package's help section.

```
?PerformanceAnalytics
```

The versatility and adaptability of R, including through its ability to use community-driven advanced analytics packages makes it an indispensable tool for the modern RM and quantitative professional. Despite its relatively steep learning curve, the R language has many benefits and unique advantages.

Project 3

Run a multiple regression on a set of variables of your choice and store it in an object. Investigate the object and extract the residuals.

Do diagnostics on the residuals. Are they randomly distributed around zero or are there any patterns? Investigate the possible diagnostic tests and conduct at least four of your choice. Make graphs and explain them.

Lecture Four: Expectations and Deviations

We have defined risk as the deviation from what we expect. Mathematically speaking risk is representative of how far the realization of a given variables of interest, x, happens to be from its mathematical expectation *E[x]*. The notion of expected value has both its theoretical interpretation as it can be calculated from the variable's distribution, as well as its practical interpretation as we can calculate it from empirical data. This lecture focuses on the latter.

The volatility of the variable of interest (the risk) can then be quantified by resorting to its variance or standard deviation - both of which can be calculated either from the assumed data distribution or from data at hand. We survey the approaches and apply them to a data set with 24 stocks from the NY stock exchange, together with the SP500 index. Then we proceed to formulate the relationship between risk and rewards and thus define the Sharpe ratio and market betas.

Expected Value

The expected value can be aptly described with a gamble. If we throw a fair coin and assume we get 1 EUR for tails, and 0 for heads, our expectation is equal to the outcomes, weighted by their probabilities, or

$$E[x] = 0.5 * 1 + 0.5 * 0 = 0.5$$

This simple example can be extended to an arbitrary number of discrete outcome i together with their respective probabilities, thus reaching the following formula:

$$E[x] = p_i x_i$$

In the case that we are not dealing with discrete outcomes and probabilities we have to take recourse to the whole distribution function of the variable, and calculate the expectation from there. In practice there are rarely cases with well-defined probabilities, and the practitioner often has only data on past realizations. In this case, the expectation can be calculated empirically by making certain assumptions on data. The simplest one is that the long-term expectation is equal to the long-run average values of the data series. While this may seem naive, we should note that a lot of financial series are characterized by mean-reversion, thus they tend to converge to their long run averages.

We are going to see this in the behavior of the SP500 index over the period 1999-2004, with the time series being at monthly frequencies. We initially load the data and rename the index.

```
library(stockPortfolio)
data(stock04)
colnames(stock04)[[25]] <- "SP500"
stock04 <- as.data.frame(stock04)
```

Then we plot index behavior.

```
plot(stock04$SP500, type="lines", col="navyblue", lwd=3,
     main="SP500 Returns", ylab= "Return", xlab ="Time Index")
```

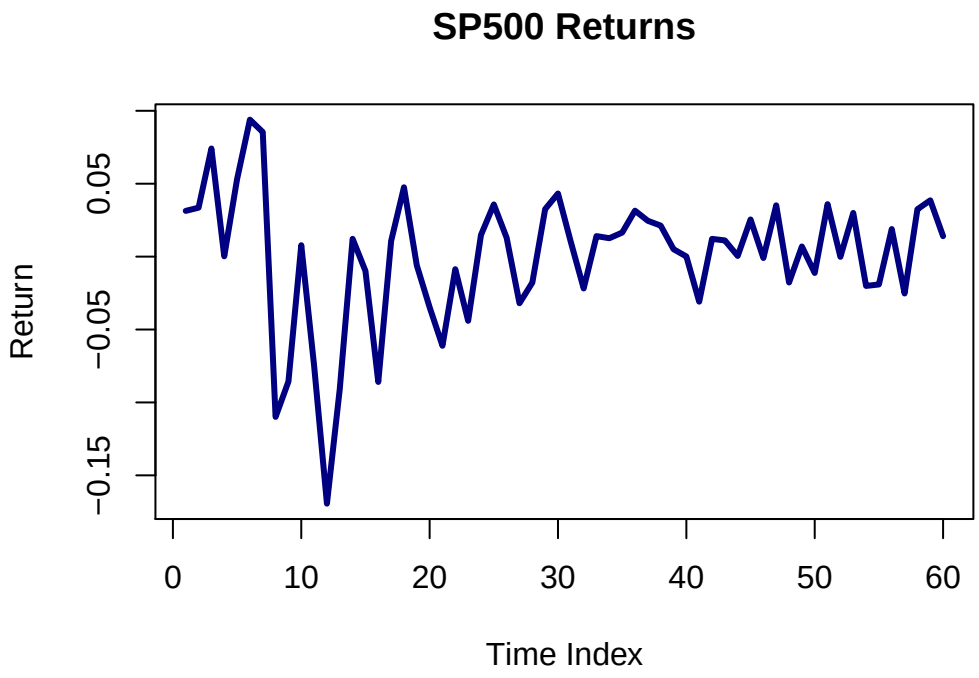
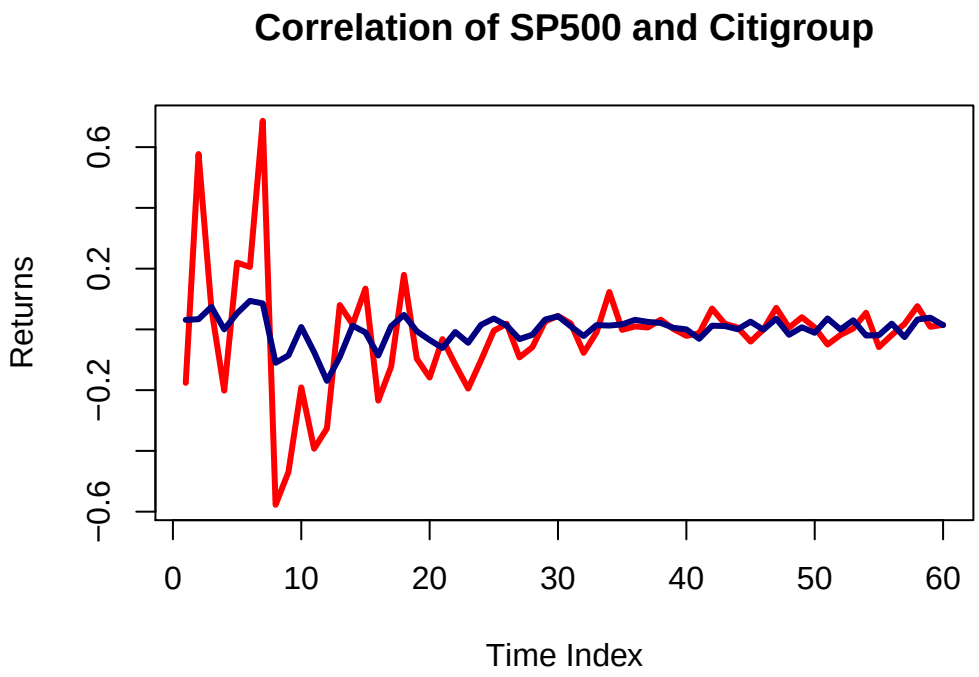
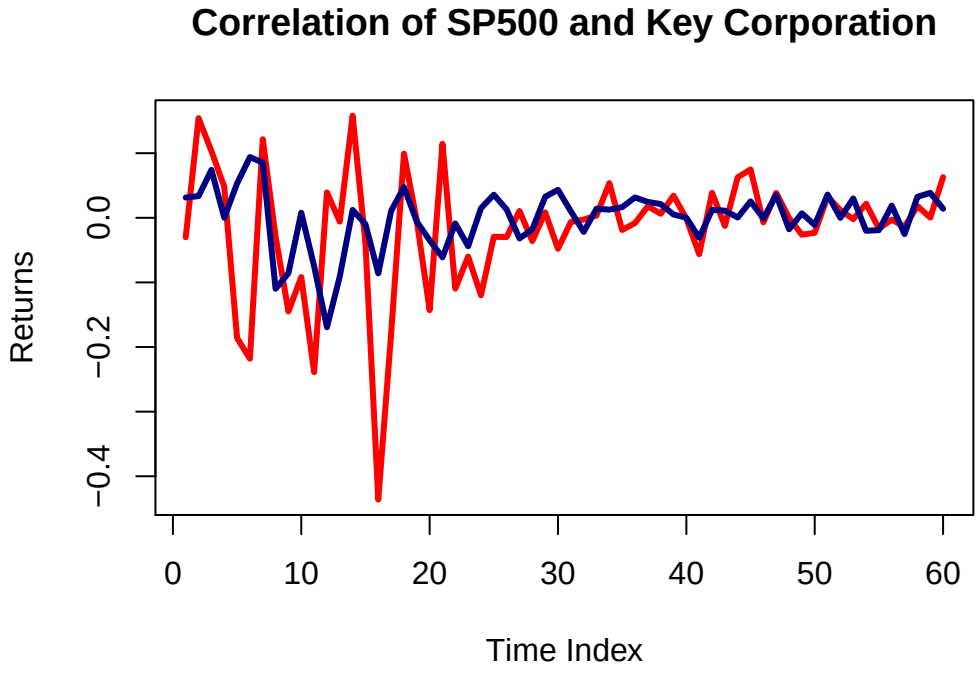


Figure: Dynamics of S&P500 Index over the period 1999-2004

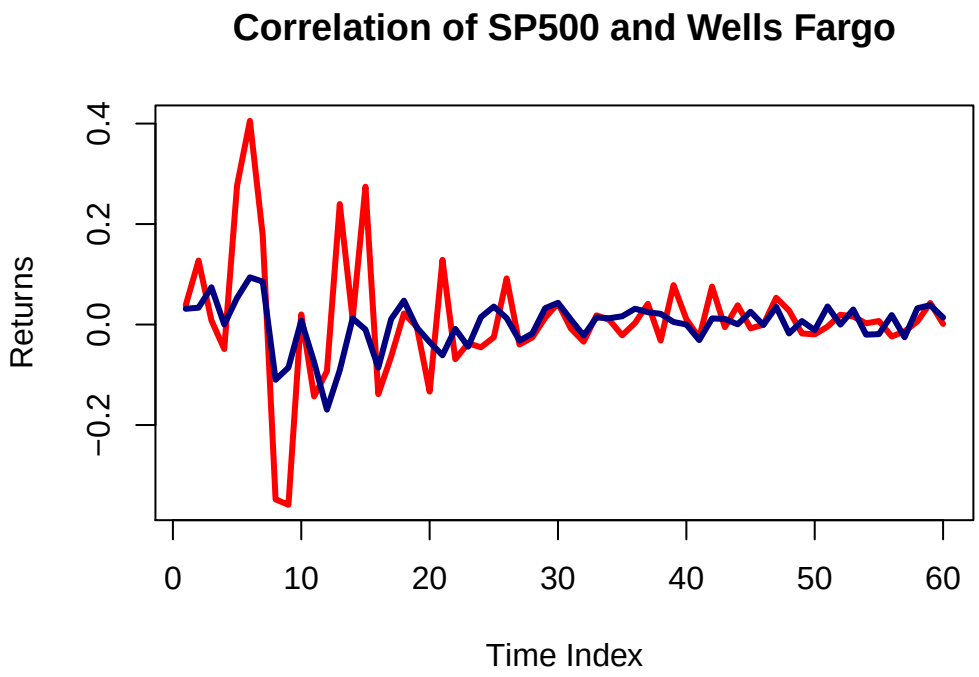
Initially we observe large deviations but they smoothen over the entire period and return to its long run average. Clearly, the stock market dynamics also reflect economic growth. When the series is detrended we would expect to see even stronger mean-reversion. In this sense also the Efficient Market Hypothesis would claim that short term fluctuations are largely unpredictable and can be best modeled either by random fluctuations around the long-run average or by a simple auto-regressive process of order 1 - AR(1). We can also see that individual stocks tend to follow overall market dynamics, albeit imperfectly.



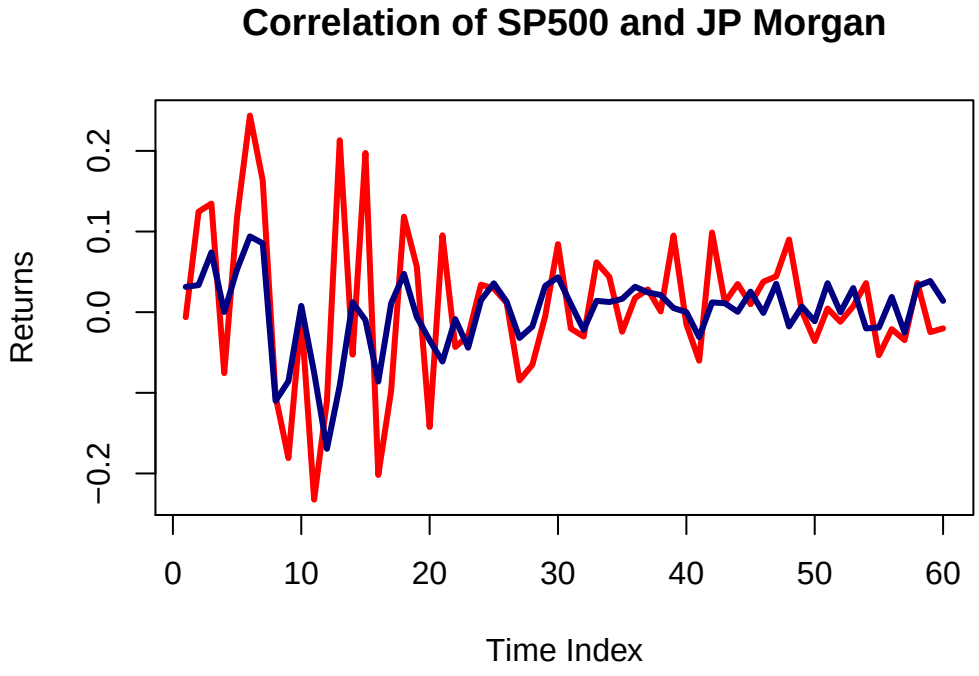

```
## integer(0)
```




```
## integer(0)
```




```
## integer(0)
```




```
## integer(0)
```

To get a better feel of stocks dynamics in the market we can also have a look at their overall descriptive statistics. For this we can use the command *describe* from the psych package or the more conventional *summary* from the base package.

```
library(psych)
describe(stock04, skew=F)
```

```
##       vars  n  mean   sd   min  max range   se
## C        1 60 -0.02 0.19 -0.58 0.69  1.26 0.02
## KEY      2 60 -0.02 0.10 -0.44 0.16  0.59 0.01
## WFC      3 60  0.01 0.12 -0.36 0.41  0.76 0.02
## JPM      4 60  0.01 0.09 -0.23 0.24  0.48 0.01
## SO       5 60  0.01 0.04 -0.09 0.10  0.19 0.00
## DUK      6 60  0.01 0.04 -0.10 0.09  0.19 0.01
## D        7 60  0.00 0.05 -0.15 0.13  0.28 0.01
## HE       8 60  0.00 0.07 -0.35 0.13  0.48 0.01
## EIX      9 60  0.01 0.06 -0.16 0.15  0.31 0.01
## LUV     10 60  0.00 0.09 -0.27 0.20  0.46 0.01
## CAL     11 60  0.03 0.17 -0.30 0.37  0.67 0.02
## AMR     12 60  0.02 0.21 -0.44 0.76  1.21 0.03
## AMGN    13 60  0.00 0.09 -0.16 0.33  0.49 0.01
## GILD    14 60  0.02 0.06 -0.13 0.16  0.29 0.01
## CELG    15 60  0.03 0.10 -0.25 0.24  0.49 0.01
## GENZ    16 60  0.00 0.07 -0.12 0.24  0.36 0.01
## BIIB    17 60  0.00 0.11 -0.41 0.25  0.65 0.01
## CAT     18 60  0.01 0.11 -0.35 0.35  0.70 0.01
## DE      19 60  0.01 0.10 -0.30 0.26  0.55 0.01
## HIT     20 60  0.00 0.09 -0.32 0.27  0.59 0.01
## IMO     21 60  0.02 0.10 -0.24 0.22  0.46 0.01
## MRO     22 60  0.01 0.10 -0.27 0.26  0.53 0.01
## HES     23 60  0.02 0.11 -0.27 0.42  0.68 0.01
## YPF     24 60  0.01 0.11 -0.33 0.32  0.65 0.01
## SP500   25 60  0.00 0.05 -0.17 0.09  0.26 0.01
```

The market as a whole over this period has a subdued return - its mean stands at 0% due to the large negative effects of the dotcom bubble in 2000-2001 which led the US economy into recession. This is why even large company stock have mean values of no more than 3%, while some venture into negative territory. Looking only at such a short period of time may be misleading and we might find it useful to look at longer time frames.

Deviations and Risk

We remarked that risk in financial markets is de facto described by fluctuations in prices (and respective returns). A measure of the average fluctuation around expected value is the variance (in absolute terms) and the associated standard deviation. The variance is defined

as follows:

$$VAR = \frac{1}{n}\sum_{i=1}^{n} E[(x_i - E[x])^2]$$

The standard deviation is the square root of the variance. While relatively unimportant from the standpoint of theory, the distinction between the two has some practical significance. The standard deviation is preferred as it retains the unit of measurement, and presents risk on a more intuitive scale. The standard deviations is defined as follows:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} E[(x_i - E[x])^2]}$$

The standard deviations can be easily obtained via means of the built-in functions in many econometric, statistical, and spreadsheet software utilities. We can also use the *sd* function in R to obtain standard deviations for all stocks in this data set.

```
sapply(stock04, sd)
```

```
##           C         KEY         WFC         JPM          SO         DUK
## 0.18853837 0.09758986 0.11690059 0.09265763 0.03830653 0.04184385
##           D          HE         EIX         LUV         CAL         AMR
## 0.05243038 0.07033330 0.05908553 0.08584715 0.17484460 0.20903094
##        AMGN        GILD        CELG        GENZ        BIIB         CAT
## 0.08809148 0.06302460 0.10197560 0.06884011 0.10545118 0.11084088
##          DE         HIT         IMO         MRO         HES         YPF
## 0.10072146 0.08828054 0.09646105 0.10217532 0.11432990 0.10540908
##       SP500
## 0.04604754
```

We can readily observe that risks across stocks vary significantly. The highest one is for Citigroup - standing at 18.9%, and the lowest for Southern Company - 3.83%, Duke Energy (DUK) - 4.18%, and the overall market itself (SP500) - at 4.61%.

If data would follow a normal distribution we can infer what ranges of values will a given variable take with a pre-determined degree of certainty. Let us take data on SP500 that we look and assume it follows a normal distribution with a mean of 0, and a standard deviation of 5. We randomly draw 1000 variables from this distribution N~(0,5). We expect that 68.2% of all cases would fall within plus or minus one standard deviation, thus 682 should take values between -5 and +5. 95.6% of all cases should be within +/-2 standard deviations (or between -10 and +10), and more than 99% of all cases will be between plus and minus three standard deviations.

```
x <- rnorm(1000,0,5)
hist(x, col="navyblue", main="Histogram of N~(0,5)",
     xlab="Value of realization")
```

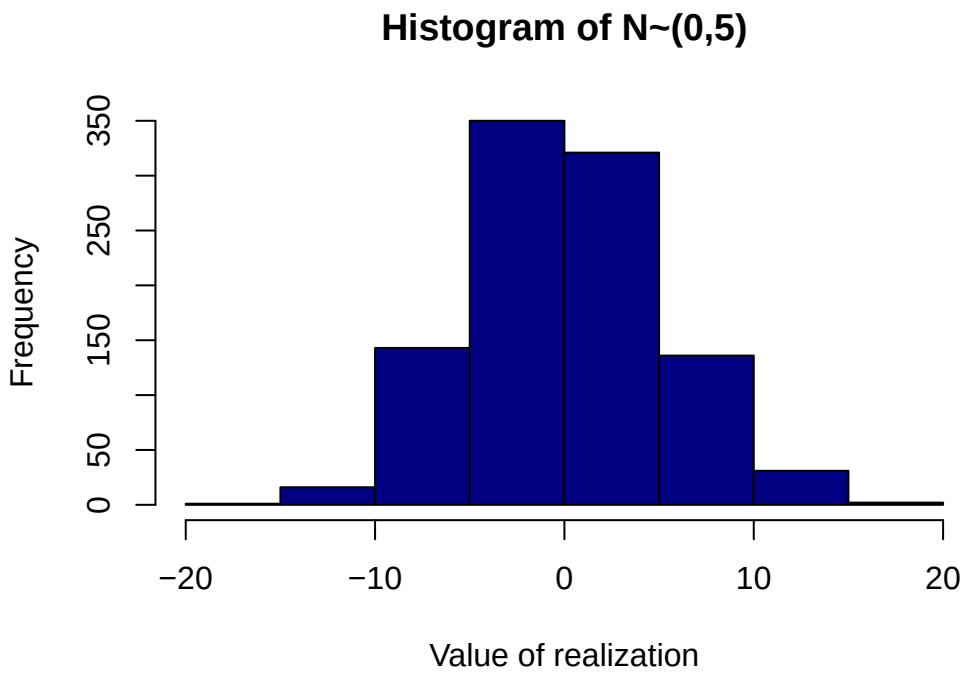


Figure: Values of randomly drawn numbers from normal distribution with S&P500 Parameters

Risk and Expected Return

When thinking about risk and expected return we should note both the theoretical rule, as well as the empirical regularity that they tend to be inversely proportional - the higher the expected return, the greater the risk is likely to be. A way to explain this is to take recourse to a market-making argument. If investors are to take up more risk they will naturally need have a bigger compensation for that. If a given asset will offer only more risk but with low relative expected return, it will never be traded and disappear from the market.

We can review the stocks expected returns and their risks to check this regularity. While the theoretical distinction between risk and expected return is of sweeping importance, in practice the analyst observes only current and past returns. If we assume mean-reversing price behavior, then past values can serve as a guide for future ones. Thus for our purposes we use the long-run average of returns as a proxy for expected future returns. To observe the correlation between risk and expected returns we calculate their standard deviations and means, and plot them in a scatter plot.

```
plot(sapply(stock04, sd), sapply(stock04, mean), col="navyblue",
     xlab="Stock Risk, SD", ylab="Expected Stock Return", pch=19)
```

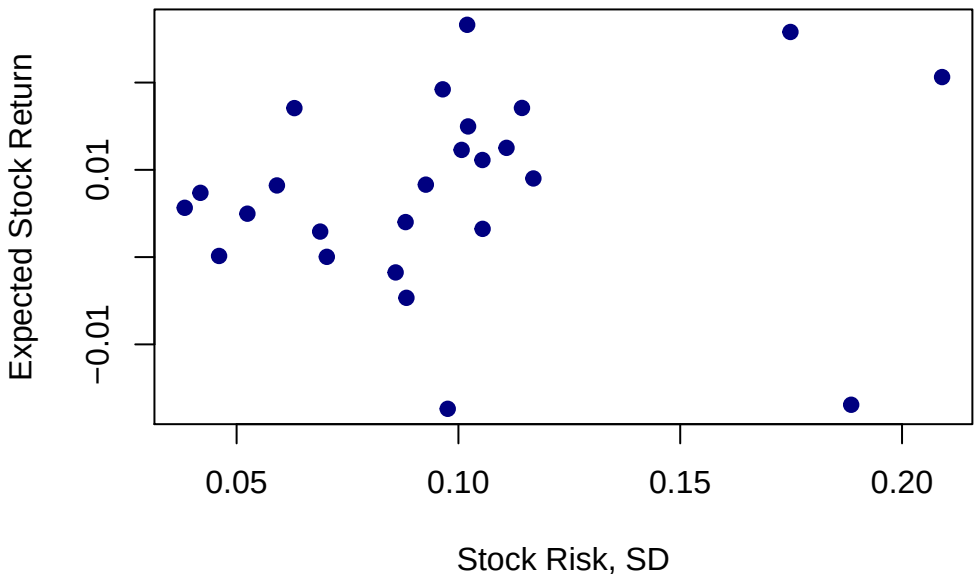


Figure: A robust positive correlation between risk and expected return

One possible way to formalize this notion is to quantify the amount of risk that an investor is willing to undertake for a given extra profit. The unit of expected return per risk is also known as Sharpe's ratio, S (for the original paper, see Sharpe 1994). It shows the connection between excess return, $r - r_f$ and the standard deviation, σ (here $r - r_f$ denotes the excess return). The Sharpe ratio is defined as follows:

$$S = \frac{E[r - r_f]}{\sigma}$$

This can either be calculated using the values at hand, or alternatively, many software utilities provide a command for direct calculation. We can alternatively use the Performance Analytics R package to obtain classic Sharpe ratio with a given confidence. If we assume a zero riskless return, then the Sharpe ratio reduces to the proportion of expected return over risk. A simple calculation follows.

```
sapply(stock04, mean)/sapply(stock04, sd)
```

```
##             C           KEY           WFC           JPM            SO
## -0.0897405764 -0.1781227215  0.0771060735  0.0895974205  0.1475191600
##           DUK             D            HE           EIX           LUV
##  0.1760079907  0.0948970508  0.0004637007  0.1390029397 -0.0204499058
##           CAL           AMR          AMGN          GILD          CELG
##  0.1475459270  0.0986906227  0.0455631714  0.2709179261  0.2610356681
```

```
##          GENZ          BIIB           CAT            DE           HIT
##  0.0424954171  0.0306943025  0.1129484643  0.1219634429 -0.0528374278
##           IMO           MRO           HES           YPF         SP500
##  0.1992902991  0.1465366419  0.1494971404  0.1055543496  0.0028584060
```

As a general rule, higher Sharpe ratios signal better risk-return trade-off and investors will tend to prefer them. In this sense the market trade-off is not very favorable (only 0.002), while other stock (e.g. GILD and CELG) offer much better trade-offs at around 0.26-0.27. At any rate we should note that those ratios are relatively low and reflect the unfavorable economic climate over the period of study.

Market Beta

When considering assets in financial markets, we can say that their risk is composed of two major components:

- *Market risk* - defined as the risk that financial market volatility will feed into the individual asset's price and return fluctuations. Financial markets tend to be very sensitive to a large set of new information and react, sometimes strongly, to developments of different importance and magnitude. Particular factors that tend to drive the whole market are news for the macroeconomy (such as growth, inflation, interest rates, unemployment, production, new construction, business and consumer sentiment), the political environment (balance of power, new elections, new policy, subsidies) and the legal environment (tax laws, regulations, rules of conduct). Those factors affect all securities and thus market risk cannot be diversified away.
- *Specific risk* pertains only to risks that are particular for a given asset or security. This may involve a number of operational, credit, liquidity, reputational, business, and strategic risks that affect only a given company. An erroneous marketing decision, or fraudulent behavior, or change of management present examples of this. Since this risk is idiosyncratic, it can be diversified by expanding the portfolio scope, and almost eliminated.

Total risk is then partly diversifiable - the specific part is, and the market one is not. Therefore it is of interest to see what part of the asset volatility is driven by market fluctuations (i.e. what is the market risk). A way to quantify this is to use so-called "market beta". While the market beta is a concept grounded in theory we can observe and estimate its proxy in practice (also called the b-coefficient). We use r_M to denote the market return and r_i to denote the return of a given security. By regressing the latter on the former, we reach the following equation:

$$r_i = \alpha + \beta * r_M + \epsilon$$

Beta is thus calculated to be:

$$\beta = \frac{Cov(r_i, r_M)}{Var(r_M)}$$

Beta thus shows by what amount will the return on this asset change if the market fluctuates by 1 unit. With a beta of 1.5, as the market grows by 1%, the return of the stock will grow by 1.5%. The alpha in this case is a constant, referring to autonomous return. Using the data we have loaded we can calculate the market beta of Citigroup by means of a simple linear regression

```
summary(lm(stock04$C ~ stock04$SP500))
```

```
##
## Call:
## lm(formula = stock04$C ~ stock04$SP500)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.25675 -0.06251 -0.00251  0.04175  0.50194
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.01728    0.01813  -0.953    0.344
## stock04$SP500  2.76019    0.39710   6.951 3.54e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1405 on 58 degrees of freedom
## Multiple R-squared:  0.4545, Adjusted R-squared:  0.445
## F-statistic: 48.32 on 1 and 58 DF,  p-value: 3.543e-09
```

As a first general observation from the results, one notices that Citi is strongly connected to the market and reacts disproportionately to its volatility. The beta stands at 2.8, meaning that a market fall of 1% produces a decrease of Citi's shares by 2.8%; the rise is also symmetric and of the same value. This connection is highly statistically significant with $p < 0.001$ and the high R^2 also indicates high practical significance. A value of 0.45 of the adjusted R-squared indicates that the market dynamics explain up to 45% of Citi's returns variance. This is likely due to the fact that Citi itself is a huge provider of financial market services and its success hinges critically on favorable developments in the market itself.

The whole set of betas can be easily calculated using either the built-in utilities of available programs or through simple scripts. Additionally, market vendors sell pre-calculated betas for the needs of business analysis. We can calculate the betas on our data at hand using a simple script.

```
betas <- rep(0,25)
for (i in 1:25) {
```

```
  betas[i] <- lm(stock04[,i]~stock04$SP500)$coefficients[2] }
betas <- as.data.frame(betas); betas <- cbind(colnames(stock04), betas)
print(betas)
```

```
##    colnames(stock04)     betas
## 1                  C 2.7601864
## 2                KEY 0.5383856
## 3                WFC 1.3305675
## 4                JPM 1.0997124
## 5                 SO 0.3457367
## 6                DUK 0.4251353
## 7                  D 0.4878024
## 8                 HE 0.5585550
## 9                EIX 0.7215669
## 10               LUV 0.9726065
## 11               CAL 1.2464687
## 12               AMR 1.7208375
## 13              AMGN 0.5071715
## 14              GILD 0.4283621
## 15              CELG 0.4238480
## 16              GENZ 0.3097018
## 17              BIIB 0.5343878
## 18               CAT 1.8795856
## 19                DE 1.6704655
## 20               HIT 1.2576868
## 21               IMO 0.9513585
## 22               MRO 1.2720837
## 23               HES 0.9257900
## 24               YPF 0.8226616
## 25             SP500 1.0000000
```

Again we observe that firms more closely connected in their business model to the financial markets are much more strongly influenced by its dynamics, while productive companies tend to be less influenced by market risk (as signified by their betas for the period).

Project 4

Imagine that you are given the choice between winning EUR 100 with a probability of p = 0.4 and EUR 50 with a probability of p = 0.6 or a sure bet of earning EUR 65. Using a random number generator generate two sets of trials:

- Ten trials of 10 lotteries each and sum your total earnings; and
- A single trial of 100 lotteries and sum your total earning.

Report on results - both standard deviation and mean earnings from trials. How are

practical results different from the theoretical expectations? Assuming risk neutrality, which of the three options should one choose (the sure bet, the ten trials, or the single trial)?

Lecture Five: Risking It in the Financial Markets

Financial markets present useful laboratories in which we can measure risk and observe its quantitative impact on financial results. The observation that risk is highly correlated with expected returns is key but rarely enough for the proactive of investment. Key questions are related to their exact relation and its numeric value, and how this knowledge can help so that investors realize their desired return with the least possible exposure to negative risks.

We look at this questions by reviewing first the Capital Asset Pricing Model and calculate it using returns data on NYSE from 1996 through 2006. We then review the benefits of diversification and delve deeper into asset correlations. This puts the building blocks for optimal portfolio selection a la Markowitz (1965, 1968).

Capital Asset Pricing Model

The Capital Asset Pricing Model (or CAPM in short) aims to answer the question what is the appropriate rate of return given an asset's volatility to the market (its beta). The concept of return is viewed in terms of the risk premium over a certain riskless rate, and thus our focus is on excess return.

The model postulates that in pursuit of this returns, the markets are characterized by the following conditions:

- Rational utility-maximizing agents
- Complete market information
- Absence of market frictions (such as delays or transaction costs)

While those assumptions are not completely met in practice, the model still presents a good first approximation. Starting from the idea that investors will not take any more risk than necessary, it is clear that the higher the volatility (beta) of the asset, the higher its excess return should be. As usually we denote return of an asset as r_i, the risk free return as r_f, and the market return as r_M. They are related in the following equation:

$$E[r_i] - r_f = \beta_i(E[r_M] - r_f)$$

The coefficient beta can be calculated via means of ordinary least squares (OLS) and is as follows:

$$\beta_i = \frac{Cov(r_i, r_M)}{Var(r_M)}$$

This can also be written as:

$$E[r_i] = r_f + \beta_i(E[r_M] - r_f)$$

CAPM thus allows us to calculate the appropriate return of an asset, given market return, riskless return, and the level of volatility.

We can estimate CAPM using either a simple linear regression or through some of the built-in functions of econometric and statistical packages. For our example we will use the data set "managers" from the PerformanceAnalytics R package.

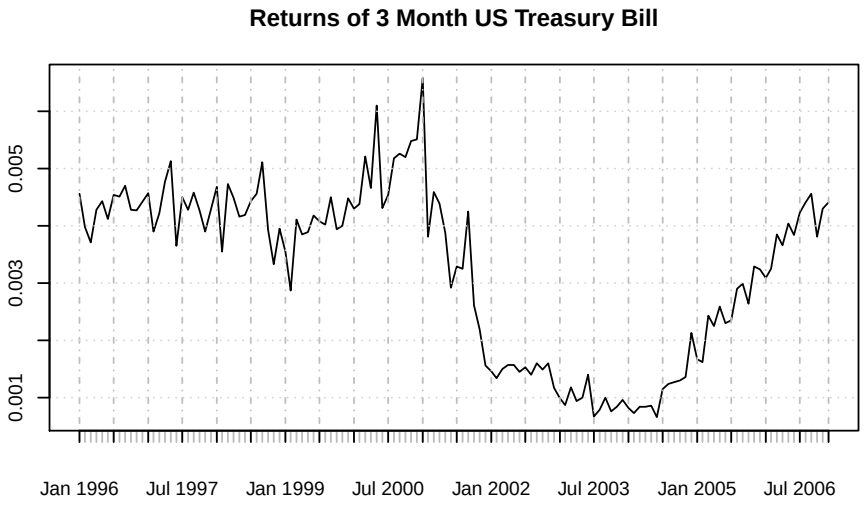
```
library(PerformanceAnalytics)
library(psych)
data(managers)
describe(managers, skew=F)
```

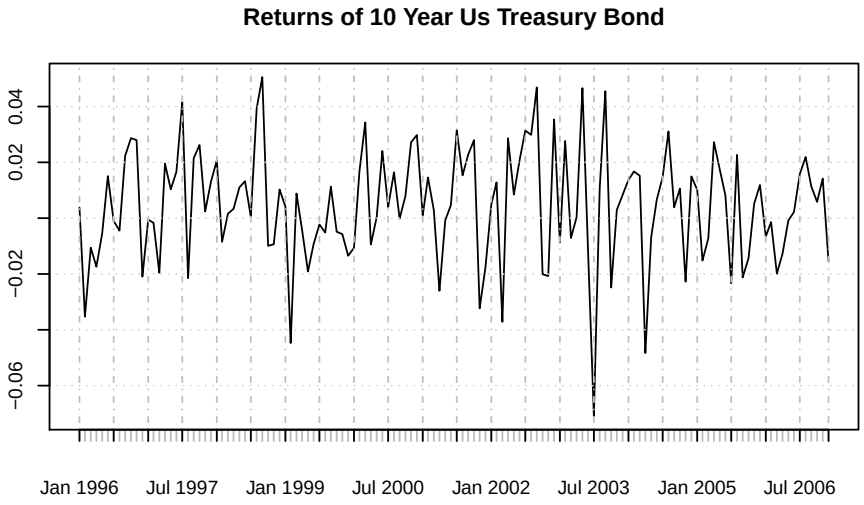
```
##              vars   n mean   sd   min  max range   se
## HAM1            1 132 0.01 0.03 -0.09 0.07  0.16 0.00
## HAM2            2 125 0.01 0.04 -0.04 0.16  0.19 0.00
## HAM3            3 132 0.01 0.04 -0.07 0.18  0.25 0.00
## HAM4            4 132 0.01 0.05 -0.18 0.15  0.33 0.00
## HAM5            5  77 0.00 0.05 -0.13 0.17  0.31 0.01
## HAM6            6  64 0.01 0.02 -0.04 0.06  0.10 0.00
## EDHEC LS EQ     7 120 0.01 0.02 -0.06 0.07  0.13 0.00
## SP500 TR        8 132 0.01 0.04 -0.14 0.10  0.24 0.00
## US 10Y TR       9 132 0.00 0.02 -0.07 0.05  0.12 0.00
## US 3m TR       10 132 0.00 0.00  0.00 0.01  0.01 0.00
```

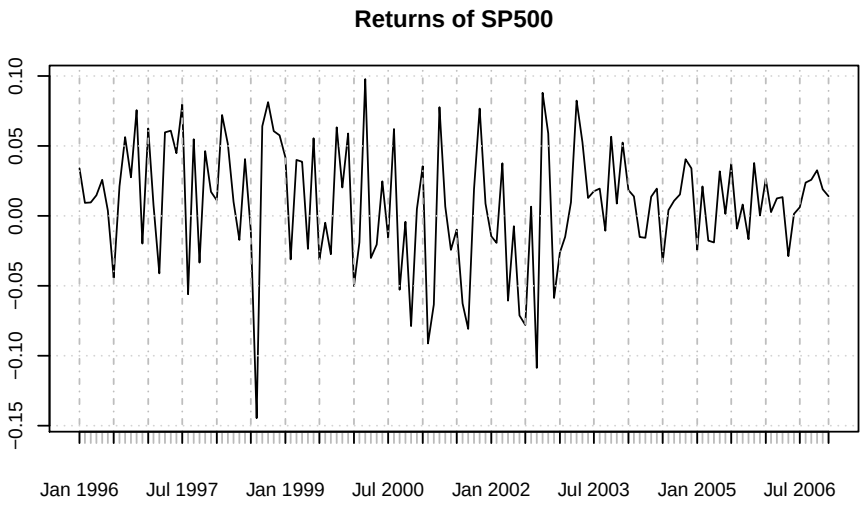
This data set contains returns on SP500, EDHEC (hedge funds index), and two proxies for the riskless rate - US treasury bills and bonds with maturities of 3 months and 10 years. In addition to that there is data on 6 Hypothetical asset managers - or alternatively six portfolios of securities over parts of the period.

We can see the relative dynamics of more risky securities and less risky securities with respect to the market. to do this we plot together the bill and bond fluctuations and the overall SP500 index volatility. We would expect to see very little return and fluctuations of the short term T-bill, more in the longer-term T-bond and the largest volatility in the market index. This is indeed what we observe.

```
par(mfrow=c(3,1))
plot(managers[,10], main="Returns of 3 Month US Treasury Bill")
plot(managers[,9], main="Returns of 10 Year Us Treasury Bond")
plot(managers[,8], main="Returns of SP500")
```

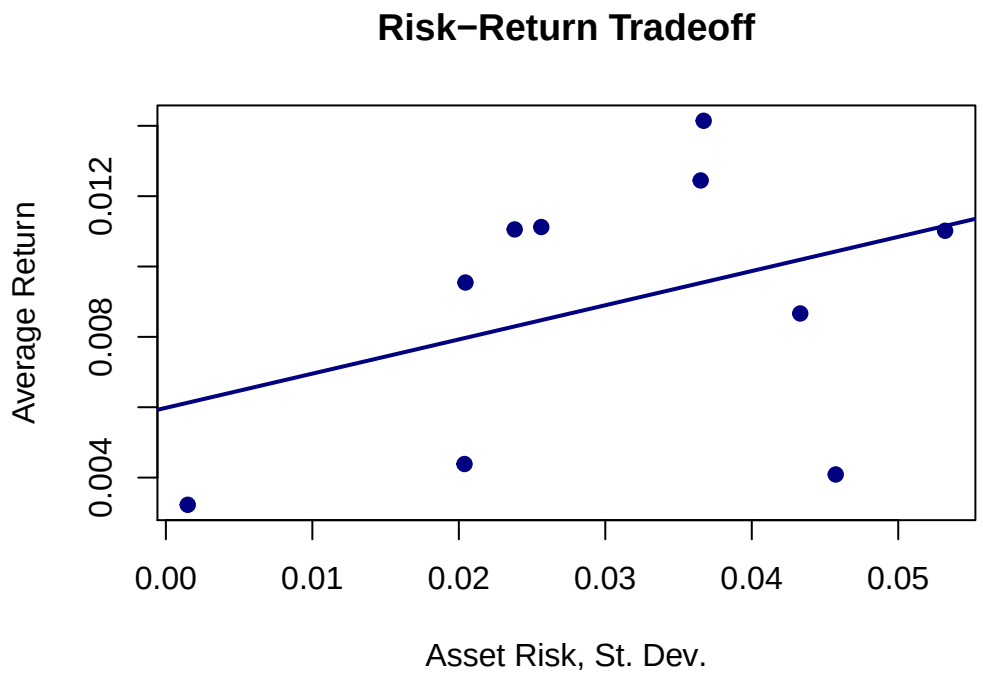

Returns of 3 Month US Treasury Bill
0.001
0.003
0.005
Jan 1996
Jul 1997
Jan 1999
Jul 2000
Jan 2002
Jul 2003
Jan 2005
Jul 2006


Returns of 10 Year Us Treasury Bond
−0.06
−0.02
0.02
0.04
Jan 1996
Jul 1997
Jan 1999
Jul 2000
Jan 2002
Jul 2003
Jan 2005
Jul 2006


Returns of SP500
−0.15
−0.10
−0.05
0.00
0.05
0.10
Jan 1996
Jul 1997
Jan 1999
Jul 2000
Jan 2002
Jul 2003
Jan 2005
Jul 2006

The three month US T-Bill has a return between 0.1% and 0.5%, the 10 year one - between 4% and -6%, and the market - between 10% and -15%. We would expect to observe even greater variability in the case of individual securities. To observe the relationship between risk and return we plot the returns and the standard deviations on a single scatter plot.

```
sd <- sapply(managers, sd, na.rm=TRUE)
r <- sapply(managers, mean, na.rm=TRUE)
plot(sd, r, col="navyblue", pch=19, xlab="Asset Risk, St. Dev.",
     ylab="Average Return", main="Risk-Return Tradeoff", size=3)
abline(lm(r~sd), col="navyblue", lwd=2)
```



Here we observe the expected positive relationship between risks and returns. To formalize it, we can calculate CAPM for all the time series in the data at hand. As a natural benchmark for the riskless rate, we take the three month US treasury bill. The market return is approximated by the S&P 500 Index. To calculate the betas, we use the CAPM.beta command from the PerformanceAnalytics package.

```
CAPM.beta(Ra = managers[,-c(8,10)],Rb = managers[,8], Rf = managers[,10])
```

```
##                      HAM1      HAM2      HAM3      HAM4      HAM5      HAM6
## Beta: SP500 TR 0.3900712 0.3383942 0.5523234 0.6914073 0.3208326 0.3235414
##                 EDHEC LS EQ  US 10Y TR
## Beta: SP500 TR   0.3341502 -0.0793304
```

The biggest beta values are found in HAM3 and HAM4 (0.55 and 0.69, respectively) indicating greatest risk levels. Other alternatives have betas of around 0.33 to 0.39, showing

less risk. Unsurprisingly, the safest asset is the ten-year US treasury bond which fluctuates only very mildly with the market, and takes on average an opposite direction. Once the appropriate asset (or group of assets) beta is calculated, the analyst can estimate its appropriate return, given the market return and the riskless rate.

Diversification and the Riskless Portfolio

So far we have focused on formally measuring and understanding risk and its trade-off with expected return. A simple measure of this is the Sharpe ratio, which we covered in Lecture Four We can calculate it in order to better appreciate the excess return per unit of risk. Using the SharpeRatio command we can obtain of set of different ratios, depending on their denominator. We use the default 95% confidence interval.

```
SharpeRatio(R = managers[,-10], Rf = managers[,10], p = 0.95)
```

```
##                                          HAM1      HAM2      HAM3       HAM4
## StdDev Sharpe (Rf=0.3%, p=95%): 0.3081020 0.2988608 0.2525301 0.14643845
## VaR Sharpe (Rf=0.3%, p=95%):    0.2306863 0.3970699 0.2504936 0.09553906
## ES Sharpe (Rf=0.3%, p=95%):     0.1295014 0.1788256 0.2093343 0.06625013
##                                         HAM5      HAM6 EDHEC LS EQ
## StdDev Sharpe (Rf=0.3%, p=95%): 0.03545540 0.3785371   0.3142695
## VaR Sharpe (Rf=0.3%, p=95%):    0.02399862 0.3022965   0.2737607
## ES Sharpe (Rf=0.3%, p=95%):     0.01664487 0.2308450   0.1855867
##                                   SP500 TR  US 10Y TR
## StdDev Sharpe (Rf=0.3%, p=95%): 0.12558293 0.05684359
## VaR Sharpe (Rf=0.3%, p=95%):    0.07957460 0.03741555
## ES Sharpe (Rf=0.3%, p=95%):     0.05760917 0.02610548
```

Those metrics allow us to understand the additional return we obtain per unit of risk (defined as either the standard deviation, the expected shortfall, or the value at risk). This is useful but the key topic of risk management is *taking the optimal amount of risk* that is connected to a certain desired return. Thus the risk management professional needs to also focus on ways to mitigate risk.

A very common approach for that is **diversification**. In the case of financial market positions this means that resources are spent on different assets, thus decreasing the probability that all of them will perform exceptionally badly (or exceptionally well) at the same time. Diverse assets are combined in a portfolio, which usually has lower risks that any of the individual assets.

We can illustrate with the data at hand. The standard deviation of HAM1 stands at $\sigma_{HAM1} = 0.03$, of HAM3 - at $\sigma_{HAM3} = 0.04$, and of HAM4 - at $\sigma_{HAM4} = 0.05$. Imagine that we combine them in equal proportions in a new portfolio, and calculate its risk, σ_c.

```
sd(managers$HAM1*0.33+managers$HAM3*0.33+managers$HAM4*0.34)
```

```
## [1] 0.03161513
```

It stands at $\sigma_c = 0.03$. Even though we now hold portfolios that used to have individual risks of 0.04, and 0.05, in combination their overall risk stands at 0.03. This is due to the fact that of the two large types of risks (market and specific), the individual risk is diversifiable. A single asset bears its own risk, described by the standard deviation. Bundling multiple assets into a portfolio may lead to a decrease in risk as assets move in different directions. As number of assets increases, portfolio risk decreases and converges to overall market risk. Thus a well-diversified portfolio leads to (near) elimination of specific risk.

More formally, we claim that when we construct a portfolio of two assets with returns r_1 and r_2, each of them taking a certain proportion w_1 and w_2 of the overall portfolio, we obtain a return of r_p. It is equal to:

$$r_p = w_1 * r_1 + w_2 * r_2$$

.

The risk of this portfolio, σ_p is:

$$\sigma_p = \sqrt{(w_1 * \sigma_1)^2 + (w_2 * \sigma_2)^2 + 2\rho(w_1 * \sigma_1)(w_2 * \sigma_2)}$$

Here with ρ we denote the Pearson correlation between assets, which is defined as follows:

$$\rho = \frac{Cov(r_1, r_2)}{\sigma_1 * \sigma_2}$$

The higher the positive correlation, the more the two assets move together. Negative correlations indicate that assets move in different directions - as the one loses value, the other one gains.

It is precisely negative correlations that allow diversification to work. Intuitively, if one of the assets loses money, the other one will gain, thus decreasing overall portfolio losses. Thus positive correlations increase risk, while negative ones decrease it.

At the extreme case of perfect negative correlation $\rho = -1$ we can construct a completely riskless portfolio. With $\rho = -1$, the portfolio risk equation reduces to the following:

$$\sigma_p = \sqrt{(w_1 * \sigma_1)^2 + (w_2 * \sigma_2)^2 - 2(w_1 * \sigma_1)(w_2 * \sigma_2)} = \sqrt{(w_1 * \sigma_1 - w_2 * \sigma_2)^2}$$

We thus obtain:

$$\sigma_p = |w_1 * \sigma_1 - w_2 * \sigma_2| = |w_1 * \sigma_1 - (1 - w_1) * \sigma_2|$$

The portfolio risk reaches zero, $\sigma_p = 0$, when the following conditions hold:

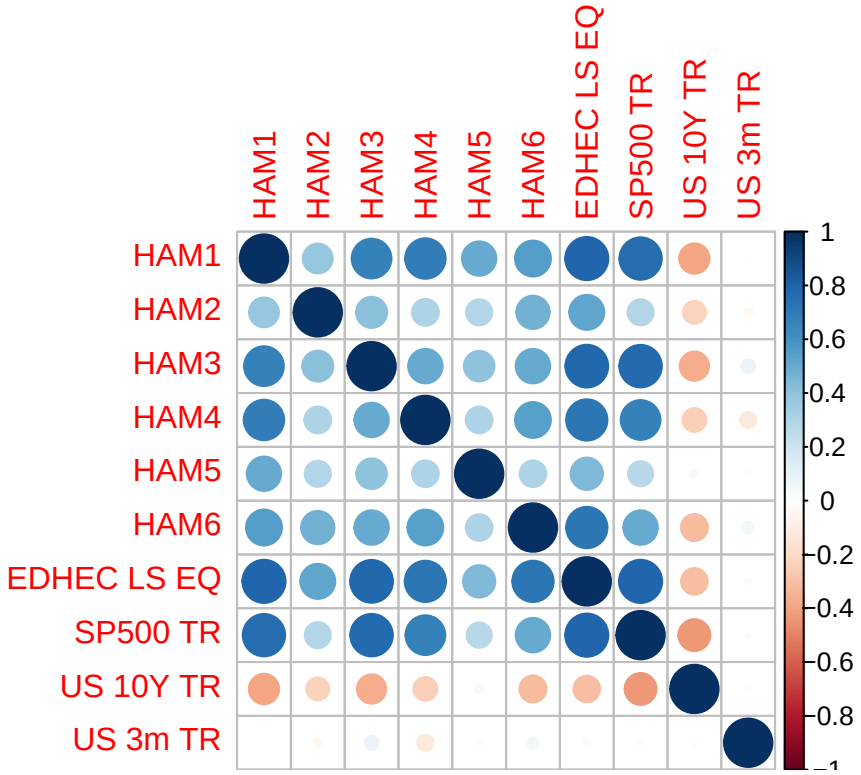
$$\frac{w_1}{1 - w_1} = \frac{\sigma_2}{\sigma_1}$$

In this case every movement in the negative by asset 1 is exactly offset by a positive movement by asset 2, and vice versa. Thus no loss is realized, but also the upside potential is minimized. In reality it is very rarely the case that any two assets can exhibit such a strong and clear-cut negative relationship, especially over a prolonged period of time. This means that while risk can be minimized, it can never be truly eliminated.

The correlation coefficients (co-movements) are crucial for RM professionals and oftentimes analysts need to quickly see the correlational patterns - either through a visual inspection or through a more formal numeric review. For the visual inspection one can use many tools, among which is the `corrplot` package in R. It presents visually the correlations and color-codes them: by default the blue ones are positive, and the red ones - negative. The circle size corresponds to the size of the correlation coefficient.

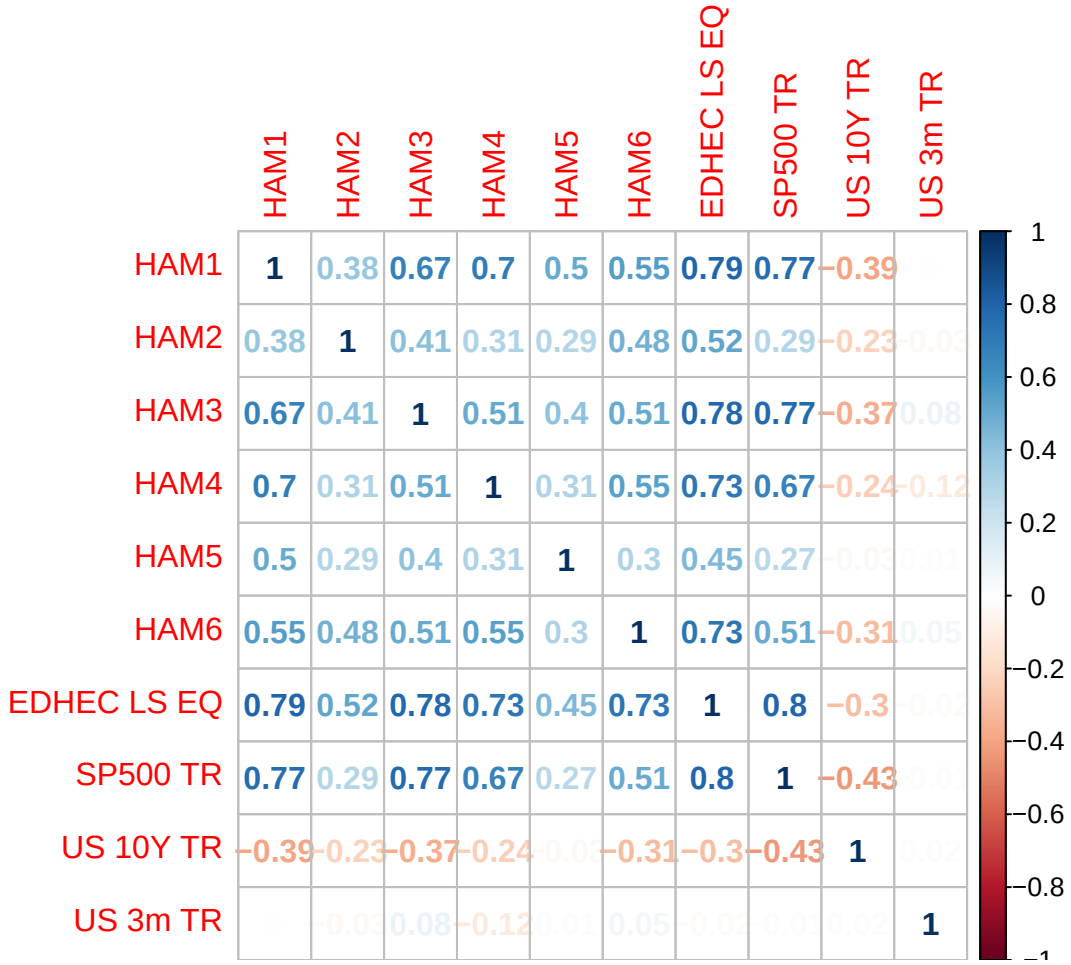
Using data from October 2001 on, so that there are no missing observations, we obtain the following.

```
library(corrplot)
corrplot(cor(managers[69:132,]), method="circle")
```



Alternatively, we can also inspect numeric values.

```
library(corrplot)
corrplot(cor(managers[69:132,]), method="number")
```



Since correlation matrices are symmetrical, it is sometimes a good idea to combine visuals with numerics. As an overall conclusion we can say that most HAMs are positively and strongly connected to market dynamics and so is the EDHEC index. Combining assets from those groups into a single portfolio has limited scope for decreasing risk. The only consistently negative correlational pattern that we observe is the one exhibited by the 10 Year US Treasury bond. Adding it to the portfolio holds some potential for risk reduction. As the market (S&P 500) will be going down by 1% the US 10Y TR will move in the opposite direction and go up by 0.43% thus eliminating half of the losses in case of equal amount of shares.

While the co-movement of assets is crucial for risk management, one should be careful as historically calculated correlations may change abruptly and thus increase risk unexpectedly. This is especially true during economic downturns when negative correlations turn to positive ones as the whole market is decreasing. In such cases when the investor needs diversification most, it is also most likely to fail.

The Efficient Frontier

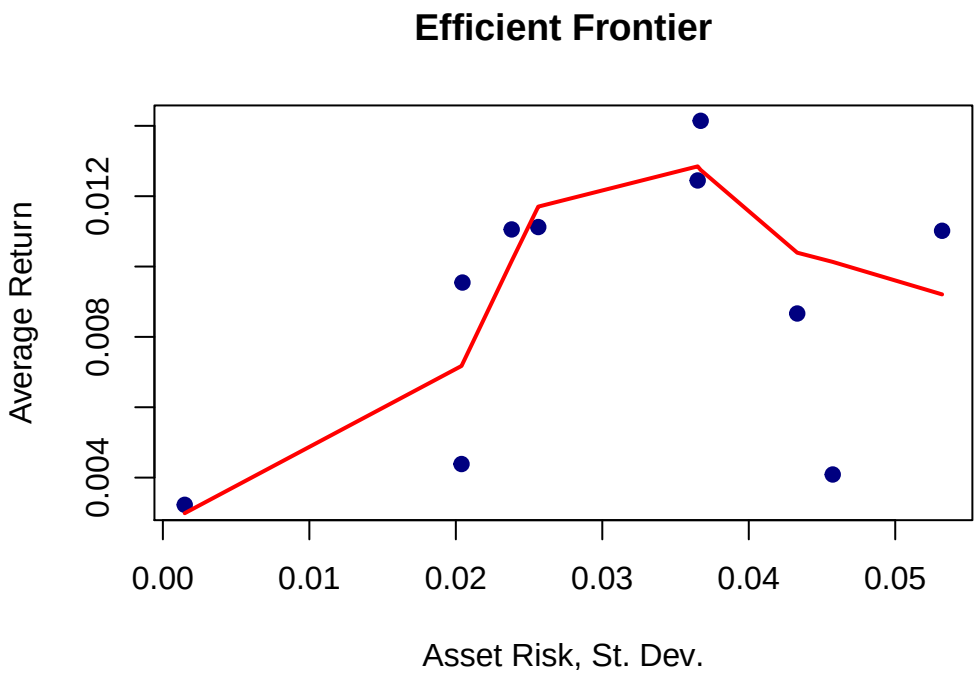
A way to formalize the intuition behind the optimum risk-return trade-off is to build the efficient market frontier. It involves plotting all possible portfolios along their risk-return dimensions and to connect all the best performers. A given portfolio A outperforms B if for a given level of risk it offers a greater return (or, conversely, it provides lower risk for a given return).

As the market develops rational investors will not buy dominated portfolios and will thus either make them disappear or rise in profitability, thus pushing them to the efficient frontier. Depending on an investor's risk preference he can pick the best portfolio for him or her along the frontier.

Alternatively, the capital allocation line provides all possibilities in which the investor combines riskless assets with risky assets to obtain the exactly desired trade-off. A rational investor will thus make his choice when the CA line is exactly tangent to the efficient frontier.

Here we will try to construct empirically the frontier using the data at hand. We plot the assets along their returns and standard deviations on a scatter plot. Then a loess line that best describes the point is drawn, thus dividing the portfolios below and above the efficient frontier. The portfolios above the line are outperforming those below and should be preferred. The only anomaly here is an asset with high risk and relatively low return at the rightmost part of the graph.

```
plot(sd, r, col="navyblue", pch=19, xlab="Asset Risk, St. Dev.",
     ylab="Average Return", main="Efficient Frontier",
     size=3) + lines(lowess(r~sd), col="red", lwd=2)
```




```
## integer(0)
```

The anomalous asset should not be selected as alternatives provide better risk-return trade-offs. In theory it should be eliminated by the market over time but it still remains in existence. Some market imperfections may persist over time, giving opportunities for particularly good (or bad) returns. This may be due to frictions, irrationality, or unimportance. At any rate, real financial markets do not conform perfectly to theory at all times, and the risk managers should be conscious to any possible deviations.

Forward-looking Numbers

So far we have been looking at risk and return as parameters that are calculated from past data. Both the average and the standard deviations can be estimated in such a way, and if, indeed, series are mean-reverting, this can be an adequate approximation. Additionally, this approach is the easiest to implement for the analyst.

Still, when investment choices are made, the risk lies not in the past but in the future, and so the relevant values are expected risk and expected return. It is possible that there can change, sometimes dramatically, as the economic or political headwinds change and that the past may be a poor guide to the future.

The RM professional can thus choose to adjust the trends from past data using information at her disposal, or giving greater weight to a group of observations. All in all, expected values may be obtained in some of the following ways:

- **Using historical data** - assuming that time series properties are time invariant will allow the usage of long-run historical figures. A typical example for that would be long-run economic growth or real interest rates.
- **Adjusting historical data** - if there is a regime shift, the analyst may choose to give more weight to more recent data, or to use smoothed data (deseasoned, detrended, moving average, etc). This is typically done in the case of unemployment or investment in volatile economies.
- **Forecasting expectations** - using statistical methods to forecast the movement of the asset, possibly supplementing them with additional non-statistical information (such as the one gleaned from the annual reports) Examples of this include forecasting market shares and sales.
- **Running simulations** - given a system with understandable behavior, the analyst can simulate it and reach the probability distribution of outcomes. This can give both an estimate of expectations and deviations, as well as the confidence intervals in which they fall. For example, exotic financial instruments can be evaluated using simulations.
- **Using expert knowledge** - if the systemic change is so large and pervasive, quantitative methods may be of little use. In this case expert knowledge and benchmark case studies can provide valuable insight. This holds particularly true in the case of extreme events such as economic and financial crises, or wars.

Thus there are many ways to reach values about expected return and risk. The RM analyst needs to be aware of all of them and appreciate both their strong and their weak points. A particularly fruitful approach is to combine the forecasts from different methods - a large body of forecasting literature shows that ensemble algorithms tend to outperform individual methods.

Project 5

Pick a financial market of your choice and obtain data for returns of at least 50 traded securities. Calculate their CAPM betas, using a reference riskless rate of your choice.

Construct the efficient frontier - are there assets below it? Why do they persist? Explain.

Lecture Six: Valuing Risk through Value at Risk

An important aspect of managing risk is being able to estimate the levels of risk and compare them against risk preference or expected return. It it thus useful to form a metric of what is the maximum expected loss, or what is the value that is under risk.

In this lecture we define formally the Value at Risk (VaR) and the Expected Shortfall (ES, or ETL) metric, and use an intuitive software implementation to calculate them. We then proceed to look at their strengths and weaknesses and put it in the context of recent controversies surrounding them. More details can be found in Crouhy et al. (2006).

Defining Value at Risk

Value at Risk is a metric that answers the question what is the maximum expected loss that can be incurred with certain probability over a given period of time. It can be applied to both individual assets and portfolios but individual VaRs do not aggregate into portfolio VaRs. The key question it answers is What is the maximum expected loss of this portfolio with a given probability for a given period of time, e.g. What is the maximum expected loss of this portfolio with probability of 99% over the next day if the market behaves normally?

More formally we can define it as follows. The value at risk VAR_α with a confidence level of $\alpha = 95\%$, for a given loss L is thus defined as:

$$VAR_\alpha = inf[l \in \mathbb{R} : P(L > l) \leq 1 - \alpha]$$

If we denote a well-defined distribution function of this random variable as F, then the VaR expression reduces to:

$$VAR_\alpha = inf[l \in \mathbb{R} : F_L(l) \geq \alpha]$$

Once the analyst has the distribution function of a given variable, she can easily find the VaR by locating the sum that lies at the desired percentile. For example if we are dealing with a normal distribution with a mean of 0 and a standard deviation of one, we can easily find what is the probability that we realize a loss of at least two units. This can be restated as the probability that the random variable takes values at minus two or less, and can be easily calculated.

```
pnorm(-2,0,1)
```

```
## [1] 0.02275013
```

An alternative way to think about it is to set a probability of a given loss, and see what amount corresponds to it. For example, what is the maximum amount that we lose with a given probability. If we lose at least 2 with a probability of p = 0.02275, then we are losing

less than that with a probability of *1 - p*, or 0.9772. In short, we are about 98% confident that losses will not exceed two units.

Using a given probability threshold we can easily calculate the sum that corresponds to it. This is done as follows.

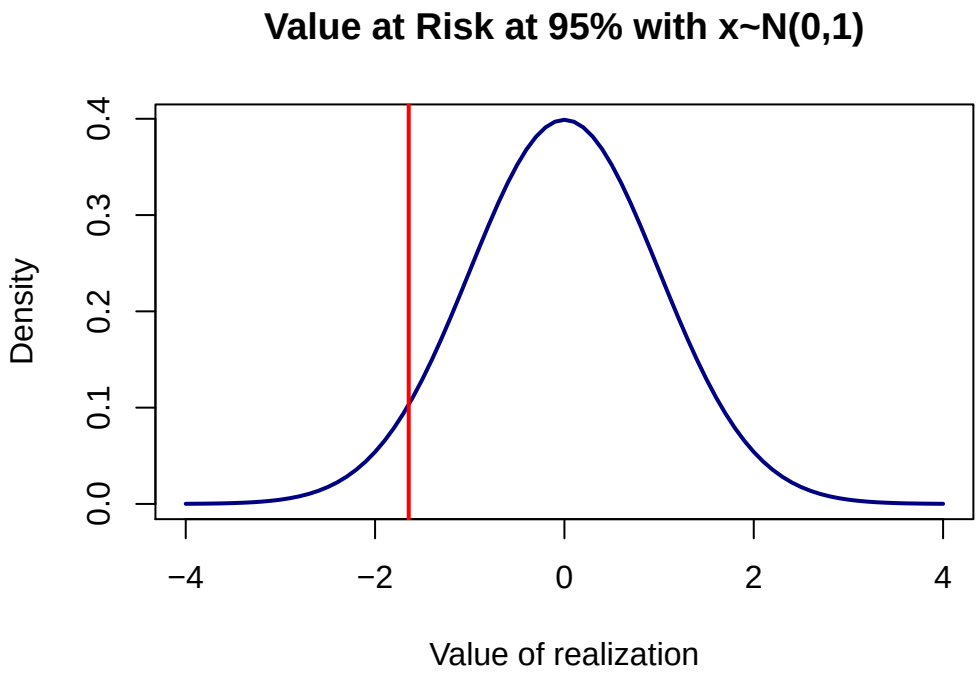
```
qnorm(0.02275013,0,1)
```

```
## [1] -2
```

The value at risk calculation thus corresponds to finding the true distribution of the variable (e.g. asset return) and calculate what is the largest possible loss 95% (or 99%) of the time. If the 95% VaR stands at 10 million, one would expect that 19 out of 20 days losses to be below that.

The concept is graphically illustrated in the following plot.

```
x <- seq(from = -4, to = 4, by=0.1)
plot(x, dnorm(x, 0, 1), col="navyblue", xlab="Value of realization",
     main="Value at Risk at 95% with x~N(0,1)", type="l", lwd=2,
     ylab="Density") + abline(v=qnorm(0.05, 0,1), col="red", lwd=2)
```

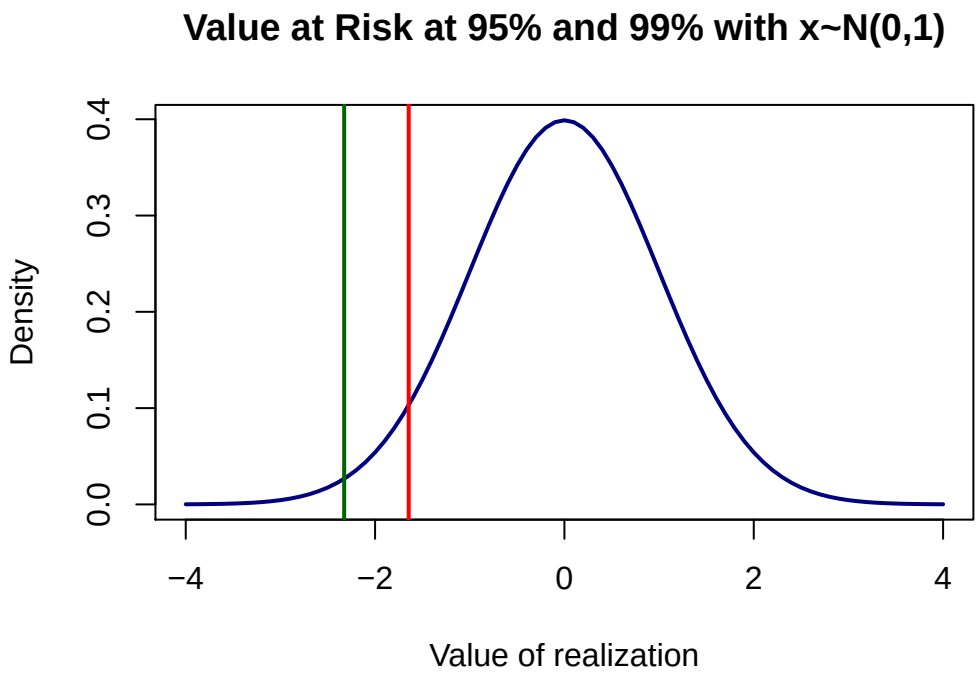



```
## integer(0)
```

We draw a normally distributed random variable with mean of 0, and standard deviation of 1, and then calculate the cut-off value at 5% (red vertical line). 95% of the time, the random variable will take values larger than that - i.e. this is the maximum expected risk 95% of the time.

Another line can be added to the graph that corresponds to the 99% level. It is to the left of the 95% line, which shows that extremely large deviations from the expectation will be relatively rare if the data is normally distributed.

```
x <- seq(from = -4, to = 4, by=0.1)
plot(x, dnorm(x, 0, 1), type="l", lwd=2, col="navyblue", ylab="Density",
     main="Value at Risk at 95% and 99% with x~N(0,1)",
     xlab="Value of realization") + abline(v=qnorm(0.05, 0,1), col="red",
     lwd=2) + abline(v=qnorm(0.01, 0,1), col="darkgreen", lwd=2)
```




```
## integer(0)
```

For convenience we have used the normal distribution for illustration. It is particularly notable in that it is symmetric around its mean and large profits and large losses are equally probable. Real-life data may follow different distributions but the RM professional can use the same logic and approach to measure the maximum expected risk with a given confidence level.

Calculating Value at Risk

The key problem in calculating Value at Risk is finding the true data distribution or at least a close approximation to it. We should not forget that returns data we often have at hand is merely a sample from the larger population of possible realizations of the data-generating process. Often, normality is assumed while returns may in reality follow a different

distribution.

There are three common ways to derive the underlying data distribution:

- **Parametric Estimation** - this approach uses historical data to estimate a distribution's type and parameters. Usually this begins by inspecting data and making assumptions about its form and then estimating the best-fitting distribution through an optimization or parameter search.
- **Non-parametric estimation** - this approach uses historical data to make simulations but without imposing assumptions on the data structure. The simulation proceeds by re-sampling past values of both returns and other relevant variables, and using them to form new scenarios. The statistical properties and the value of these simulations are then used to construct an estimate of returns distribution.
- **Monte Carlo Methods** - this approach consists of constructing and parametrizing a model that captures well the dynamics of the variable under interest (e.g. stock price). The model is then run through many simulations and aggregate statistics about its behavior are used to construct the distribution.

Probably the easiest approach to calculation is to use parametric historical data but the analyst should be careful as the usual pitfalls with time series are present - both technical and conceptual. At any rate, VaR calculation is widely available in numerous statistical packages and can be easily done using R.

For this purpose we are going to use the EDHEC data set - this is data for composite hedge fund index returns. Each of the funds contains 152 monthly observations with varying degrees of risk - ranging form almost none in the Fixed Income Arbitrage fund, to quite a bit in the Short Selling fund.

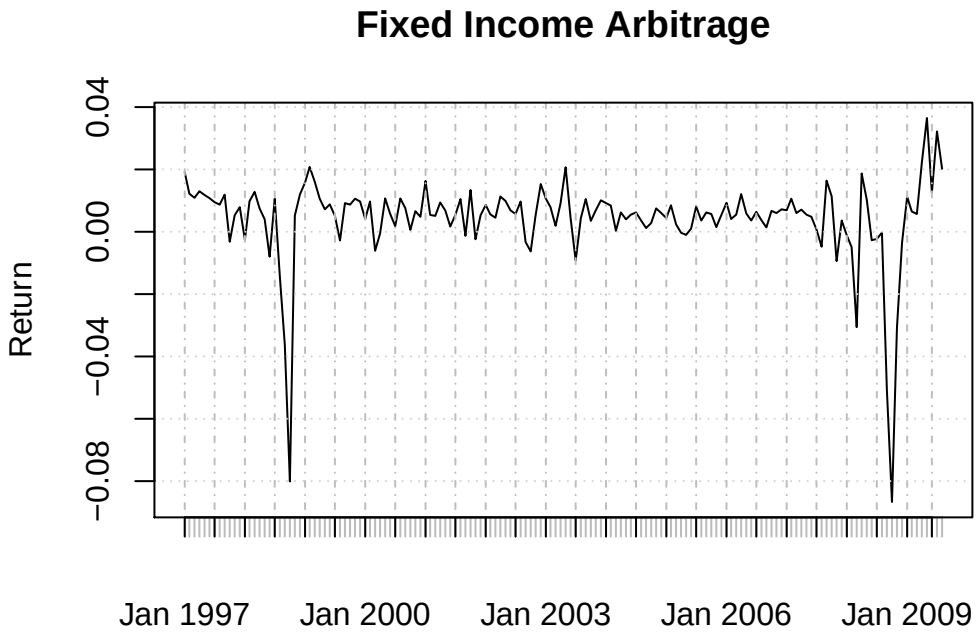
```
library(PerformanceAnalytics)
library(psych)
data(edhec)
describe(edhec, ranges = F, skew=F)
```

```
##                          vars   n mean   sd se
## Convertible Arbitrage       1 152 0.01 0.02  0
## CTA Global                  2 152 0.01 0.03  0
## Distressed Securities       3 152 0.01 0.02  0
## Emerging Markets            4 152 0.01 0.04  0
## Equity Market Neutral       5 152 0.01 0.01  0
## Event Driven                6 152 0.01 0.02  0
## Fixed Income Arbitrage      7 152 0.00 0.01  0
## Global Macro                8 152 0.01 0.02  0
## Long/Short Equity           9 152 0.01 0.02  0
## Merger Arbitrage           10 152 0.01 0.01  0
## Relative Value             11 152 0.01 0.01  0
## Short Selling              12 152 0.00 0.06  0
```

```
## Funds of Funds            13 152 0.01 0.02  0
```

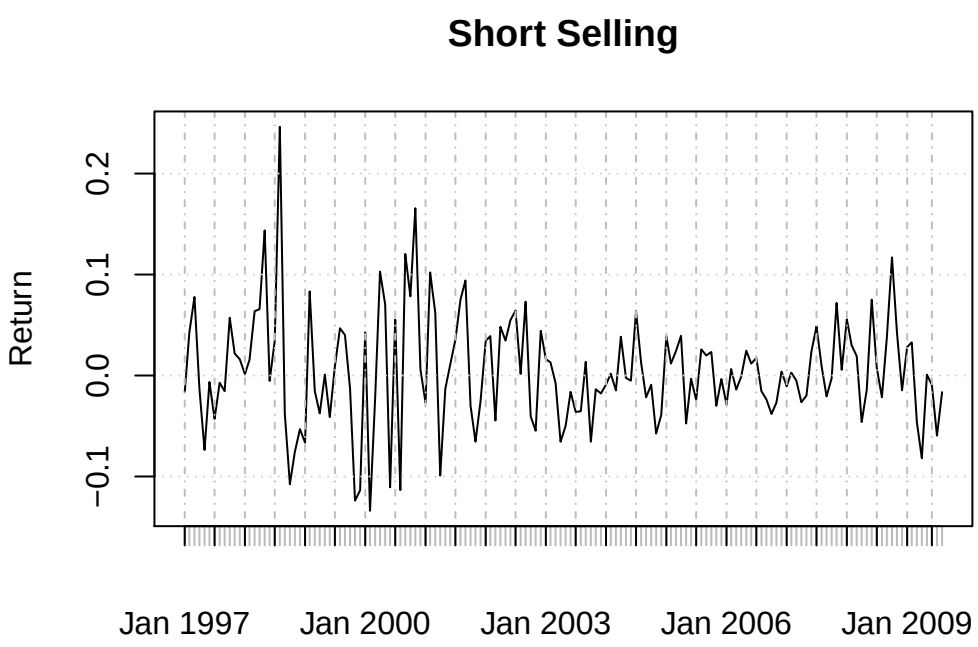
We can easily discern the two most contrasting funds with highest and lowest risk as measured by the standard deviation. Again, we make the observation that fixed income securities are much less risky and only fluctuate significantly in the presence of global risks or issues with large sovereign issuers.

```
plot(edhec$"Fixed Income Arbitrage", main="Fixed Income Arbitrage",
     ylab="Return")
```



Short selling is typically more risky due to the high level of speculation and to the fact that sometimes short selling is financed through leverage. We must particularly note that leverage serves to significantly amplify risks and should be utilized with care.

```
plot(edhec$"Short Selling", main="Short Selling", ylab="Return")
```



Value at risk at the desired overall level for the data at hand can be easily calculated by the VaR command in the package *PerformanceAnalytics*. We calculate the 95% VaR using the assumption that data is normally distributed. For convenience we transpose the results and multiply by 100 to reach percentage values.

```
100*t(VaR(edhec, p = 0.95, method="gaussian"))
```

```
##                                VaR
## Convertible Arbitrage  -2.6457816
## CTA Global             -3.4710978
## Distressed Securities  -2.2126899
## Emerging Markets       -5.4989269
## Equity Market Neutral  -0.8761813
## Event Driven           -2.2462022
## Fixed Income Arbitrage -1.9001981
## Global Macro           -2.0230181
## Long/Short Equity      -2.8592642
## Merger Arbitrage       -1.1524776
## Relative Value         -1.4930492
## Short Selling          -8.6170270
## Funds of Funds         -2.3938882
```

The calculation here shows that we are 95% certain that we will not lose more than 2.54% of the holdings in the Convertible Arbitrage fund, more than 0.87% in the Equity Market Neutral Fund, and no more than 8.62% in the Short Selling fund. We can recalculate those

numbers to reach 99% of a given loss.

```
100*t(VaR(edhec, p = 0.99, method="gaussian"))
```

```
##                                VaR
## Convertible Arbitrage   -4.007498
## CTA Global              -5.178111
## Distressed Securities   -3.458969
## Emerging Markets        -8.118887
## Equity Market Neutral   -1.487900
## Event Driven            -3.492656
## Fixed Income Arbitrage  -2.862782
## Global Macro            -3.179074
## Long/Short Equity       -4.365418
## Merger Arbitrage        -1.911081
## Relative Value          -2.389296
## Short Selling          -12.359632
## Funds of Funds          -3.630933
```

In this case we are 99% confident that we will not lose more than 4% of the holdings in the Convertible Arbitrage fund, more than 1.49% in the Equity Market Neutral Fund, and no more than 12.36% in the Short Selling fund. Alternatively, the actual losses will exceed those numbers in one out of a 100 trading days. Now that those expected losses are calculated, the firm can multiply them by their holding and reach a total value of its expected monetary loss.

There are three common variations of the VaR calculation, which take into account different assumptions about the data distributions. Since the numbers may differ significantly across those calculations, the analyst is well advised to be conscious of the choice.

- **Historical Method** - no assumption is made of the data distribution, but merely uses the historical data at hand to calculate the VaR metric.
- **Gaussian** - the data are assumed to follow the normal distribution and the VaR is calculated parametrically.
- **Modified** this approach uses the Cornish-Fisher modification that corrects for non-normality of the data and possible skewness.

Those three approaches produce different numbers and there needs to be a deliberation on which to use. Oftentimes there is official company policy, methodology or guidelines that answer this question. In their absence, the RM professional will need to make a judgment call. As an illustration to those possible differences we calculate VaR using those methods at the 99% confidence.

```
x <- cbind(100*t(VaR(edhec, p = 0.99, method="historical")),
           100*t(VaR(edhec, p = 0.99, method="gaussian")),
           100*t(VaR(edhec, p = 0.99, method="modified")))
colnames(x) <- c("Historical VaR", "Gaussian VaR", "Modified VaR"); print(x)
```

```
##                         Historical VaR Gaussian VaR Modified VaR
## Convertible Arbitrage         -6.6592    -4.007498   -10.092227
## CTA Global                    -5.0191    -5.178111    -4.847019
## Distressed Securities         -6.4393    -3.458969    -6.533764
## Emerging Markets             -11.5301    -8.118887   -13.971949
## Equity Market Neutral         -2.0850    -1.487900    -4.404136
## Event Driven                  -6.2598    -3.492656    -6.385154
## Fixed Income Arbitrage        -6.5055    -2.862782    -5.850228
## Global Macro                  -2.8309    -3.179074    -2.437999
## Long/Short Equity             -5.8973    -4.365418    -5.508705
## Merger Arbitrage              -2.7141    -1.911081    -3.630211
## Relative Value                -4.3753    -2.389296    -5.053100
## Short Selling                -11.8698   -12.359632   -12.223601
## Funds of Funds                -6.0784    -3.630933    -5.500037
```

Depending on assumptions, the VaR metric can produce widely divergent results. The range from the lowest to the highest VaR estimates in the table can sometimes vary by up to 5-6%. Over large sums this means loss uncertainty of tens or even hundreds of millions. Such large differentials in the VaR metric can impact executive decisions and lead the organization into taking unwanted risk. The RM process should therefore take account of multiple indicators and make intelligent decisions on the most appropriate set of indicators to be used.

Strengths and Weaknesses

The VaR metric is widely criticized but no less widely used and even required for some companies under regulation (e.g. banks and financial intermediaries). It naturally has both strengths and weaknesses and its usage requires an intelligent understanding of both.

Among its many **strengths** we should particularly note the following:

- *Provides a common, consistent and integrated measure of risk* - VaR is a measure that is commonly understood and provides an integrated high-level overview of total organizational risk. This allows for both better RM practice and more streamlined executive decisions. If the methodology remains invariant (e.g. confidence levels, assumptions) it will also yield consistent deterministic estimates.
- *It can provide an aggregated measure of risk and risk-adjusted performance* - this is useful not only for strategic and tactical decision-making but also for designing corporate policy and incentive schemes. The VaR can guide management into understanding their risk profile and incentivize or disincentivize certain actions or individuals depending on their risk profile.
- *Provides firms the opportunity to assess the benefit of portfolio and activity diversification* - as diversification increases into different assets or activities, the VaR metric will move, pointing at the beneficial effects of these actions on the overall risk profile.

- *It is easy to communicate and understand, compressing risk in a single number* - finally, and not trivially, the metric is easy to explain and understand, thus making for a large possible audience that can use it and appreciate it. The overall risk of the organization (or portfolio, position, activity, or department) is contained in a single number with clear business implications (e.g. loss of 5 million at most in 99 of 100 days).

Its many strengths are counterbalanced by a few key **weaknesses** that should be taken into account whenever using VaR:

- *Requires serious assumptions that may not be met in practice* - the VaR calculation imposes the assumptions of well-defined (usually normal) distribution, normal market conditions, and time invariant parameters. Those may not be realistic especially during times of economic and financial turbulence. However, the assumptions will skew the VaR number (sometimes significantly) and provide an inaccurate picture of risk.
- *The VaR metric focuses on quantitatively measurable risks* - the focus of the metric is on quantifiable and quantitatively measured risks, that are fed into an econometric model. In practice however qualitative risks may have large impact (e.g. new government in Russia) and indeed drive volatility, returns, and losses.
- *Encourages excessive risk taking and gives false sense of precision and comfort* - the very presence of a metric may lead management to take excessive risk that they believe to be well-calculated while in fact they are not. This illusion of knowledge may in fact worsen the risk profile of the organization and executives are led to think they are walking on the knife's edge of their risk preference, while they are far above it.
- *It is a rough guide to expected maximum loss – realized loss may well exceed it* - finally one should not forget the VaR metric is only a rough approximation to risk and what is observed in practice may well be very different, even if all caution is taken for the VaR number to be calculated and interpreted correctly.

It is precisely of those downfalls that the metric has come under serious criticism as the global financial crisis of 2008-2009 engulfed the world economy.

Expected Shortfall

A major drawback of the Value at Risk metric is that while it says what will be maximum loss in α percent of the cases, it fails to appreciate what is the expected loss in those $1-\alpha$ percent of cases. The Expected Shortfall (ES) metric can be constructed to remedy that. If the VaR calculates what will happen if things do not get too bad, the ES metric calculates what happens in the cases that they do.

For example, the 95% VaR metrics on two assets may be both 1 million. This means that in 95% of the cases losses will not exceed 1 million. However in the 5% of the cases that they do, the first asset's expected loss is 2 million, while the second asset's is 5 million. Clearly the former is less risky than the latter. The ES estimate gives a measure exactly of the

expected loss given we have broken the VaR value. More formally it can be defined as follows (we use the VaR_γ to denote VaR at the γ level and α to denote ES at level α):

$$ES_\alpha = \frac{1}{\alpha}\int_0^\alpha VaR_\gamma(X)d\gamma$$

Alternatively, if the underlying distribution is a continuous one, then the ES is equal to the tail conditional expectation (TCE):

$$TCE_\alpha(X) = E[-X|X \leq VaR_\alpha(X)]$$

By its way of calculation an expected shortfall at a given level is equal to, or (typically) larger than the VaR at that level. In this sense, ES is a more conservative metric and should be used whenever greater uncertainty prevails. It also provides a fuller and more nuanced picture of risk, especially suitable for fat-tailed distributions, that could aid the RM professional in better understanding and mitigating risks.

The ES can be easily calculated in numerous software utilities. We should keep in mind that it is known under different names the most common being Expected Shortfall (ES), Expected Tail Loss (ETL), and Conditional Value at Risk (cVaR). Here we illustrate its R implementation in the `PerformanceAnalytics` package. Again, we use the `edhec` data and calculate the ES (or ETL) at 99%.

```
100*t(ETL(edhec, p = 0.99, method="gaussian"))
```

```
##                                ES
## Convertible Arbitrage   -4.684598
## CTA Global              -6.026907
## Distressed Securities   -4.078670
## Emerging Markets        -9.421637
## Equity Market Neutral   -1.792072
## Event Driven            -4.112443
## Fixed Income Arbitrage  -3.341417
## Global Macro            -3.753911
## Long/Short Equity       -5.114338
## Merger Arbitrage        -2.288289
## Relative Value          -2.834946
## Short Selling          -14.220605
## Funds of Funds          -4.246042
```

We can clearly see the much large numbers in terms of expected loss. Yet again, the overall risk profile remains intact - Short Selling and Equity Markets pose by far the greatest risks, while Equity Market Neutral - the smallest. This method is convenient due to the same interface as the `VaR` command. It is of particular note that it can use the same three methods as the Value at Risk estimation - historical, normal, and modified Cornish-Fisher. We calculate all of them in turn:

```
x <- cbind(100*t(ETL(edhec, p = 0.99, method="historical")),
           100*t(ETL(edhec, p = 0.99, method="gaussian")),
           100*t(ETL(edhec, p = 0.99, method="modified")))
colnames(x) <- c("Historical ES", "Gaussian ES", "Modified ES"); print(x)
```

```
##                          Historical ES Gaussian ES Modified ES
## Convertible Arbitrage          -11.320   -4.684598  -10.092227
## CTA Global                      -5.375   -6.026907   -5.572595
## Distressed Securities           -8.055   -4.078670   -6.533764
## Emerging Markets               -16.265   -9.421637  -13.971949
## Equity Market Neutral           -4.360   -1.792072   -4.404136
## Event Driven                    -7.565   -4.112443   -6.385154
## Fixed Income Arbitrage          -8.340   -3.341417   -5.850228
## Global Macro                    -3.085   -3.753911   -2.437999
## Long/Short Equity               -6.520   -5.114338   -6.957140
## Merger Arbitrage                -4.100   -2.288289   -3.630211
## Relative Value                  -6.150   -2.834946   -5.053100
## Short Selling                  -12.895  -14.220605  -16.974586
## Funds of Funds                  -6.170   -4.246042   -5.500037
```

The ES is also very dependent on its precise methods and assumptions and those should be carefully selected. The outcomes will heavily depend on that - e.g. the Convertible Arbitrage and Emerging Markets funds vary by as much as 6% between the Gaussian method and the Modified method. Such discrepancies may significantly affect the decision-making and potentially skew it. Again, the RM professional should be careful to avoid this and select optimal methods and assumptions that fit data best.

The Global Crisis and the VaR and RM controversy

As the global financial crisis raged in the period 2008-2011, many analysts interpreted the bank failure and financial system problems as clear indication of (among others) mismanaged risk and failure of the traditional methods to manage risk. A hedge fund manager, David Einhorn, captured the general mood by saying that VaR is "an airbag that works all the time, except when you have a car accident."

There may be some truth to that, and we will review the controversy that ensued. At the start of the crisis, there was widespread use of VaR models for risk management, capital adequacy and regulatory purposes. Models were built by utilizing short-term data (one or two decades), which included only relatively mild business cycles. Enthusiasm and risk-taking behavior were rampant and institutions and traders seemed to be deriving comfort from models. The widespread adoption of VaR models means that likely not everyone was appreciating their limitations because of either ignorance or conscious choice.

As the crisis engulfed the financial sector, VaR-based models became largely irrelevant due to data limitations, parameter variance, and overall shift in market dynamics. VaR models

seem to work in situations of low and predictable risk but cannot accommodate high volatility. What is more, calculated risks that the companies were thinking they were taking quickly turned into enormous losses far beyond expectation.

As large systemic institutions were taking unexpectedly large hits, their positions became untenable and losses had to be socialized. Governments stepped in, propping large banks and other financial intermediaries to prevent financial markets collapse. This took place both in the USA, and across Europe. As public money was spent public sensitivity and scrutiny increased, pointing to the failure of traditional RM methods and RM systems. While this holds some truth, few viable alternatives have been devised.

A large reform has taken place in terms of regulators with both increased supervision at the systemic level and reform in key regulations, such as the Basel III standards (2010, revised 2011). Still, VaR metrics remain an integral part of this framework. The RM analyst should therefore use the metric judiciously and amplify its strengths to ensure its positive impact on the risk management process.

Project 6

Imagine that you are operating on the Bulgarian Stock Exchange and are holding a well-diversified portfolio that resembles the market.

Assume that data follows the Normal distribution. Build the distribution of returns based on historical data. What is the mean and standard deviation?

What is the 90% VaR over the next day? What is the 95% and 99% VaR? What are the limitations of the obtained results?

Lecture Seven: Random Variables and Distributions

So far we have been thinking of risk as the probability that some variable of interest (e.g. price, return, profit, loss) will have a realization that is far away from what we expect. This is essentially the question of a realization of a random variable, drawn from a particular distribution. We will put this problem in context by reviewing a number of common distribution that can be useful when modeling data and explaining how their general form can be calibrated to better fit empirical observation.

We look at the normal distribution as a natural benchmark, and then proceed to investigate the exponential one. The lecture finishes with a look at distributions that can emulate the fat tails phenomenon.

Understanding Distributions and Their Role

All the possible realizations of a given random variable can be summarized in its probability density function (PDF) and counterpart - the cumulative density function (CDF). The PDF gives the likelihood of a given value, or alternatively it can give the likelihood that the realization will be above or below a given value. This is of obvious interest to risk management, as it provides a quantitative estimate of the risk that some event (or value) will occur.

For any distribution we can measure this by using Chebyshev's inequality. It states that in any distribution with mean μ and standard deviation of σ, then at least $1 - \frac{1}{k^2}$ of the observations are within k standard deviations from the mean, or:

$$P(|X - \mu| \leq k\sigma) \geq \frac{1}{k^2}$$

An alternative way to think about this is to say that no more than $\frac{1}{k^2}$ observations are k standard deviations from the mean. The RM implications and this gives an indication of the amount of risk taken, regardless of the underlying data distribution.

Using the inequality, we can calculate the probability of a given observation lying k standard deviations away from its expected value, which is done in the following table.

k	Min % within k SD of mean	Max % beyond k SD from mean
1	0%	100%
1.5	55.56%	44.44%
2	75%	25%
3	88.8889%	11.1111%
4	93.75%	6.25%
5	96%	4%
6	97.2222%	2.7778%
7	97.9592%	2.0408%

k	Min % within k SD of mean	Max % beyond k SD from mean
8	98.4375%	1.5625%
9	98.7654%	1.2346%
10	99%	1%

We note that in an unknown distribution there may be no realization to be found in the range of one standard deviation away from the mean, whereas in the Normal distribution 68% of data is found in exactly this interval. Thus, depending on the distribution chosen, the numeric value of the risk taken may differ dramatically. On the practical side this gains in importance as it is often difficult to decide on the precise distribution given a relatively small sample of empirical data. While there are econometric tests of the normality of distribution, the results are not always definitive.

It is not overstated to say that the data distribution is probably the most important single parameter in risk management models. It gives the range of possible outcomes, their likelihood, and the uncertainty surrounding them. This is why the RM professional needs to understand the wide range of modeling possibilities and judiciously choose the most appropriate. There are literally hundreds of different distributions but in practice three large groups are most common:

- *Variants of the normal distribution* - we have explicitly or implicitly used them so far.
- *Power law and exponential distribution* - for positive values only, and with very long tails.
- *Groups of fat-tails distributions* - those that approximate the normal bell curve but give larger probability of extreme events.

We review all of them in turn.

The Normal Distribution

Traditionally, the most widely used distribution is the Normal or the Gaussian one. This kind of distribution describes particularly well realizations of events that are bounded from above and below and tend to cluster around a well-defined middle point. Such data can be natural regularities or human-made ones. Examples include height, weight, intelligence (IQ), etc. While there are some people with exceptional e.g. height, there are few extremely tall or short individuals, and most people will be of average height. Furthermore, there are both upper and lower bounds, as no one can have negative height or be 5 meters tall. While the Gaussian distribution itself is not strictly bounded, extreme cases are very unlikely.

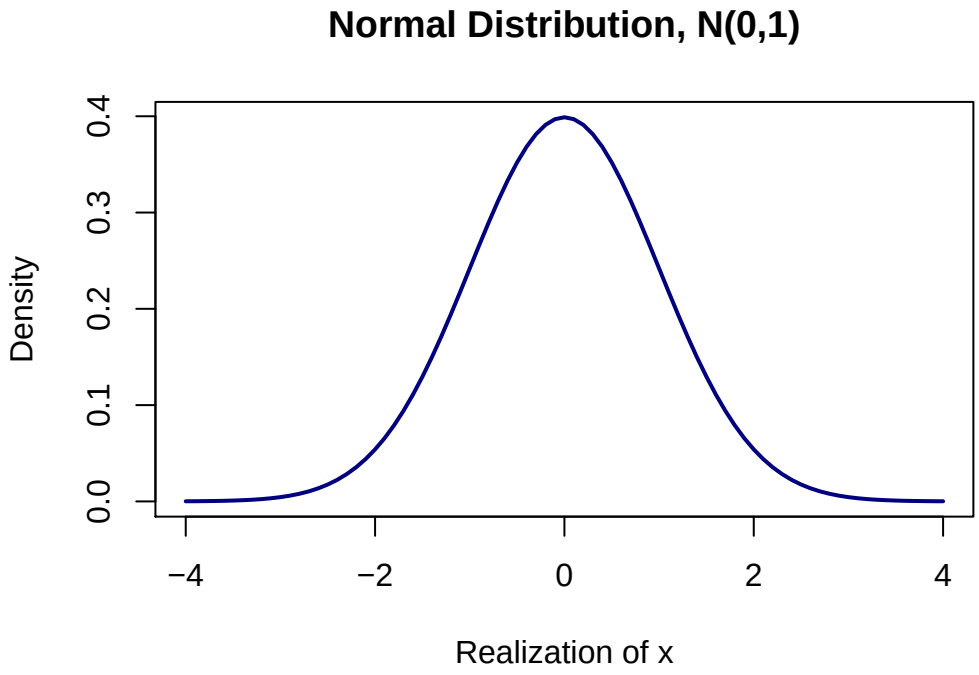
Apart from such obvious instances, the normal distribution is used to model a lot of phenomena, whose true data distribution is not exactly but approximately normal, or it is unknown. This is largely due to the fact that the Central Limit Theorem posits that under certain conditions the means of a large enough pool of random variable iterates will converge to a normal distribution.

A lot of research in financial markets has found that returns do tend to be approximately normally distributed, which is due to their mean-reversion properties, as well. The probability density function of the normal distribution is thus defined to be:

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Such a function produces the well-known bell-shape characteristic of Gaussian random variables. Here we can see how most observations cluster around the mean. The further away values are, the more unlikely they are to occur.

```
x <- seq(from = -4, to = 4, by = 0.1)
plot(x, dnorm(x, 0,1), type="l", col="navyblue", ylab="Density",
     xlab="Realization of x", main = "Normal Distribution, N(0,1)",
     lwd=2)
```



For this distribution a lot of observations are close to the mean, and almost all fall within up to three standard deviations from it. Extreme events are thus very unlikely.

k	Min % within k SD of mean	Max % beyond k SD from mean
1	68.2%	31.8%
2	95.5%%	4.5%
3	99.7%%	0.3%

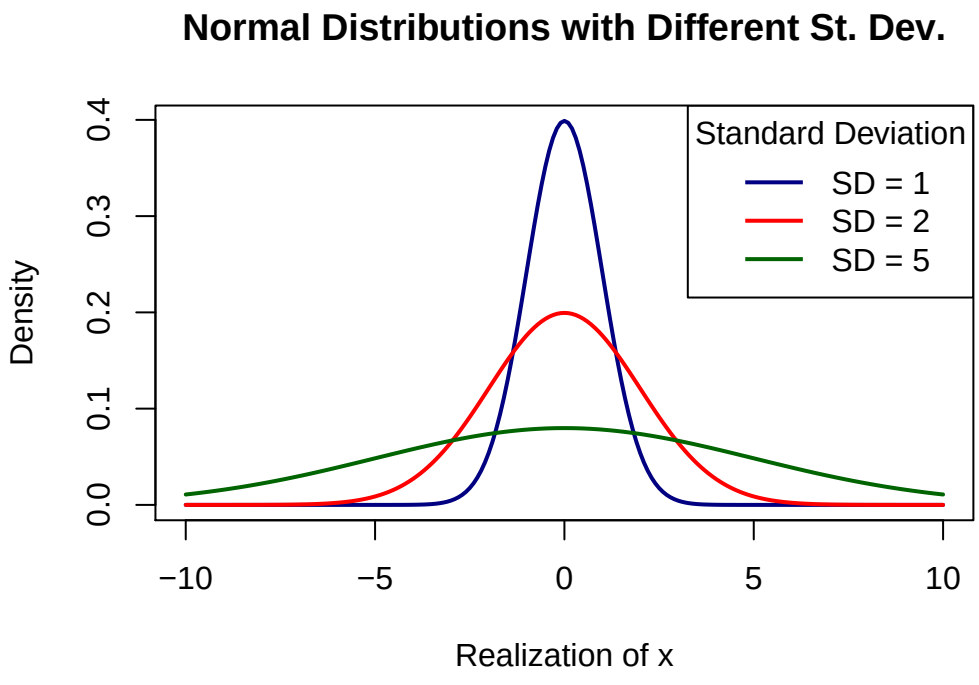
It is from those properties of the normal distribution that the term six sigma comes from. A six sigma (6σ) event will be equal in probability to 3.4 in a million. In production, the six sigma quality control prescribes that only 3.4 items in a million could have a defect. Overall, the term implies a very rare event, unlikely to occur.

The key parameter of the Normal distribution is its standard deviation (the mean merely translates horizontally the bell). The larger the standard deviation, the more dispersed data is. This can be graphically seen in the following figure that plots three normal distributions with means of 0, but with standard deviations of 1, 2, and 5, respectively.

```
x <- seq(from = -10, to = 10, by = 0.1)
labels = c("SD = 1", "SD = 2", "SD = 5")
colors = c("navyblue", "red", "darkgreen")
plot(x, dnorm(x, 0,1), type="l", col="navyblue", ylab="Density",
     xlab="Realization of x",
     main = "Normal Distributions with Different St. Dev.",
     lwd=2) + lines(x, dnorm(x, 0,2), col="red", lwd=2) + lines(x,
     dnorm(x, 0,5), col="darkgreen", lwd=2)
```

```
## integer(0)
```

```
legend("topright", title="Standard Deviation",
       labels, lwd=2, col=colors)
```



As usual, larger standard deviations make up for larger risks and should be properly accounted by the RM professional. A way to understand the effect of the SD is to

interactively change it. The following code snippet will produce such an interactive graph in R.

```
library(manipulate)
manipulate(
  plot(seq(-10,10,.01), dnorm(seq(-10,10,.01), 0,x),
       type="l", col="navyblue", lwd=2, ylab="Density",
       xlab="Value of Random Variable"),x = slider(0.5,5))
```

The Power Law and Exponential Distributions

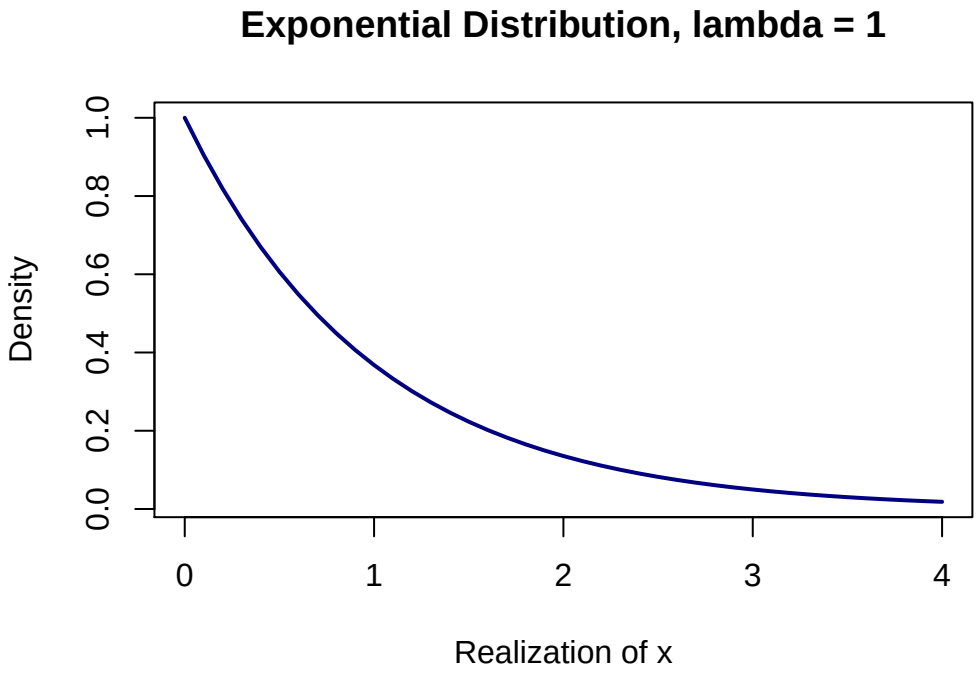
Some empirical phenomena cannot be described well by a normal distribution. There are cases in which a given non-extreme event (value) appears very often but there also there are some truly extreme events (values) that do take place (so called "long tails"). For instance, such distribution is characteristic for wealth, popularity, web site visits, etc. In the case of wealth, this means that a lot of individuals have relatively little assets, while very few individuals are extremely rich. In the case of internet sites, this means that a lot of sites have very few (if any) visitors, while a few generate most of the traffic (like Google, Facebook, etc.) This empirical regularity is referred to a power law as it can be approximated by power functions.

An easy way to describe such data would be to use the exponential distribution, a special type of the Poisson distribution. If we denote with λ the rate at which a given even occurs, then the exponential distribution is defined as follows for $x \geq 0$:

$$f(x|\lambda) = \lambda * e^{-\lambda x}$$

The graph of the distribution for $\lambda = 1$ is presented in the following plot.

```
x <- seq(from = 0, to = 4, by = 0.1)
plot(x, dexp(x, 1), type="l", col="navyblue", ylab="Density",
     xlab="Realization of x", main = "Exponential Distribution, lambda = 1",
     lwd=2)
```



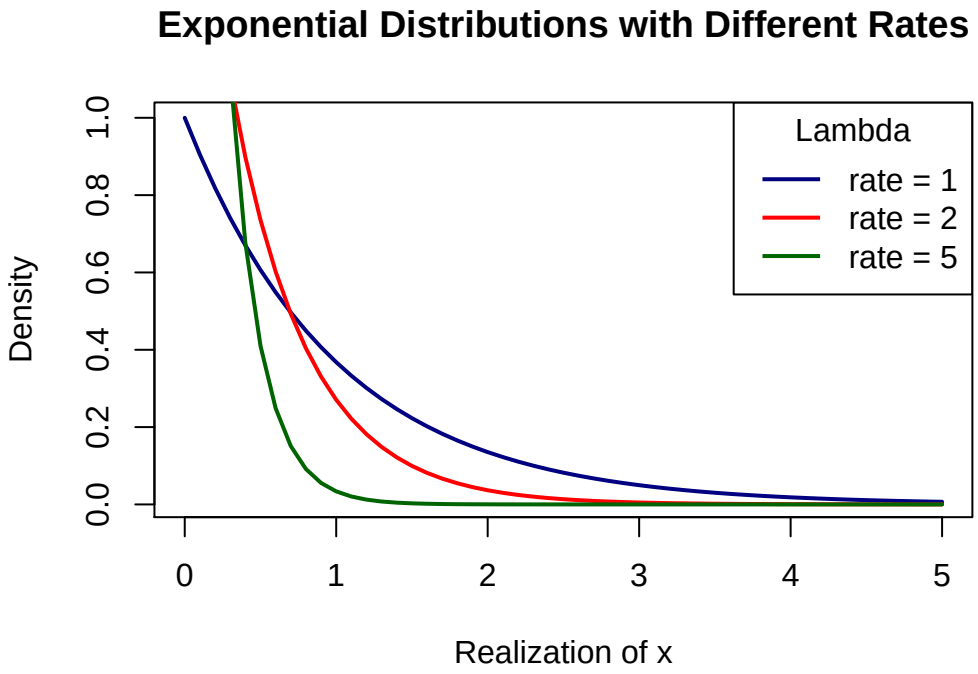
Such a distribution can capture very well the process of inequality. Pareto was referring to such a type of distribution when he posited his 20-80 principle - that 20% of the people in Italy hold 80% of the land in the country.

The key parameter in this distribution is the rate at which a given event occurs - λ. The more the rate increases, the more skewed to the left the distribution becomes, thus signalling greater inequality. This is illustrated in the following plot with the rate set to 1, 2, and 5.

```
x <- seq(from = 0, to = 5, by = 0.1)
labels = c("rate = 1", "rate = 2", "rate = 5")
colors = c("navyblue", "red", "darkgreen")
plot(x, dexp(x, 1), type="l", col="navyblue", ylab="Density",
     xlab="Realization of x",
     main = "Exponential Distributions with Different Rates", lwd=2) +
     lines(x, dexp(x, 2), col="red", lwd=2) + lines(x, dexp(x, 5),
     col="darkgreen", lwd=2)
```

```
## integer(0)
```

```
legend("topright", title="Lambda",
       labels, lwd=2, col=colors)
```



One can better understand the change of the exponential distribution by interactively changing the rate. The following code snippet constructs such a tool in R.

```
library(manipulate)
manipulate(
  plot(seq(-10,10,.01), dt(seq(-10,10,.01), df=x),
       type="l", col="navyblue", lwd=2, ylab="Density",
       xlab="Value of Random Variable"),x = slider(0.5,5))
```

Distributions with Fat Tails

Despite its ubiquity, the Normal distribution features a major flaw with important implications for risk management. It understates the importance of unlikely events with potentially large impact. While data supports that the overall form of financial market returns approximates the Gaussian distribution, it seems that extreme events have a higher realized probability than they would under it. This phenomenon is known as *fat tails*. Thus there is benefit to model some returns data not as normally distributed, but with some alternative that takes into account this phenomenon.

One of the popular distributions that can mimic this behavior is the Cauchy one. Very popular in the natural sciences, it is a stable distribution that has a well-defined analytic form. If we denote with x_0 the peak of the distribution and with γ - a scale parameter, its probability density function can be expressed as follows:

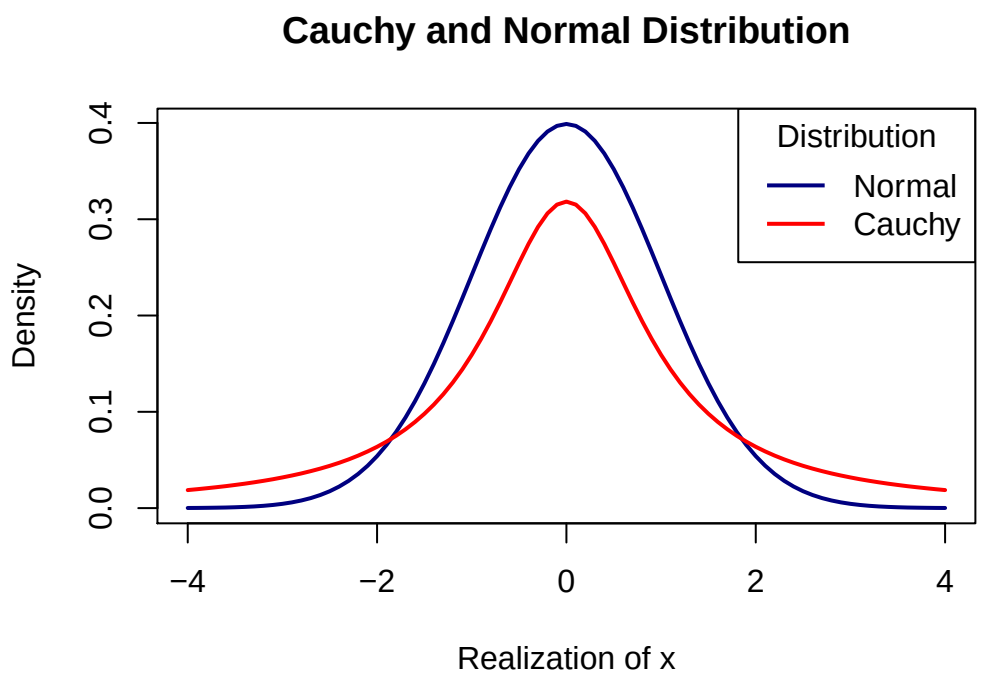
$$f(x|x_0,\gamma) = \frac{1}{\pi\gamma}[\frac{\gamma^2}{(x-x_0)^2+\gamma^2}]$$

We can compare a Cauchy distribution with location 0, and scale of 1 with a normal distribution with mean of 0 and s.d. of 1, presented on the following plot. The tails of the Cauchy distributions are much thicker, thus giving a larger probability to outliers, and less probability of realizations near its mean. While the Cauchy distribution resembles the normal one in shape, it can provide very different numeric estimates for risk, especially for rare events.

```
x <- seq(from = -4, to = 4, by = 0.1)
labels = c("Normal", "Cauchy")
colors = c("navyblue", "red")
plot(x, dnorm(x, 0, 1), type="l", col="navyblue", ylab="Density",
     xlab="Realization of x", main = "Cauchy and Normal Distribution",
     lwd=2) + lines(x, dcauchy(x, location=0, scale=1), col="red", lwd=2)
```

```
## integer(0)
```

```
legend("topright", title="Distribution", labels, lwd=2, col=colors)
```



Such a behavior is naturally useful from a risk management perspective. It is often the case that the largest risk to an organization lies not around expectation but far from it, and the outliers need to be accounted more carefully for. Since those events could possibly have a catastrophic impact they should not be underestimated.

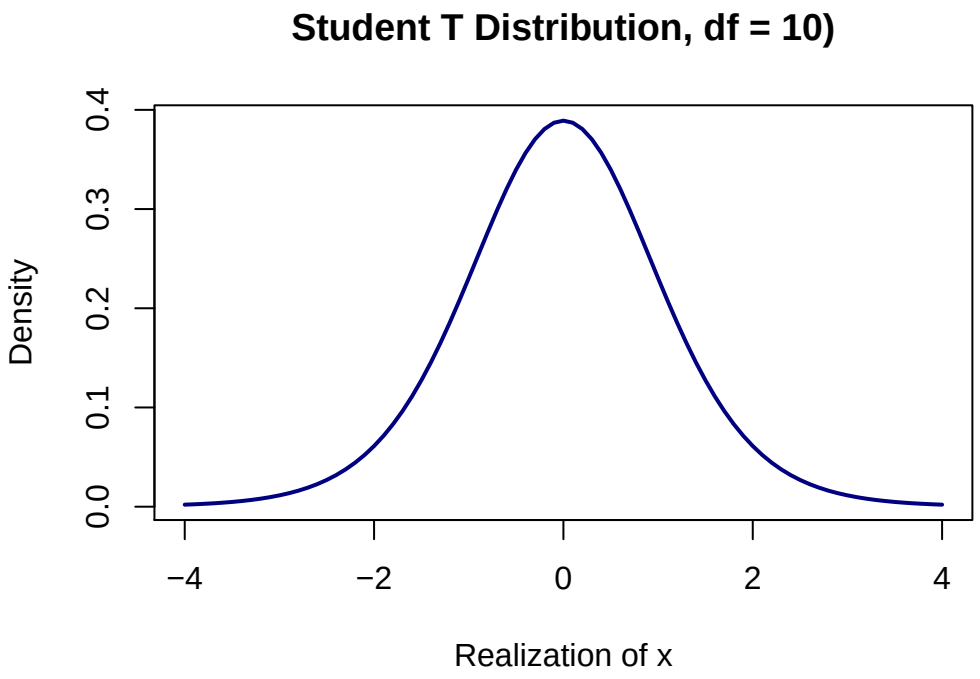
Here we can also make an argument from a management perspective. Even if the true distribution of data were normal, and the RM manager is using a fat tails one, this means that he somewhat overstates the risk. In this way the RM process provides a buffer against the uncertainty of calculation and leads to lower risk within the organization. In some particularly sensitive operations (e.g. financial regulations, or nuclear power plant operations) this may be desirable.

Another distribution that is in wide use for RM perspective is the Student T distribution. As we will see, it has a lot of desirable properties. To define it, we use the Gamma function, Γ, where $\Gamma(n) = (n-1)!$. We also denote the degrees of freedom of the distribution with ν, where $\nu = n - 1$. The probability density function is thus:

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1 + \frac{x^2}{\nu})^{-\frac{\nu+1}{2}}$$

Visually, the Student's T distribution resembles the normal one with its typical bell curve. Such a plot is presented in the following figure, with a centered T distribution with 10 degrees of freedom.

```
x <- seq(from = -4, to = 4, by = 0.1)
plot(x, dt(x, df=10), type="l", col="navyblue", ylab="Density",
     xlab="Realization of x", main = "Student T Distribution, df = 10)",
     lwd=2)
```



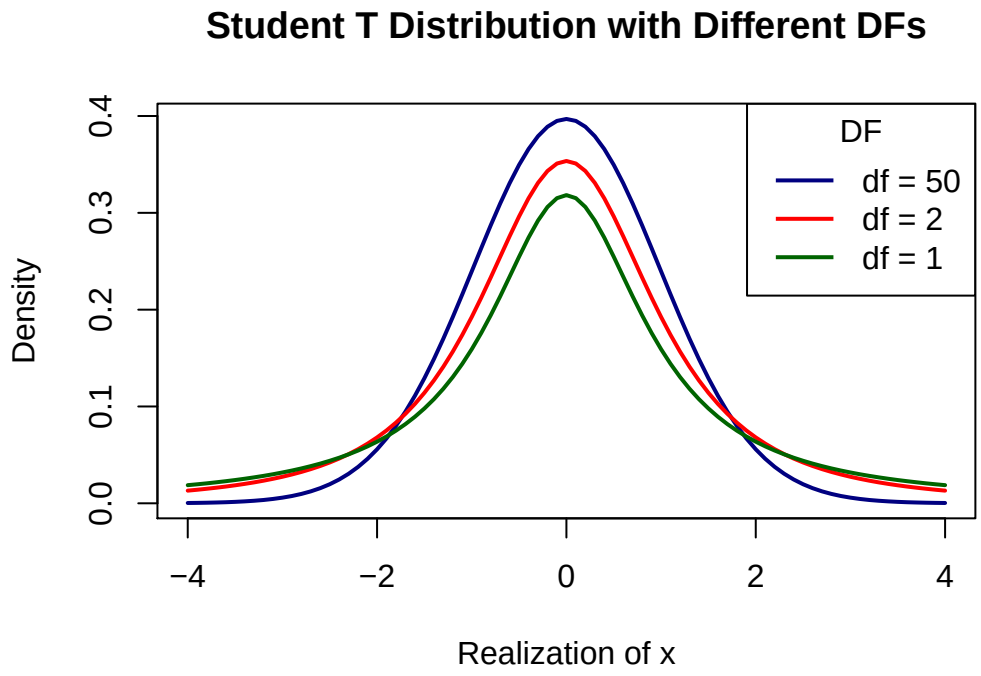
We can see how the distribution changes as the degrees of freedom change in the following

plot for $\nu = 1, 2, 50$.

```
x <- seq(from = -4, to = 4, by = 0.1)
labels = c("df = 50", "df = 2", "df = 1")
colors = c("navyblue", "red", "darkgreen")
plot(x, dt(x, df=50), type="l", col="navyblue", ylab="Density",
     xlab="Realization of x",
     main = "Student T Distribution with Different DFs", lwd=2) +
     lines(x, dt(x, df=2), col="red", lwd=2) + lines(x,
     dt(x, df=1), col="darkgreen", lwd=2)
```

```
## integer(0)
```

```
legend("topright", title="DF",
       labels, lwd=2, col=colors)
```



An especially useful property of the student T-distribution is that by changing its key parameter - the degrees of freedom, it mimics the behavior of other ones. In particular:

- A Student's T distribution with $\nu = 1$ is the same as the Cauchy distribution.
- A Student's T distribution with ν approaching infinity is the same as the Normal distribution. Above $\nu > 30$ it very closely approximates it.
- For values of $1 < \nu < 30$ in between, distributions tend to have fatter tails than the Gaussian one, thus making it useful for some classes of data.

Naturally, if the analyst believes strongly in the high probability of extreme events, she can

set the degrees of freedom to a fractional number that best fits data and information.

Deciding on Distributions

We have seen how parametric and simulation modeling strongly depends on the assumption of the data distribution and its parametrization. Naturally, even if perfectly correct those distributions are merely approximations to the true data-generating process in real life. Thus, they are at best a rough guide to reality. On the other hand, they give a sense of mathematical precision that can hardly be justified under careful scrutiny.

This has led some observers such as N. N. Taleb to note that modeling encourages needless risk-taking by providing erroneous numbers, and that it should either be improved significantly before it becomes useful, or abandoned altogether. Such criticism has widely been applied not only to models in finance, but also to models of economic decisions, and of the macroeconomy. While models could use some improvement, there is hardly any viable alternative to them. We will postpone the discussion of their problems to Lecture Twelve but will instead focus on ways to better decide on data distributions and their parameters in order to have more accurate view of data, and so - better estimates of risk.

The first key question is what distribution most closely resembles the data at hand. A natural benchmark would be to assume that the distribution is normal and test this hypothesis against data. There are many tests for normality among which the analyst can choose. The Jarque-Bera one is a well-known one that is commonly implemented in many statistical and econometric packages. We demonstrate it on the EDHEC data in R.

```
library(tseries); library(PerformanceAnalytics)
data(edhec)
jarque.bera.test(edhec$"CTA Global")
```

```
##
##  Jarque Bera Test
##
## data:  edhec$"CTA Global"
## X-squared = 0.53946, df = 2, p-value = 0.7636
```

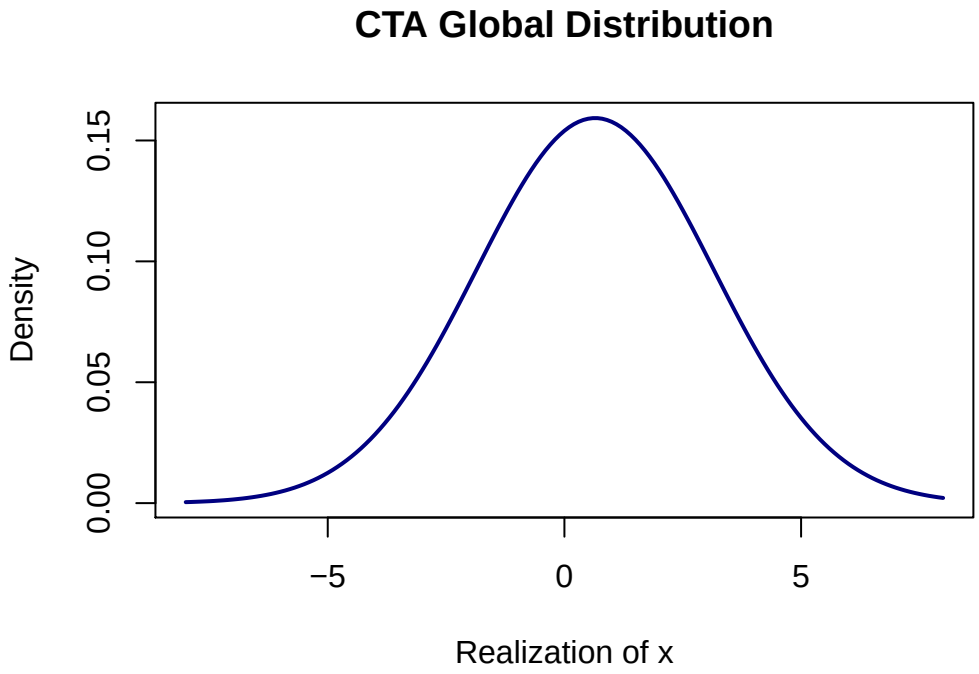
We test each distinct time series against the null hypothesis of normality. In the case of CTA Global we cannot reject the null, since the p-value stands at 0.76. Therefore we retain our hypothesis of Gaussian data. The next step is to fit the parameters of the distribution, and for that purpose the command `fitdistr` from the R `MASS` package can be used.

```
library(MASS)
fitdistr(edhec$"CTA Global", "normal")
```

```
##       mean           sd
##   0.006489474   0.025048096
##  (0.002031669) (0.001436607)
```

It seems that the best fitting distribution would be with a mean in percent of $\mu = 0.64$, and a standard deviation of $\sigma = 2.50$. We plot this distribution to see what is the optimal approximation to the dynamics of the CTA Global returns, based on the data we collected. We should note that data needs to be as representative as possible in order to estimate a distribution which is close to the real one.

```
x <- seq(from = -8, to = 8, by = 0.1)
plot(x, dnorm(x, 0.6489474,2.5048096), type="l", col="navyblue",
     ylab="Density", xlab="Realization of x",
     main = "CTA Global Distribution", lwd=2)
```



After the estimation phase is complete, the RM professional can use the returns distribution for his purposes, to calculate risks, to predict or simulate. However, the limitations of this data should also be kept in mind. While it is econometrically relatively easy to fit distribution parameters and there are many tests for specific types of distributions we need to be wary that no amount of calculation can substitute common sense and use the number judiciously.

Project 7

Find data on at least six real-world phenomena. Provide descriptive statistics (including histogram) for them.

- What distributions can best approximate them?

- Find the distribution parameters.

Feel free to use either the `fitdistr` command or any other utility of your choice.

Lecture Eight: Monte Carlo Methods for Risk Management

A Short Note on Monte Carlo Methods

A wide range of statistical methods can be used to run comprehensive computer-aided simulations for complex problems. By using this approach the analyst simulates every conceivable state of the world, and given the resulting numbers, can draw conclusions for the likelihood of events of interest. In this chapter we look at some RM examples that use this approach and show how this is sometimes the only feasible way to do pricing and analyze risk in a rigorous manner.

Risk, as we have seen, can be modeled using the probability distributions of different events or outcomes. It is often the case that modeling only one event is insufficient as the final risk-weighted decision depends on a chain of events with different probability of realization. For example, if the question is what is the expected portfolio return at the end of the quarter, this can be answered by modeling the whole set of probabilities of each stock. In the simple case of two stock there are four distinct cases - both go up, both go down, and the two instances of reverse movements. As we quantify all the possible returns and their respective chances, the problem quickly becomes very complex.

With the addition of more and more uncertain variables, the number of possible outcomes grows exponentially (as probabilities are multiplied) and the analytic solution becomes practically intractable. A possible approach to modeling such cases is to run simulations of **almost all** possible outcome paths and look at aggregate statistics. This will give both the expected values as well as their ranges (standard deviations) for the purposes of managing risk. The class of such simulations is known as Monte Carlo Methods. For a more detailed introduction, the reader is referred to Rober & Casella (2009).

The usual way to approach a given problem using Monte Carlo methods is as follows:

- Define the decision problem and the key variables
- Quantify the relationships between the variables (generally in mathematical form)
- Define the key variables of interest that are not deterministic
- Model those key variables - what are their conceivable outcomes, and how likely they are
- Simulate the whole model by randomly drawing the values for the key variables from their respective distributions
- Iterate numerous times so that the results from the drawn samples converge to the whole population
- Calculate the statistics and distribution of outcomes and draw conclusions for RM purposes

By doing this the analyst develops confidence of the possibilities but this comes at the price that potential users may conceive this exercise as a statistically-driven black box. It is therefore of urgent importance that during the model-building phase experts with domain

knowledge be involved - both to ensure better model quality and improve the buy-in of results.

A Simple Monte Carlo Simulation

To illustrate the principles of Monte Carlo Methods (MCM), we can simulate a simple business case. Assume a company that needs to calculate its expected profit. The profit is therefore:

```
profit <- revenue - cost
```

The firm has already signed contracts for 20 Million which it expects to fulfill them, and also has additional sales it will realize over the year. In addition that that there might be unexpected external shocks (revshock), thus we reach:

```
revenue <- 20 + sales + revshock
```

We similarly model cost as the sum of fixed cost, variable cost, and a shock:

```
cost <- fcost + varcost + costshock
```

If the fixed cost is 25 Million, and the variable one depends on the number of produced units (where this is equal to total sales over their price of 1) multiplied by the cost of producing one unit (0.8), we reach:

```
p <- 1
cost <- 25 + 0.8*(sales/p) + costshock
```

Thus this relatively simple decision problem depends on three key variables - the expected sales, and the unexpected shocks to revenue. In a relatively stable industry with constant sales, we can model them using past data. Suppose the past sales do follows a Normal distribution with a mean of 40, and a standard deviation of 8. Note that in a rapidly growing industry the firm will model its expectations on a very different premise (e.g. a power distribution).

Thus we can model sales as:

```
sales <- rnorm(1, mean = 40, sd = 8)
```

The random shocks can be also be modeled using historical data or expert judgment. Suppose they are uniformly distributed from -3 to +5 Million, thus being a bit positively skewed.

```
revshock <- runif(1, min = -3, max = 5)
costshock <- runif(1, min = -3, max = 5)
```

We can put all this together and thus get the values for one possible scenario. For illustrative purposes we calculate the profit margin:

```
p <- 1
sales <- rnorm(1, mean = 40, sd = 8)
revshock <- runif(1, min = -3, max = 5)
costshock <- runif(1, min = -3, max = 5)
revenue <- 20 + sales + revshock
cost <- 25 + 0.8*(sales/p) + costshock
profit <- revenue - cost; print(paste0("Margin is ",100*profit/sales," %."))
```

```
## [1] "Margin is 17.5812909995863 %."
```

In this case the analyst only sees one realization of all the probable outcomes. It may be a particularly favorable or unfavorable one, and thus can skew decision-making. This is why it is often the case that such simulations are run a lot of iterations. This gives much more information and is not very computationally expensive for problems of smaller scale. We thus put this whole scenario into a loop.

First we create the empty object *profits* and then iterate for every i in the program 10,000 times. In this way we simulate 10,000 possible realizations for the scenario.

```
margins <- rep(0, 10000)
for (i in 1:10000) {
p <- 1
sales_i <- rnorm(1, mean = 40, sd = 8)
revshock_i <- runif(1, min = -3, max = 5)
costshock_i <- runif(1, min = -3, max = 5)
revenue_i <- 20 + sales_i + revshock_i
cost_i <- 25 + 0.8*(sales_i/p) + costshock_i
profit_i <- revenue_i - cost_i
margins[i] <- 100*(profit_i/sales_i)
  }
```

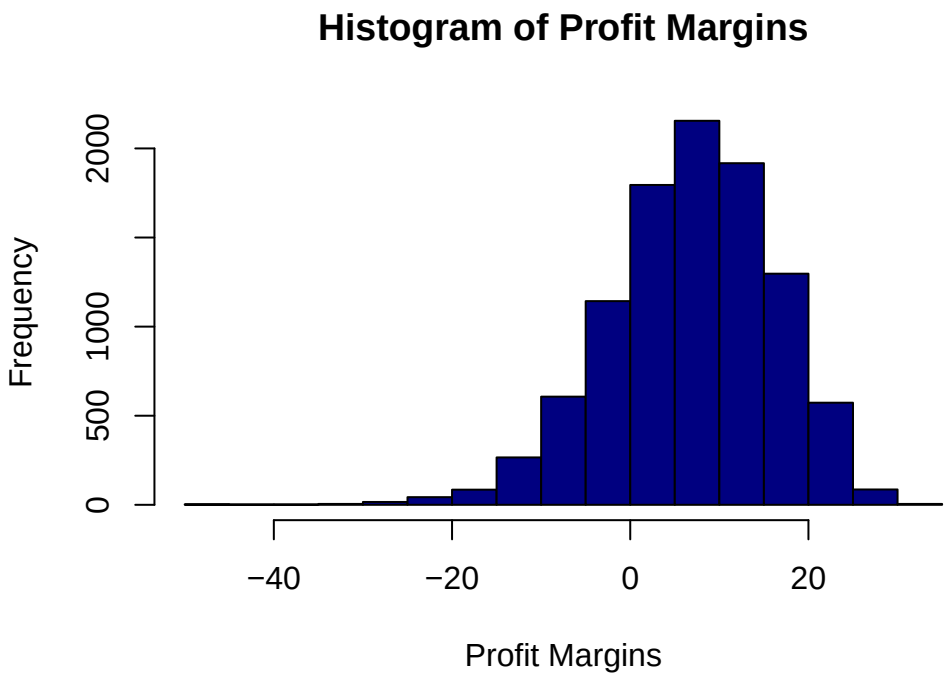
Overall statistics for all the simulated scenarios can be derived from the resulting object using familiar functions as per preference.

```
library(psych)
describe(margins,skew = F)
```

```
##    vars     n mean   sd    min   max range   se
## X1    1 10000 6.85 9.27 -48.85 33.91 82.76 0.09
```

A particularly useful tool for MCM analyses is the histogram. It present the results for the variable of interest in an easy to grasp and intuitive way. This is shown in the following graph.

```
hist(margins, xlab="Profit Margins", main="Histogram of Profit Margins",
     col="navyblue")
```



Obviously, with every re-run of the analysis, the results will be different, as new values for the key variables are randomly drawn. However, if the iterations are enough in number (in this case 10,000), the results will differ only slightly as the sample size provides for convergence. This is also one of the key metrics for the quality of a MCM model - relative stability during re-calculation. In terms of the data we have in this example, it seems that the firms risks are skewed to the upside and it should mostly expect positive margins between 0 and 10%, but there is also a non-trivial probability for a small negative margin over the next year.

Applying Monte Carlo Methods to Stock Returns

We can also use MCM to evaluate the returns of a given stock, portfolio, or a set of positions. This gives us the flexibility to simulate many runs of the eventualities, related to stock behavior and derive a concrete probability distribution for returns (or any other variable of interest). We can use the assumption that stock returns follow a Brownian motion type of dynamics (e.g. Pelsser, 2000) and model their returns r as function of their long-run average μ, their risk (or standard deviation) σ, and time t. The path of returns is then defined as follows:

$$r = \mu\Delta t + \sigma Z + \sqrt{\Delta t}$$

In this case the mean μ is the drift and σ is then a marker of volatility. The variable Z is a standardized normally distributed random variable (with a mean of 0, and a standard deviation of 1). Assume that we are dealing with a stock with a mean monthly return of 5%,

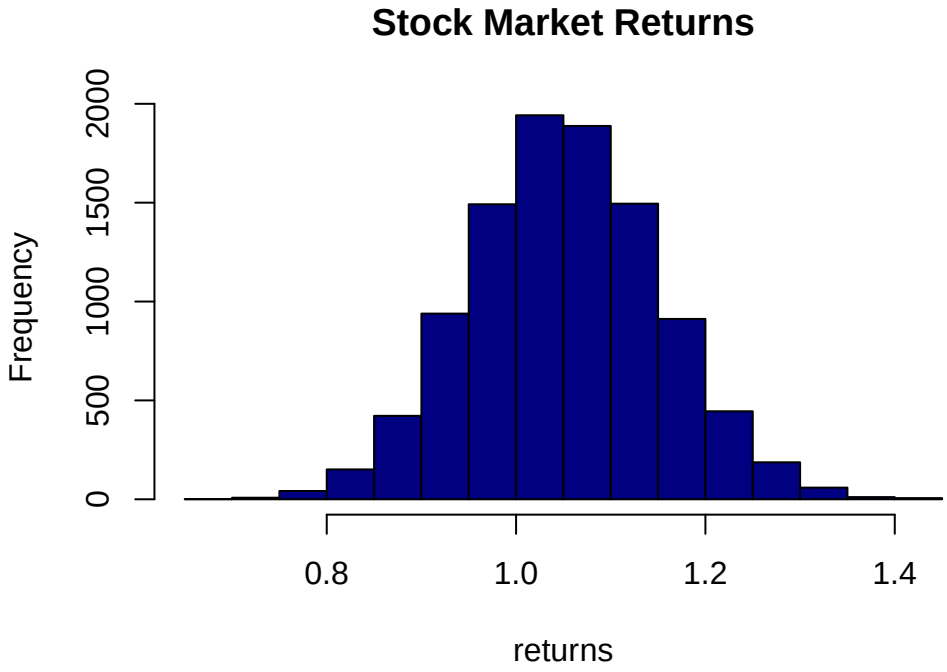
a standard deviation of 10%. If dealing with annual data, then $\Delta t = 1$, for quarterly data it is 1/4 (or 0.25), and 1/12 for monthly data.

To provide for the same randomly generated numbers every time, we can use the `set.seed()` command so that results do not change from simulation to simulation. We now generate 10,000 iterations of this equation, using the following code:

```
set.seed(456)
n <- 10000
z <- rnorm(n)
mu <- 0.05
sd <- 0.10
delta_t <- 1
returns <- mu*delta_t + sd*z + sqrt(delta_t)
```

To obtain an impression of the stock market dynamics in aggregate we can plot the histogram of the results.

```
hist(returns, main="Stock Market Returns", col="navyblue")
```



Using this simulation we can derive implications for the possible risks of holding this stock and obtain of quantitative estimate of its expected behavior. As usual, this can be summarized in its descriptive statistics:

```
library(psych)
describe(returns)
```

```
##    vars     n mean  sd median trimmed mad  min  max range skew kurtosis se
## X1    1 10000 1.05 0.1   1.05    1.05 0.1 0.69 1.43  0.74 0.06    -0.01  0
```

The simulation generated a distribution with exactly the same aggregate statistics - mean of 5% and a standard deviation of 10%, thus showing that results are consistent across the simulated scenario. If this was not the case - i.e. if we were aiming to simulate the behavior of a stock with a mean return of 5% but then the simulation yielded a different number, this means that results are not robust and there may be some model errors.

Monte Carlo Pricing of Options

Options are particular financial instruments that give the investor the opportunity but not the obligation to purchase or sell a given stock for a given price at a certain point in the future. Such an instrument is particularly useful for risk management purposes as it effectively allows the investor to hedge a given risk by purchasing a set of options. Only the simplest options can be priced using analytic formulas, and more complex ones need to be priced using simulations of their price paths and returns. Thus option pricing is a major application for MCM methods in finance. As an example here we will look at so-called European options (those can be exercised only at a specified expiration date), but the major conclusions extend beyond that.

- **The European call option** - it gives the owner to right (but not the obligation) to buy a given asset at a pre-specified date for a pre-specified price.
- **The European put option** - it gives the owner to right (but not the obligation) to sell a given asset at a pre-specified date for a pre-specified price.

The exercise of this right is called "strike" (denoted K), and the value of the underlying asset at given time t is denoted $S(t)$. For rational investors, the profit from striking an option is as follows.

For call options:

$$\pi = max[(S(t) - K), 0]$$

For put options:

$$\pi = max[K - (S(t)), 0]$$

Options pricing is particularly challenging and sometimes possible only numerically but not analytically. The Black-Scholes formula gives a general differential equation that show the dynamics of options, given their strike prices K, spot prices of the underlying asset $S(t)$, the time period t, the price of the option $V(S, t)$ (a put or a call), the riskless rate r, and the asset volatility σ. The Black-Scholes equations is then as follows:

$$\frac{\partial V(S,t)}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V(S,t)}{\partial S^2} + rS\frac{\partial V(S,t)}{\partial S} - rV = 0$$

A key conclusion from this equations is that is posits the possibility for a perfect riskless hedge by buying and selling the underlying asset. Therefore it also implies a single unique option price. By solving the above equation given the parameters of the European options we can obtain formulas for their pricing.

However, it is only a small subset of options that can be priced this way, which is why Monte Carlo methods are so useful. We will illustrate how they can be leveraged to price options, along with the respective pricing, implied by the above equation. Using the Black-Scholes equation, we can reach analytic formulations of the price of a call option, $C(S,t)$. Denoting the cumulative distribution function (cdf) of the normal distribution as $N()$ and the strike time as T, it is as follows:

$$C(S,t) = N(d_1)S_t - N(d_2)K^{-r(T-t)}$$

Where:

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}[ln\frac{S_t}{K} + (r + \frac{\sigma^2}{2})(T-t)]$$

and:

$$d_2 = d_1 - \sigma\sqrt{T-t}$$

The price of the equivalent put option, $P(S,t)$ can then be defined by the following equation:

$$P(S,t) = K^{-r(T-t)} - S_t + C(S,t)$$

Which then simplifies to:

$$P(S,t) = N(-d_2)K^{-r(T-t)} - N(-d_1)S_t$$

Since all the variables in those equations are known, we can easily calculate the price of the option. In this simple case the analyst can do that both analytically and through simulation. As a concrete example, assume that we are dealing with an underlying asset that currently sells at 140 with a calculated standard deviation of 20%, and the option allows us to buy it at a price of 100 in 3 years time. The riskless return at this point stands at 2%. Those numbers are stored in their respective variables.

```
stock <- 140
sigma <- 0.2
strike <- 100
TTM <- 3
rf <- 0.02
```

We then calculate the ratios for d_1 and d_2 in R:

```
d1<-(log(stock/strike)+(rf+0.5*sigma^2)*TTM)/(sigma*sqrt(TTM))
d2<-d1-(sigma*sqrt(TTM))
```

Using the Black-Scholes formula we obtain values for the call option:

```
BS.call<-stock*pnorm(d1,mean=0,sd=1)-strike*exp(-rf*TTM)*pnorm(d2,mean=0,sd=1)
BS.call
```

```
## [1] 48.29649
```

The price for obtaining a call option under those conditions stands at 48.296. The equivalent can also be done for the put option, thus reaching:

```
BS.put<-BS.call-stock+strike*exp(-rf*TTM)
BS.put
```

```
## [1] 2.472939
```

The put option costs 2.473. Those calculations are deterministic and represent the unique optimal price in a frictionless world (or at least in financial markets bordering on efficiency). An alternative way to reach valuation would be to simulate the return dynamics of the option and then discount that to the present so that the discounted cash flow gives the option price. We do that leveraging Monte Carlo methods. Suppose we decide to run 10,000 simulations.

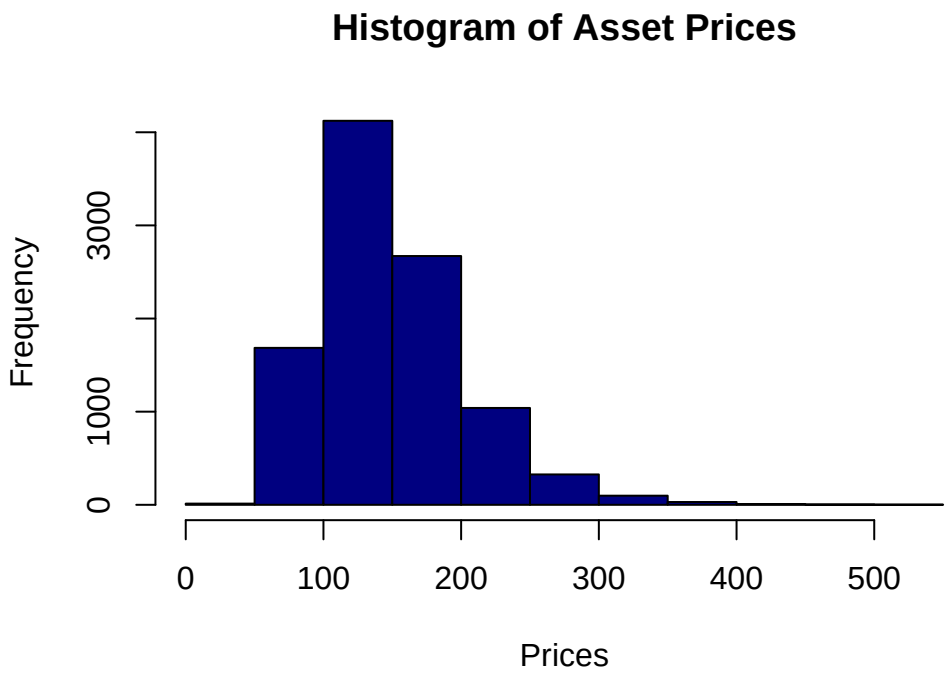
```
n <- 10000
```

We define the dynamics of the return and standard deviation and then simulate the price over 10,000 iterations.

```
set.seed(123)
R <- (rf-0.5*sigma^2)*TTM
SD <- sigma*sqrt(TTM)
TTM.price <- stock*exp(R+SD*rnorm(n,0,1))
```

The histogram again gives an idea of the expected distribution of prices of the underlying asset after three years.

```
hist(TTM.price, col="navyblue", xlab="Prices",
     main="Histogram of Asset Prices")
```



With an overwhelming chance we expect the price of the underlying asset to fall between 100 and 200 and be skewed significantly towards 100. More precise estimates can be obtained by looking at the respective descriptive.

```
library(psych)
describe(TTM.price, skew = F)
```

```
##    vars     n   mean    sd   min    max  range   se
## X1    1 10000 148.52 53.08 36.95 530.89 493.94 0.53
```

In the case of a put option, the investor clearly exercises it only when the realized price in the market is above the price that she will buy the asset. If the option gives the right to buy this stock at 100, this will only be exercises if the spot price is above that, so that profit can be realized.

This is the intuition behind the profit condition for call options. The proportion of these cases is also easy to calculate.

```
sum(TTM.price > 100) / length(TTM.price)
```

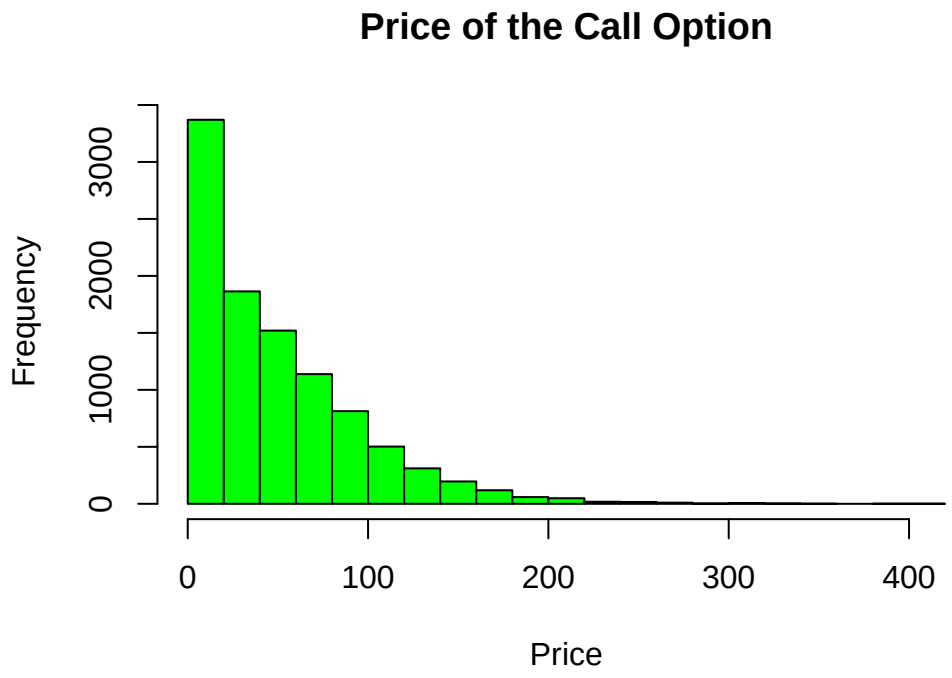
```
## [1] 0.8302
```

This option will then be exercised in 83% of the cases, and left aside in 17%. We enter this condition in the program and then discount to present value only the cases in which this option in actually exercised.

```
TTM.call<-pmax(0,TTM.price-strike)
PV.call<-TTM.call*(exp(-rf*TTM))
```

With this, we can investigate the possible prices of the call option in all our 10,000 simulated cases.

```
hist(PV.call, col="green", xlab="Price", main="Price of the Call Option")
```



The price of the option should then be the expectation (mean) of all realized prices, which turns out to be 48.179.

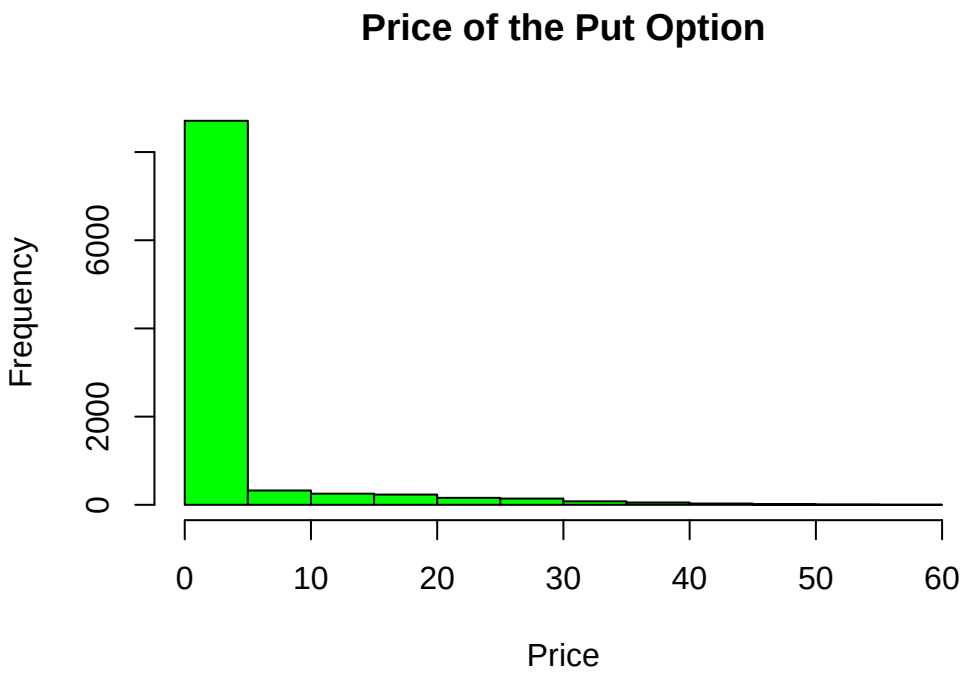
```
mean(PV.call)
```

```
## [1] 48.17931
```

We should note how close is the analytically-derived price (48.296) to the one obtained in the simulation (48.179).

In a similar fashion, one can calculate the price of the put option and construct its distribution for risk management purposes.

```
TTM.put<-pmax(0,strike-TTM.price)
PV.put<-TTM.put*(exp(-rf*TTM))
hist(PV.put, col="green", xlab="Price", main="Price of the Put Option")
```



The price a risk-neutral investor is willing to pay would therefore be the mathematical expectation (mean) of this price.

```
mean(PV.put)
```

```
## [1] 2.485097
```

This turns out to be equal to 2.485, which is again very close to the number, obtained by using the Black-Scholes formula (2.473). This is largely due to the Law of Large Numbers and the effects of the Central Limit Theorem. MCM simulations prove to be versatile and robust alternatives to analytically-derived solutions. What is more, in more complex cases numeric simulations tend to be the only option.

Conclusion

Monte Carlo methods encompass a comprehensive set of principles and practices that can account more fully for the probabilistic nature of the world. They focus on simulating thousands and even millions of iterations of one and the same process in order to understand the distribution of outcomes, as well as the expected values.

An analytic solution to a given problem produces merely a point estimate (in the case of Black-Scholes option pricing - the price of the option), but a simulation also provides a plethora of other information - the conceivable ranges of the price realization, the level of risks (standard deviation), the skewness of those realizations. What is more, there are many complex problems that simply cannot be solved analytically (as are many differential

equations and systems of differential equations), and thus MCM tend to be the only tools. Here, we showed how using MCM methods converges to the BS solutions but yield much more information in the process.

Monte Carlo simulations are particularly useful for pricing and risk management of complicated financial instruments such as more advanced options and derivatives which are not necessarily fully understood by the market. As such they are an indispensable tool for the RM analyst but care should be taken that business knowledge is well integrated in the underlying modeling process.

Project 8

In a financial market of your choice select three different types of options and valuate them according to the BS equation (if applicable) and a simulation model.

Are the generated prices close to what is observed in the market. Why or why not?

Lecture Nine: Operational Risk

Operational risk stems directly from the staffing and processes in any given organization. It is underlined by the fact that action or inaction that occurs in the run of normal business operations may create conditions for loss. This can have either benevolent or malevolent causes but the key fact of importance to management is that it generates negative cash flow, reputational damage or decreases efficiency. While it is not always straightforward to delineate what exactly constitutes operational risk and differentiate it from other types, we will formulate a working definition of operational risk, see how it is managed and then proceed to model it formally via means of a numeric simulation. A more detailed overview can be found in Crouhy et al. (2005) or in the Basel accords (2004, 2011).

Definitions and Types

The Basel Accord definition is a natural starting point for this journey. There operational risk is defined as **the risk of loss, resulting from inadequate or failed internal processes, people, and systems, or from external events**. This brings us to three large groups of sources of this type of risks.

These are as follows:

- **People risk** - here we include both problems pertaining to inadequate training or insufficient background and effectiveness, as well as malevolent or fraudulent activity by employees. The effects can vary from marginal to quite dramatic. A vivid example of the last is Societe Generale's Jerome Kervier who has convicted in 2008 for forgery, breach of truest, and misuse of the bank's resources that resulted in a loss of 4.9 billion EUR for the bank.
- **Process risk** - connected with the structuring and execution of business processes in the organization
  - *Model Risk* - connected with the formulation, execution, interpretation and use of quantitative and qualitative models for business purposes
  - *Transaction Risk* - problems pertaining to the execution of transactions, the complexity of offering, booking and settlement errors, poor documentation issues, and risks stemming from contractual obligations
  - *Operational Control Risks* - stemming from the inability to fully control execution. In a financial organization this will have to do with exceeding limits, security and volume risks, whereas in a production one this will pertain mostly to the compliance and quality of execution.
- **Systems and technology risk** - the advent and increase in usage of information and communication technologies means that more processes are digitized and more (in some cases - all) transactions are electronic. This in turn leads to a hike in technology-related risks - e.g. system failures, programming errors, information risks, breach of security, system attacks by external parties, telecommunication risks. A

striking example is the series of cyber attack against Estonia in 2007, that led to a massive denial of service by government and public web-sites and led to heightened security concerns and the inability of citizens to fully use the country's sophisticated electronic services.

The best practices for managing operational risk closely parallel the generic risk management cycle we already developed. The three key groups of drivers for effective RM processes can be defined as follows:

- **Create a receptive risk management environment** - executive management must be actively involved in setting the RM framework for operations, and be aware of the importance of operational risk on general business performance. This starts from creating processes, procedures, and rules for ORM, as well as providing adequate resources and staffing levels, to incentivizing senior management to implement the RM policy in a comprehensive way.
- **Manage Risk** - this include the identification of relevant risks, their measurement and rigorous modeling, as well as their continuous monitoring and taking steps for avoiding or mitigating the negative effects. A key point here is that any organization needs to be aware of its "operational risk catalog" - a list of all possible risks that stem from its people, processes, and technology and to estimate what each risk contribution to the overall performance (or bottom line) is.
- **Dissminate Information** - a natural final stage is the dissemination of information. The extent of disclosure varies widely depending on the type of organization. Publicly listed ones (such as large banks, funds, large companies) are legally required to make full disclosure as their ORM can have crucial implications for their cash flow and thus affect share prices. Non-public ones can afford to be more selective in their disclosure and only fully inform relevant stakeholders.

Measuring Operational Risk

A first and important step in measuring operational risk is the precise definition of what types of events fall in this group. Leveraging the risk catalog and pinpointing at what process junction those occur via usage of business process maps is extremely helpful in this case. A common way to quantify the risk is by considering the loss stemming from a given event.

The Basel capital accord (2004, 2011) considers the following groups of operational risk loss events:

- *Internal Fraud*
- *External Fraud*
- *Employment practices and workplace safety*
- *Client, product and business practices*
- *Damage to physical assets*
- *Business disruption and system failure*
- *Execution, delivery, and process management*

Depending on the type of organization some groups will be naturally more relevant than others. A financial organization will be more exposed to fraud activities, while a production one will suffer from risk associated with workplace safety and process execution and delivery. Once the event has occurred it can be quantified and this number can be used to characterize the magnitude of the operational risk. Those costs come into three groups.

The first one is the **cost-to-fix** that refers to the direct external costs that are paid in order to rectify the event. Here we may include legal, labor, and resource cost. The **write-down costs** are those associated with the loss of value that a company's asset experiences in the case of the risk event. This may be either financial or non-financial asset. Finally, the **resolution cost** refers to the totality of all expenses incurred to fully rectify the event (external payments and write-downs) as well as the cost to return to normal operations (restitution). Those definitions are somewhat narrow as they fail to include the foregone benefits and revenue due to the risk event but tend to be more operational as they are easier to quantity and more intuitive from accounting perspective. The RM analyst may nevertheless decide to include foregone profit as an additional quantification metric either on its own or as a part of the resolution cost.

The Log-normal Distribution

The exact distribution of the operation loss events has key implications for risk management and should therefore be carefully selected by the risk manager. An option that is commonly used is to leverage the Log-normal distribution. The Log-normal is a continuous distribution where the logarithm of the variable is normally distributed. More specifically, if X is log-normally distributed and μ and σ are the mean and standard deviation of the variable's natural logarithm, then:

$$X = e^{\mu + \sigma Z}$$

On a log scale the μ and σ parameters are the location and scale parameter, respectively. Denoting the mean and standard deviation of the non-log sample values as m and s, the following relationships hold:

$$\mu = \ln\left(\frac{m}{\sqrt{1 + \frac{s^2}{m^2}}}\right)$$

And:

$$\sigma = \sqrt{\ln\left(1 + \frac{s^2}{m^2}\right)}$$

The overall probability density function of this distribution is defined as follows:

$$P(X) = \frac{1}{s\sqrt{2\pi}x} e^{-(\ln x - m)^2/(2s^2)}$$

From here we can calculate the mean and standard deviation of the log-normal distribution. The mathematical expectation is:

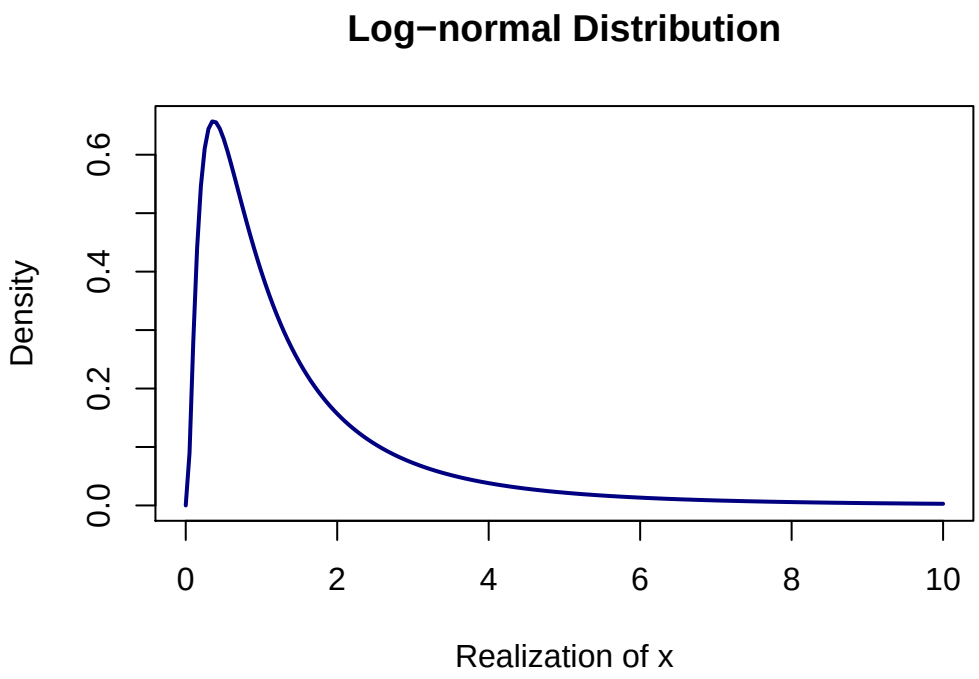
$$E[x] = e^{\mu + \frac{1}{2}\sigma^2}$$

And the standard deviation is defined as follows:

$$SD[X] = E[X]\sqrt{e^{\sigma^2} - 1}$$

We can plot the log-normal distribution in R using the `dlnorm` command. If no μ or σ are given, then in defaults to $\mu = 0$ and $\sigma = 1$, respectively. The graph is presented below.

```
x <- seq(from = 0, to = 10, by = 0.05)
plot(x, dlnorm(x, 0,1), type="l", col="navyblue", ylab="Density",
     xlab="Realization of x", main = "Log-normal Distribution",
     lwd=2)
```



The overall form of the distribution clearly shows that the overwhelming number of events tend to be of relatively small value, but there are certain events with extremely high values that happen with non-zero probability (long tails). This captures the reality of many situations in which operational risk is prevalent. The overwhelming number of errors tend to

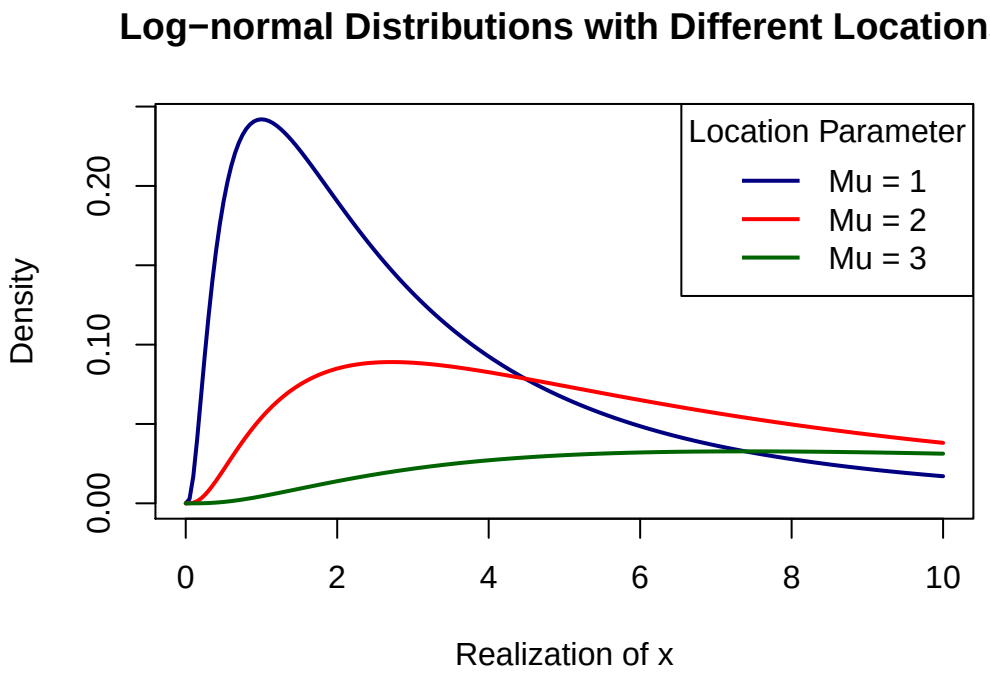
be relatively minor and easy to fix (e.g. typing errors), but there are also a small number of errors with potentially catastrophic consequences (e.g. IT system failure).

The two key parameters that change the form of the distribution are the location (μ) and the scale (σ) ones. Different values of the location effect the shape of the distribution. Here we present the three cases in which μ is equal to 1, 2, and 3, respectively.

```
x <- seq(from = 0, to = 10, by = 0.05)
labels = c("Mu = 1", "Mu = 2", "Mu = 3")
colors = c("navyblue", "red", "darkgreen")
plot(x, dlnorm(x, 1,1), type="l", col="navyblue", ylab="Density",
     xlab="Realization of x",
     main = "Log-normal Distributions with Different Locations",
     lwd=2) + lines(x, dlnorm(x, 2,1), col="red", lwd=2) + lines(x,
     dlnorm(x, 3,1), col="darkgreen", lwd=2)
```

```
## integer(0)
```

```
legend("topright", title="Location Parameter",
       labels, lwd=2, col=colors)
```



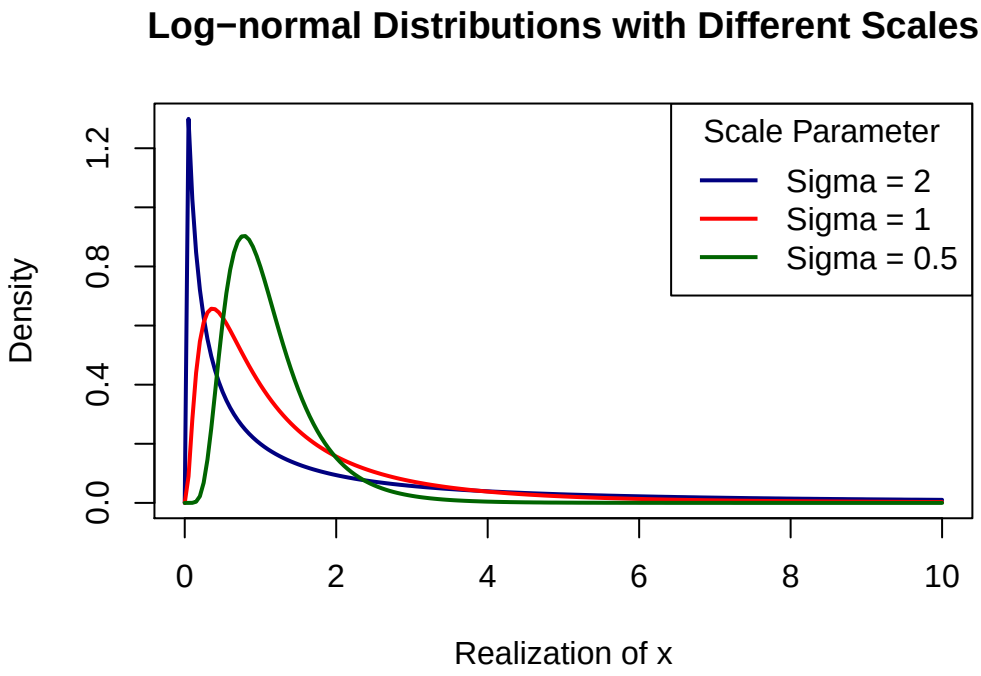
The other crucial parameters is the σ (scale) that determines the overall shape and peak of the distribution. Here we present the shape with σ equal to 2, 1, and 0.5.

```
x <- seq(from = 0, to = 10, by = 0.05)
labels = c("Sigma = 2", "Sigma = 1", "Sigma = 0.5")
colors = c("navyblue", "red", "darkgreen")
```

```
plot(x, dlnorm(x, 0,2), type="l", col="navyblue", ylab="Density",
     xlab="Realization of x",
     main = "Log-normal Distributions with Different Scales",
     lwd=2) + lines(x, dlnorm(x, 0,1), col="red", lwd=2) + lines(x,
     dlnorm(x, 0,0.5), col="darkgreen", lwd=2)
```

```
## integer(0)
```

```
legend("topright", title="Scale Parameter",
       labels, lwd=2, col=colors)
```



An important thing to note when using the log-normal distribution is that its behavior varies significantly with different numbers for the scale. What is more, the change is not monotonic and predictable but exhibits shifts at some σ values. This can be best seen when we work interactively with this parameter and observe changes. The following code snippet leverages the `manipulate` library and allows the user to interactively observe changes by varying the scale parameter.

```
library(manipulate)
manipulate(plot(x, dlnorm(x = x, meanlog = 0, sdlog = k), type="l",
                col="navyblue", lwd=2, ylim=c(0,1.5),
                main = "The Log-normal Distribution"), k=slider(0.5, 15))
```

At any rate, when operational risk is modeled through the use of the log-normal distribution, the RM analyst needs to ensure that the parameters are as precise as possible since even

small variations can lead to large shifts and produce conclusions that obscure rather than support the making of a risk-adjusted decision.

Modelling Operational Risk

The RM analyst should first decide on the specific base for risk measurement. A natural approach is to divide operations into different Lines of Business and measure operational risk separately as it is highly likely that different activities have a correspondingly different levels of riskiness. In each line of business, we should consider the following:

- **Exposure Indicator, EI** - what is the precise amount of exposure or the base upon which risks can happen. In the case of credit card fraud this would be the total amount of all credit card accounts combined. In the case of the risk of lawsuits by employees, it is the total number of all employees.
- **Probability of event, PE** - this is the chance or likelihood that a given risk event occurs. For example, in the case of credit card fraud it is the number of frauds divided by the number of accounts. In the case of employee lawsuits it is the number of lawsuits divided by total employees. In a sense, this is the empirical probability of the event happening that can help us construct the overall probability distribution.
- **Loss given event, LGE** - is the average loss over realized events - i.e. the total sum of losses divided by the number of realized events. In the case of credit card fraud this is the total sum of losses over the total number of frauds. For the risk of employee lawsuits, this is the total expenses incurred divided by the number of lawsuits. The analyst is well advised to also calculate the maximum loss in the case of event to gain a better appreciation of the risk boundaries.

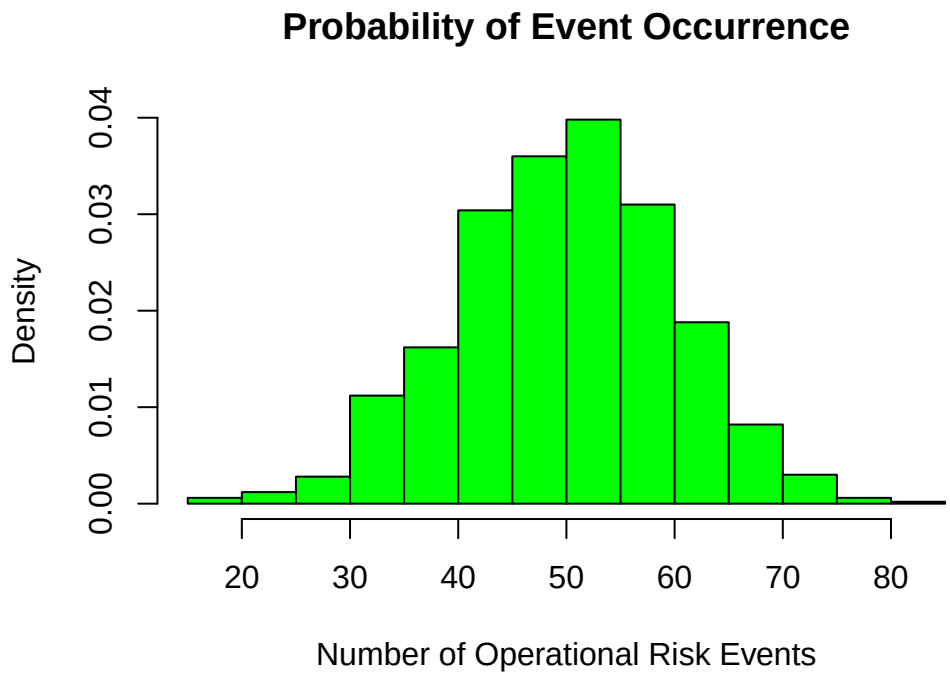
The total expected operational risk loss **OpR** is easily calculated once data on these quantities is obtained. It is the multiple of all possible events that can produce loss by their probability to do so (or total occurrences) times the average expected loss given an occurrence. In short:

$$OpR = EI * PE * LGE$$

We can show how using time series data on previous probabilities and losses we can model the annual probability of occurrence of given operational risk events, their expected losses, and the distribution for the annual expected loss. Assume that we have data on credit card frauds, and over the last years there are on average 50 of those events. Looking at the standard deviations of the numbers, the s.d. is equal to 10, and their distribution looks like the normal one. We can then randomly generate the probability distribution for the occurrence of a specific number of events. To this end we use the `rnorm` function in the following code snippet.

```
set.seed(789)
n.events <- rnorm(1000, 50, 10)
```

```
hist(n.events, col="green", freq = F,
     main = "Probability of Event Occurrence",
     xlab = "Number of Operational Risk Events")
```

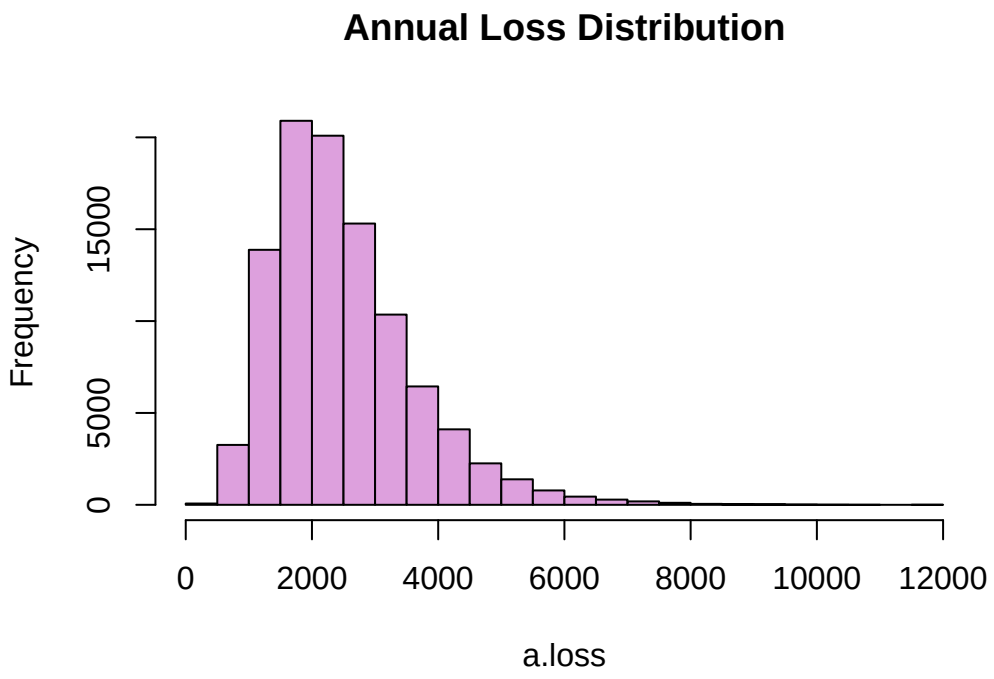


In addition to that we know that the average losses of a given fraud event are around 50, with a standard deviation of 20. If the distribution looks like a log-normal one, we can easily calculate its two key parameters - the μ and σ from this data and then randomly generate the whole distribution. This provides us with a notion of the chance of any given average loss per event in the year to come.

```
m <- 50
s <- 20
loss <- rlnorm(1000, log(50/sqrt(1+(s^2/m^2))), sqrt(log(1+(s^2/m^2))))
hist(loss, col="navyblue", freq = F, main = "Losses Given Event",
     xlab = "Average Loss Per Event")
```

Finally the annual loss is the product of the number of events that do occur times the average realized loss per event. We can simulate this using Monte Carlo methods. We first model the annual loss as the product of a randomly drawn number of events, multiplied by a randomly drawn average loss per event from the two distribution that we modeled. This gives us one possible realization of the expected annual loss. Then we iterate the same procedure 100,000 times to construct the full probability distribution.

```
set.seed(111)
a.loss <- rep(0, 100000)
for (i in 1:100000) {
  a.loss[i] <- rnorm(1, 50, 10) * rlnorm(1, log(50/sqrt(1+(s^2/m^2))),
                                         sqrt(log(1+(s^2/m^2))))
}
hist(a.loss, col="plum", freq = T, main = "Annual Loss Distribution")
```



Descriptive statistics can be used to effectively summarize this data.

```
library(psych)
describe(a.loss, skew = F)
```

```
##    vars     n    mean      sd    min      max    range   se
## X1    1 1e+05 2496.64 1127.63 278.45 11918.22 11639.78 3.57
```

All in all, given the information we had, we should expect a mean annual operational risk loss from credit card fraud to amount to 2496.6 with a standard deviation of 1127.6, and a minimum and maximum of 278.5 and 11,918.2, respectively.

Using the `fitdistr` command the analyst can also derive the μ and σ parameters of the log-normal distribution and reconstruct it as needed.

```
library(MASS)
fitdistr(a.loss, "lognormal")
```

```
##      meanlog         sdlog
##   7.7274090866   0.4396460454
##  (0.0013902829) (0.0009830784)
```

Having this distribution, we can also derive metrics for the operational Value at Risk **OpVaR** at a desired percentage (90%, 95%, 99%) by applying the quantile function for this distribution - `qnorm` - in exactly the same way as we did with the VaR metric.

While this example is relatively simple and models of operational risk can greatly increase in

complexity, it it still useful to overview the major ideas on how to think probabilistically and rigorously model risk so that management implications are clear and useful.

Providing for Risk

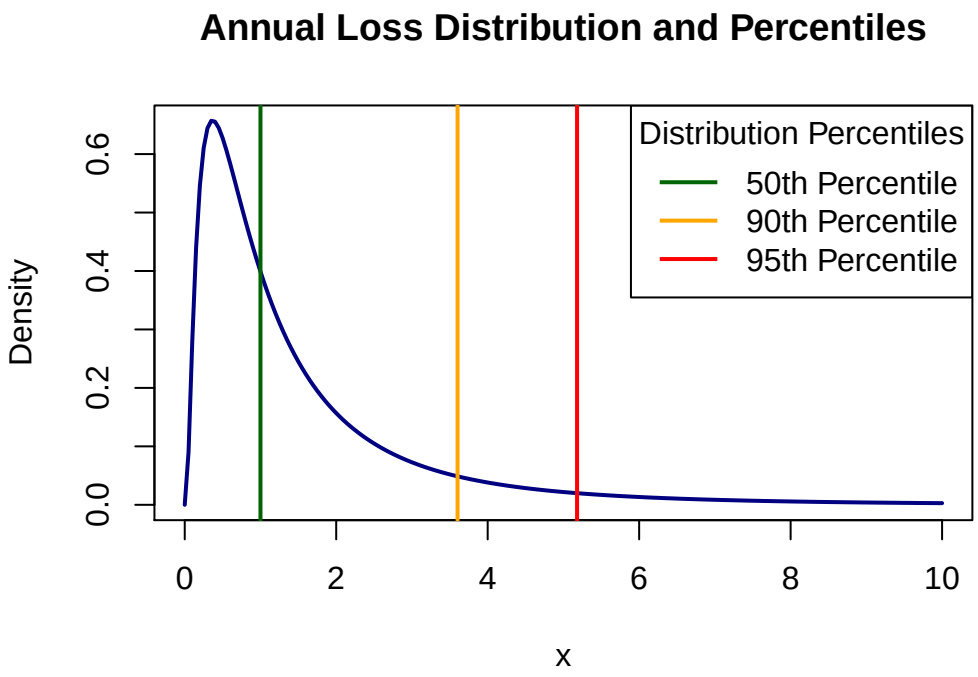
A final note concerning operational risk is about what part of the operational risk should be explicitly counted and provided for in management decision-making. Essentially all activities connected to people, processes, technology, or the external environment bring about a certain level of inherent risk. It is often the case that the largest fraction of operational risk is part of the normal running of the business and is therefore explicitly or implicitly calculated in the annual budgeting and planning process. It is only the unexpected (and possibly large impact) risks than should be given a particular note. We can illustrate this graphically.

Assume that the annual operational risk losses follows a log-normal distribution with $\mu = 0$, and $\sigma = 1$. We draw its 50th, 90th, and 95th percentile lines in the following graph.

```
x <- seq(0, 10, 0.05)
labels = c("50th Percentile", "90th Percentile", "95th Percentile")
plot(x, dlnorm(x,0,1), type="l", col="navyblue", lwd=2, ylab="Density",
     main="Annual Loss Distribution and Percentiles") +
  abline(v=qlnorm(0.5, 0,1), col="darkgreen", lwd=2) +
  abline(v=qlnorm(0.9, 0,1), col="orange", lwd=2) +
  abline(v=qlnorm(0.95, 0,1), col="red", lwd=2)
```

```
## integer(0)
```

```
legend("topright", title="Distribution Percentiles", labels, lwd=2,
       col=c("darkgreen", "orange", "red"))
```



Risks up a certain point are reasonably expected and form the staple of regular processes with their numerous but minor and relatively inexpensive to fix errors. Usually these are part of the business understanding and resources are devoted to monitor and correcting them (e.g. direct supervision, performance review, quality control processes, etc.). In the graph we have delineated those as the risk left from the 50th percentile.

The risks between the 50th and the 90th percentile are less expected and can have potentially somewhat larger impact. The RM analyst needs to devote specific attention and communicate clearly that these are not only possible but rather probable. Even more unexpected are the risks between the 90th and the 95th percentile (orange and red lines). The organization experiences losses from them only on an irregular basis and are those are thus very unexpected but can cost a much larger loss than minor operational mistakes.

Finally, the risks in the top 5% of the distribution are very unlikely but due to their extremely large loss potential these can be extremely damaging and even catastrophic. What is more, internal data can be scarce on these events and thus may show them as less likely than they actually are. In this domain, the RM analyst would do wisely to leverage external data, run simulations and even switch to another distribution to model this tail. At any rate, the risk manager needs to be particularly vigilant for risks above the 90th percentile as they are unexpected but with potentially extreme impact.

Project 9

Collect data on a risk event from the four key risk groups:

- Process Risks
- People Risks
- Technology Risks
- External Events Risks (Exogenous shocks)

Calculate their respective EI, PE, LGE. Model their probability distributions and derive the annual loss distribution for each of these risks. Combine them and derive their joint distribution. What is the combined expected loss from those four events, and what is the risk (standard deviations, ranges) associated with it? What is the OpVaR (operational risk value at risk) at 95% and 99%?

Lecture Ten: Classifying Credit Risks

One of the common tasks in risk management is classifying cases into those who are likely to yield favorable outcomes and those likely to yield unfavorable outcomes. Once this is done, the organization can increases its exposure to the former, and decrease it to the latter. The classical example is managing credit risk - the lender (e.g. a bank) tries to classify those applying for credit as to whether they will pay back or not.This leads to higher approval rates of good payers, thus decreasing credit risk.

In this lecture we have an overview of this process. Initially we begin with credit scoring to outline the philosophy behind managing credit risk. Then we present traditional classification methods such as the logistic regression and the linear discriminant analysis. We then proceed to overview a few popular machine learning models, and outline the implications of big data for RM. All the classification methods reviewed here and many more that can be of use to the RM professional can be found in the outstanding work of Hastie et al. (2011). Another example of such classification problems is found in Gerunov (2016) together with plenty of references for the interested reader.

Credit Scoring and Coefficients

Credit Scoring is the process of using qualitative and quantitative information in order to give an assessment of an individual's probability to pay back his credit. It uses a wide array of publicly and privately available information to make a risk profile, given age, income and wealth, credit history and a host of other relevant indicators. Such a credit score essentially helps vendors to classify applicants into those who are likely to perform on their loans, and those who are not. This can be either done through a binary classifier (e.g. Yes/No) or a continuous probabilistic score (e.g. individuals with a score of 6.7 have 90% probability to repay).

The quality of classification can be judged on a number of different metrics. Those are as follows:

- *Sensitivity* - measures the proportion of positives that are correctly identified (true positives).
- *Specificity* - measures the proportion of negatives that are correctly identified (true negatives).
- *Positive Predictive Value* - measures the proportion of positive results that are true positives.
- *Negative Predictive Value* - measures the proportion of negative results that are true negatives.
- *Prevalence* - measures the probability of being positive in the population.

- *Detection Rate* - measures the rate of positives that are predicted to be positive in the population.
- *Detection Prevalence* - measures the prevalence of detected positive events.
- *Balanced Accuracy* - measures the average accuracy of model (mean of sensitivity and specificity).
- *Overall accuracy* - measures the total number of correctly predicted observation over the total number of observation. This is probably the most popular metric, giving a sense of model quality and applicability.

We should keep all of them in mind as sometimes the price of missclassification differs depending on the case of misclassification. For example, classifying a borrower as false negative (meaning that you think he will not repay, while he will) and rejecting the loan leads to foregone profit. On the other hand classifying him as a false positive means direct loss. The latter may have more gravity than the former. The RM professional therefore needs not only to perform the analysis but to calibrate the model so that costly classification mistakes are avoided (even if this means incurring more low-cost ones).

Apart from the applications in banking, this scoring/classification can be used to model the conditions in a large variety of different environments. In the case of health insurance, the health and demographic profile of an individual may change his contribution. Auto insurance dealers may use historical information to sort out safe drivers and give them better premiums. Leasing offices may want to differentiate their offerings depending on the riskiness of their customers. Overall, classification problems are ubiquitous and their successful solution helps the organization better understand and manage the risks it faces. We now turn to traditional statistical methods for classifications.

Traditional Classification Approaches

In the case of discrete choice or classification problems, the **logistic regression** is a popular choice and was pioneered early in modeling problems. In this case the probability of choosing a certain outcome is approximated by the logistic function, or:

$$P(y|x_i) = \frac{\exp\left(\beta_0 + \sum_{i=1}^{n} \beta_i x_i\right)}{1 + \exp\left(\beta_0 + \sum_{i=1}^{n} \beta_i x_i\right)}$$

The estimated beta coefficients show the strength of association between a given independent variable like a demographic or a situational factor, and the dependent one – the choice. Those coefficients have the interpretation of increasing the odds of selection. The simple regression can be expanded to a multinomial logistics regression and has been widely used in modeling applications.

To illustrate its use, we leverage the `caret` packages, which contains an extremely large number of alternative modeling methods with a similar interface and command structure. From there we have a look at German credit data - data from 1000 loans from a south

German bank, of which 700 were performing, and 300 were not. We present data statistics as usual.

```
library(caret)
library(psych)
data(GermanCredit)
describe(GermanCredit, skew=F, ranges = F)
```

```
##                                      vars    n    mean      sd    se
## Duration                                1 1000   20.90   12.06  0.38
## Amount                                  2 1000 3271.26 2822.74 89.26
## InstallmentRatePercentage               3 1000    2.97    1.12  0.04
## ResidenceDuration                       4 1000    2.85    1.10  0.03
## Age                                     5 1000   35.55   11.38  0.36
## NumberExistingCredits                   6 1000    1.41    0.58  0.02
## NumberPeopleMaintenance                 7 1000    1.16    0.36  0.01
## Telephone                               8 1000    0.60    0.49  0.02
## ForeignWorker                           9 1000    0.96    0.19  0.01
## Class*                                 10 1000    1.70    0.46  0.01
## CheckingAccountStatus.lt.0             11 1000    0.27    0.45  0.01
## CheckingAccountStatus.0.to.200         12 1000    0.27    0.44  0.01
## CheckingAccountStatus.gt.200           13 1000    0.06    0.24  0.01
## CheckingAccountStatus.none             14 1000    0.39    0.49  0.02
## CreditHistory.NoCredit.AllPaid         15 1000    0.04    0.20  0.01
## CreditHistory.ThisBank.AllPaid         16 1000    0.05    0.22  0.01
## CreditHistory.PaidDuly                 17 1000    0.53    0.50  0.02
## CreditHistory.Delay                    18 1000    0.09    0.28  0.01
## CreditHistory.Critical                 19 1000    0.29    0.46  0.01
## Purpose.NewCar                         20 1000    0.23    0.42  0.01
## Purpose.UsedCar                        21 1000    0.10    0.30  0.01
## Purpose.Furniture.Equipment            22 1000    0.18    0.39  0.01
## Purpose.Radio.Television               23 1000    0.28    0.45  0.01
## Purpose.DomesticAppliance              24 1000    0.01    0.11  0.00
## Purpose.Repairs                        25 1000    0.02    0.15  0.00
## Purpose.Education                      26 1000    0.05    0.22  0.01
## Purpose.Vacation                       27 1000    0.00    0.00  0.00
## Purpose.Retraining                     28 1000    0.01    0.09  0.00
## Purpose.Business                       29 1000    0.10    0.30  0.01
## Purpose.Other                          30 1000    0.01    0.11  0.00
## SavingsAccountBonds.lt.100             31 1000    0.60    0.49  0.02
## SavingsAccountBonds.100.to.500         32 1000    0.10    0.30  0.01
## SavingsAccountBonds.500.to.1000        33 1000    0.06    0.24  0.01
## SavingsAccountBonds.gt.1000            34 1000    0.05    0.21  0.01
## SavingsAccountBonds.Unknown            35 1000    0.18    0.39  0.01
## EmploymentDuration.lt.1                36 1000    0.17    0.38  0.01
```

```
## EmploymentDuration.1.to.4                 37 1000    0.34    0.47  0.01
## EmploymentDuration.4.to.7                 38 1000    0.17    0.38  0.01
## EmploymentDuration.gt.7                   39 1000    0.25    0.43  0.01
## EmploymentDuration.Unemployed             40 1000    0.06    0.24  0.01
## Personal.Male.Divorced.Seperated          41 1000    0.05    0.22  0.01
## Personal.Female.NotSingle                 42 1000    0.31    0.46  0.01
## Personal.Male.Single                      43 1000    0.55    0.50  0.02
## Personal.Male.Married.Widowed             44 1000    0.09    0.29  0.01
## Personal.Female.Single                    45 1000    0.00    0.00  0.00
## OtherDebtorsGuarantors.None               46 1000    0.91    0.29  0.01
## OtherDebtorsGuarantors.CoApplicant        47 1000    0.04    0.20  0.01
## OtherDebtorsGuarantors.Guarantor          48 1000    0.05    0.22  0.01
## Property.RealEstate                       49 1000    0.28    0.45  0.01
## Property.Insurance                        50 1000    0.23    0.42  0.01
## Property.CarOther                         51 1000    0.33    0.47  0.01
## Property.Unknown                          52 1000    0.15    0.36  0.01
## OtherInstallmentPlans.Bank                53 1000    0.14    0.35  0.01
## OtherInstallmentPlans.Stores              54 1000    0.05    0.21  0.01
## OtherInstallmentPlans.None                55 1000    0.81    0.39  0.01
## Housing.Rent                              56 1000    0.18    0.38  0.01
## Housing.Own                               57 1000    0.71    0.45  0.01
## Housing.ForFree                           58 1000    0.11    0.31  0.01
## Job.UnemployedUnskilled                   59 1000    0.02    0.15  0.00
## Job.UnskilledResident                     60 1000    0.20    0.40  0.01
## Job.SkilledEmployee                       61 1000    0.63    0.48  0.02
## Job.Management.SelfEmp.HighlyQualified    62 1000    0.15    0.36  0.01
```

We use the `train` command to fit a model. Detailed syntax can be found by typing `?train`. First we fit a logistic regression model, selecting a subset of predictors (duration, amount of credit, age of applicant, credit history, and house ownership), as follows:

```
log <- train(Class~ Duration + Amount + Age + CreditHistory.Critical +
      Housing.Own, data=GermanCredit, method="glm", family="binomial")
summary(log)
```

```
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2044  -1.1773   0.6446   0.8502   1.7026
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)             5.614e-01  2.932e-01   1.915  0.05550 .
```

```
## Duration                -3.207e-02  7.485e-03  -4.284 1.83e-05 ***
## Amount                  -2.146e-05  3.143e-05  -0.683  0.49475
## Age                      1.397e-02  6.743e-03   2.072  0.03822 *
## CreditHistory.Critical   8.518e-01  1.799e-01   4.735 2.19e-06 ***
## Housing.Own              5.188e-01  1.546e-01   3.356  0.00079 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1221.7  on 999  degrees of freedom
## Residual deviance: 1130.1  on 994  degrees of freedom
## AIC: 1142.1
##
## Number of Fisher Scoring iterations: 4
```

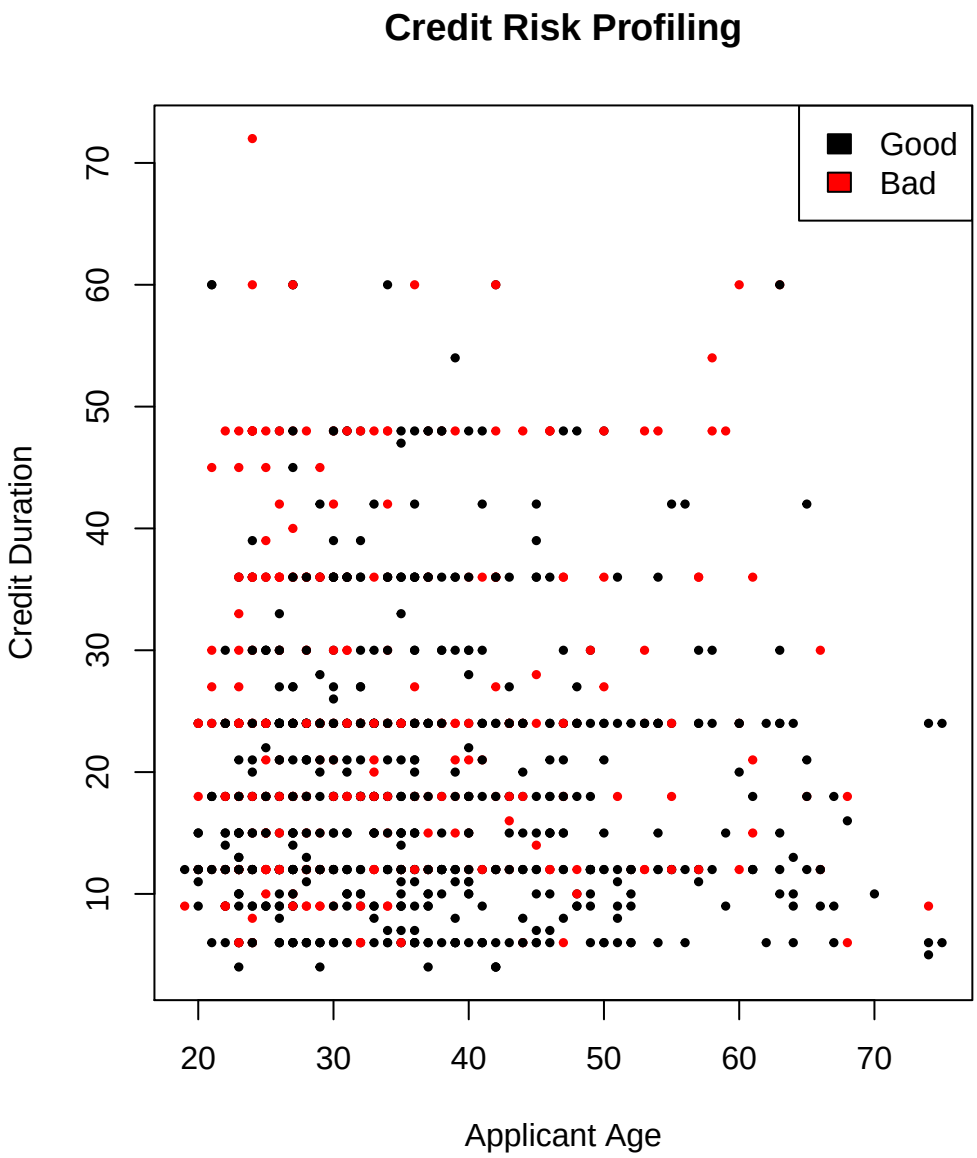
This model shows classification accuracy of around 72%. The relative importance of different predictors can be obtained by using the `varImp` command:

```
varImp(log)
```

```
## glm variable importance
##
##                         Overall
## CreditHistory.Critical  100.00
## Duration                 88.88
## Housing.Own              65.98
## Age                      34.30
## Amount                    0.00
```

We can also plot the class of credits across key dimensions. Here we investigate the link between the applicant age, and the credit distribution.

```
library(car)
plot(GermanCredit$Age, GermanCredit$Duration,
     col=recode(var = GermanCredit$Class,
     recodes="'Good'='Black'; 'Bad'='Red'"), pch=19, xlab="Applicant Age",
     ylab="Credit Duration", main="Credit Risk Profiling", cex=0.5)
legend(x="topright", legend=c("Good", "Bad"), fill=c("Black", "Red"))
```



This graph for example shows that younger people, taking long-term credits are likely to be unable to pay them (maybe due to misalignment of expectations about future income). At any rate, the lender should pay careful attention to such patterns.

It turns out that credit history is the most important factor, followed by the credit duration. Naturally, the RM professional will want to investigate and experiment with all data at her disposal in order to build an optimal model.

An alternative but still very popular approach is the **linear discriminant analysis**. It aims to classify a binary dependent variable by constructing the best linear combination of the observed variables. Let us assume the two conditional distributions of the outcome y and predictor x to be: $p(y|x)$ and $p(x|y)$, and further that these two follow a normal distribution

with means μ_y and μ_x, and with covariations of σ_{yx} and σ_{xy}, respectively. If the condition $\sigma_{yx} = \sigma_{xy}$ holds, then classification can be obtained via the following condition:

$$(x - \mu_y)^T \sigma_{xy}^{-1} (x - \mu_y) - (x - \mu_x)^T \sigma_{yx}^{-1} (x - \mu_x) < T$$

Here T is a parameter that is given some threshold value. Due to its simplicity and relative ease of interpretation, (linear) discriminant analysis continues to be of use for classification in risk management. The LDA will use the complete set of predictors in the data set, which will give us the opportunity to survey the importance of all the data listed. We illustrate its use here:

```
lda <- train(Class~Duration + Amount + Age + CreditHistory.Critical +
               Housing.Own, data=GermanCredit, method="lda")
lda
```

```
## Linear Discriminant Analysis
##
## 1000 samples
##    5 predictor
##    2 classes: 'Bad', 'Good'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 1000, 1000, 1000, 1000, 1000, 1000, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.7048043  0.1341179
```

```
varImp(lda)
```

```
## ROC curve variable importance
##
##                          Importance
## Duration                     100.00
## CreditHistory.Critical        47.98
## Age                           21.40
## Housing.Own                   15.69
## Amount                         0.00
```

For classification purposes the most crucial variables are the status of checking accounts, the duration of credit asked, the credit history, the age of applicant, and house ownership. Those are the ones that can be used for credit scoring. Alternatively, the analyst can use the trained model to predict what the outcome is (command `predict`), using all variables, as it feeds new data into the model.

A confusion matrix can be used to assess the quality of the model and its prediction accuracy.

```
confusionMatrix(data = predict(lda, GermanCredit), GermanCredit$Class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction Bad Good
##       Bad   52   37
##       Good 248  663
##
##                Accuracy : 0.715
##                  95% CI : (0.6859, 0.7428)
##     No Information Rate : 0.7
##     P-Value [Acc > NIR] : 0.1585
##
##                   Kappa : 0.1508
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.1733
##             Specificity : 0.9471
##          Pos Pred Value : 0.5843
##          Neg Pred Value : 0.7278
##              Prevalence : 0.3000
##          Detection Rate : 0.0520
##    Detection Prevalence : 0.0890
##       Balanced Accuracy : 0.5602
##
##        'Positive' Class : Bad
##
```

The predictive accuracy of the model stands at 72%, which is only a very, very modest increase over the no-information rate of 70%. We thus conclude that the model is not particularly good. Similar results are shown by the logistic regression, which points that traditional models sometimes have very limited success in complex environments.

Machine Learning Algorithms

Naïve Bayes classifiers are applications of the Bayes theorem, whereby the classification problem is solved by constructing the joint probability distributions of variables under interest and then using that for the purposes of class assignment. More specifically, we are interested in the conditional probability distribution of observations y_i over classes C_k given a number of features x_i that are pertinent to the classification problem. Assuming that any feature is independent of the others, this conditional distribution can be described as:

$$p(C_i|x_i) = \frac{1}{Z} p(C_k) \prod_{i=1}^{n} p(x_i|C_k)$$

Here we use Z to denote a scaling factor with $Z = p(x_i)$. Once the conditional probability distribution is algorithmically constructed, the classifier is complete once a decision rule is specified. A common choice is to accept the most likely class, thus assigning an observation y_i to class C_k under the following condition:

$$y_i = argmax[p(C_k) \prod_{i=1}^{n} p(x_i|C_k)]$$

The naive Bayes classifiers are not as computationally intensive as other machine learning methods but still perform well in practice. While their major assumption of feature independence is rarely achieved in reality and the posterior class probabilities may thus be imprecise, overall classification results are sometimes at par with more sophisticated approaches. Such a model can be fit using the `nb` method as follows:

```
nb <- train(Class~., data=GermanCredit, method="nb")
```

```
varImp(nb)
```

```
## ROC curve variable importance
##
##   only 20 most important variables shown (out of 61)
##
##                                 Importance
## CheckingAccountStatus.none          100.00
## Duration                             74.80
## CheckingAccountStatus.lt.0           73.13
## CreditHistory.Critical               52.49
## SavingsAccountBonds.lt.100           50.00
## Age                                  41.09
## Housing.Own                          38.64
## Property.RealEstate                  34.07
## CheckingAccountStatus.0.to.200       33.66
## Amount                               31.91
## SavingsAccountBonds.Unknown          31.72
## Purpose.Radio.Television             30.47
## Property.Unknown                     28.81
## OtherInstallmentPlans.None           27.98
## Purpose.NewCar                       26.04
## Personal.Male.Single                 25.48
## EmploymentDuration.lt.1              25.48
## InstallmentRatePercentage            25.24
## Housing.Rent                         22.58
```

```
## Personal.Female.NotSingle           22.16
```

Here we clearly see that the same factors are important for classification in the Naive Bayes algorithm as they are in the more traditional ones. The consistency of results across methods is an important check and validation for their robustness.

Neural networks are models that are heavily influenced by the way the human brain works. It is structured by neurons that send activation impulses to each other, and so is the overall architecture of the neural network model. The different input nodes (independent variables) send activation impulses in the form of mathematical weighting functions to hidden layers of other nodes, which then produce impulses towards the final output node (dependent variable). Thus each input neuron (independent variable) can affect the class under study through a series of weighted functions, a representation known as the nonlinear weighting sum. If the node is denoted as x, and its activation function is K, the neuron's network function f can be expressed mathematically as:

$$f(x) = K(\sum_{i=1}^{n} w_i g_i(x))$$

In this case K is some pre-defined function, and $g_i(x)$ is a series of functions weighted by w_i. Estimation methods (learning) calculate the optimum weights given a set of conditions such as the functions used and the number of layers. Many neural networks can be trained on the same set of data with varying degrees of complexity. The optimal choice is guided by computational tractability and parsimony. Here we fit the simplest version of a neural network with only one hidden layer, using the sigmoid function. This can be done using the following command:

```
nnet <- train(Class~., data=GermanCredit, method="nnet")
```

To understand the improvement in classification accuracy, we estimate the confusion matrix for the neural network model.

```
pred <- predict(nnet, GermanCredit)
confusionMatrix(data = pred, GermanCredit$Class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction Bad Good
##       Bad  200   76
##       Good 100  624
##
##                Accuracy : 0.824
##                  95% CI : (0.799, 0.8471)
##     No Information Rate : 0.7
##     P-Value [Acc > NIR] : < 2e-16
##
```

```
##                   Kappa : 0.5712
##  Mcnemar's Test P-Value : 0.08297
##
##             Sensitivity : 0.6667
##             Specificity : 0.8914
##          Pos Pred Value : 0.7246
##          Neg Pred Value : 0.8619
##              Prevalence : 0.3000
##          Detection Rate : 0.2000
##    Detection Prevalence : 0.2760
##       Balanced Accuracy : 0.7790
##
##        'Positive' Class : Bad
##
```

The accuracy now sees an improvement over the traditional methods such as logistic regression and linear discriminant analysis. This is a recurrent result - machine learning approaches tend to produce better performance especially in the context of complex and non-linear data.

Decision trees provide an alternative approach to modeling. Their machine learning operationalization is merely an extension of the classical decision tree model, used in decision science. Given a test data, an algorithm splits classification cases at different nodes by picking the best classifier among the features tested. Let us assume a classification problem with k classes and N_m observations in the region R_m. Thus the probability of observation y_i belonging to class C_k at node m equals:

$$p_m k = \frac{1}{N_m} \sum_{i=1}^{n} I(y_i = C_k)$$

In that case an algorithm splits the classification region to find the best classifier x_i to put a class prediction C_k on observation y_i. Then every observation is classified at node m as the majority class of this node:

$$C(m) = argmax[p_{mk}]$$

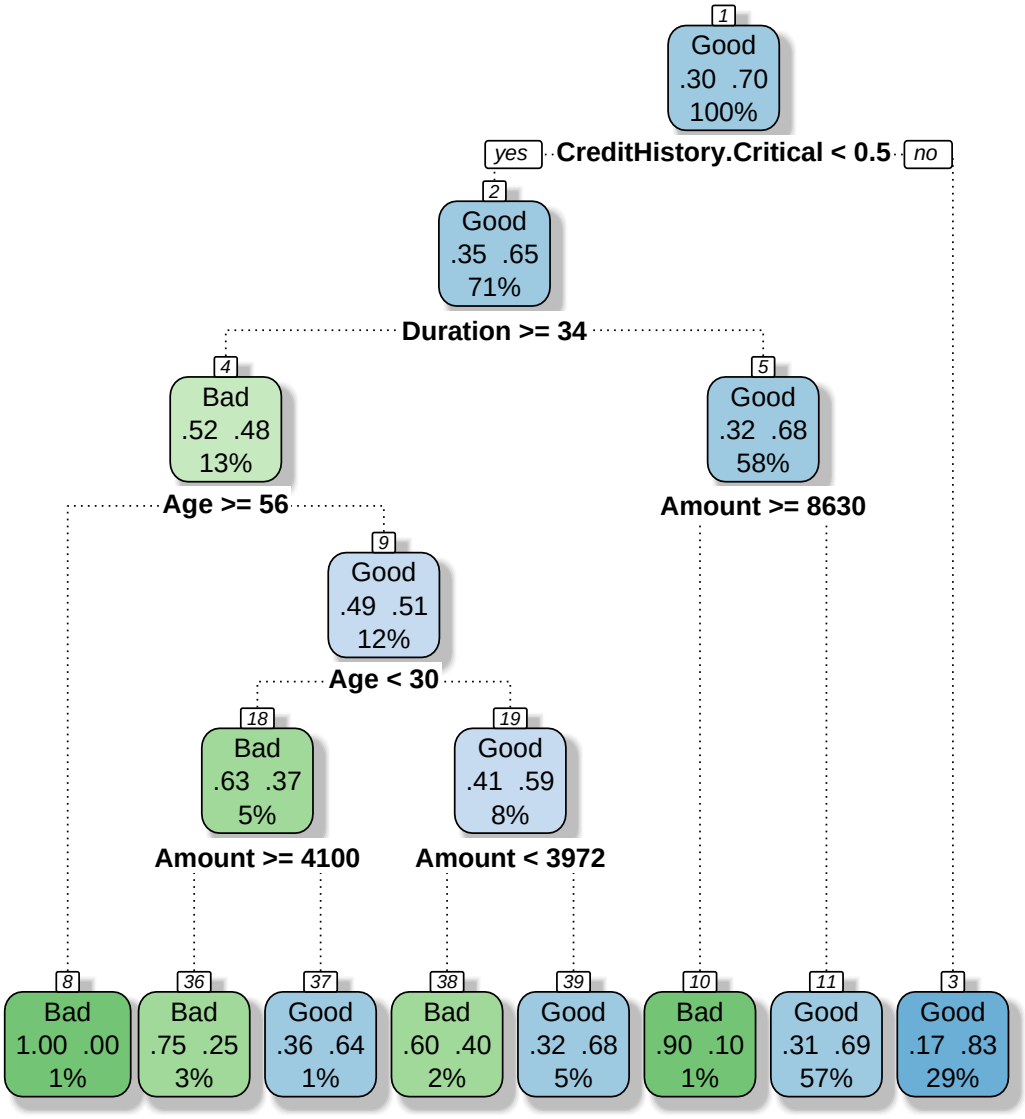
More classification nodes are created until some pre-determined number of is reached. After they are so grown, trees can be used to perform classification tasks in validation or test sets. Decision trees have the great advantage that they are easy to build, yet intuitive to interpret. In their visual form they can also be used for decision-making on the spot, especially in the case of a compact tree. Their major drawback stem from the fact that trees tend to overfit data, produce large variation and can be misled by local optima.

Trees can be fitted using the `rpart` method in the `train` command. Here, we will focus on a few indicators in order to obtain a more intuitive visualization. This trained model can be used for further classification.

```
tree <- train(Class~Duration + Amount + Age + CreditHistory.Critical +
                Housing.Own, data=GermanCredit, method="rpart")
```

We create a visualization of the model, using utilities from the Rattle package. Such a graph can be used for interactive decision making on the spot or communicated to a wide audience which will need to perform the risk classification task. The RM professional should be particularly careful to select a representative tree that adequately characterizes the decision problem.

```
library(rattle)
fancyRpartPlot(rpart(Class~Duration + Amount + Age + CreditHistory.Critical
                       + Housing.Own, data=GermanCredit))
```



Rattle 2017−May−17 12:59:28 ageru

Random forests essentially combine a pre-determined number of trees into a large ensemble that can be used for classification or regression. This process initiates by first selecting a bootstrapped subsample of the data from $b = 1 to B$, and then selecting a random number of features to be used for classification. The algorithm then builds a tree in such a way that the best variable/split point is created at node m in the same way as with classification and regression trees. Once the maximum number of trees and their respective number of terminal nodes is reached, those are combined in the forest ensemble T_b^B. Just like neural networks, random forests can model both continuous and discrete choice. The rule for continuous choice is:

$$f_k^B = \frac{1}{B}\sum_{b=1}^{B} T_b(y)$$

Under discrete choice, the classification problem is solved by taking the majority vote of trees as to the class of given observation y_i, thus obtaining:

$$C_k^B(y_i) = majority.vote[C_b(y_i)]_1^B$$

Random forest model are notable for their ease of interpretation and relatively limited parameters and tend to perform extremely well even in out-of-the-box applications. Here we fit a random forest model and investigate its properties.

```
rf <- train(Class~., data=GermanCredit, method="rf")
```

While the `caret` packages supports Random Forest models, the `randomForest` package provides much larger control over the model parameters. In out cases due to the small number of observations we might want to decrease the number of trees to 10 in order to avoid overfitting.

```
library(randomForest)
rf <- randomForest(Class~., data=GermanCredit, ntree = 10)
rf
```

```
##
## Call:
##  randomForest(formula = Class ~ ., data = GermanCredit, ntree = 10)
##                Type of random forest: classification
##                      Number of trees: 10
## No. of variables tried at each split: 7
##
##         OOB estimate of  error rate: 30.71%
## Confusion matrix:
##      Bad Good class.error
## Bad  118  181   0.6053512
## Good 123  568   0.1780029
```

We can observe how the model error rate changes as the number of trees grow. Typically, the error rate decreases fast as trees increase but at a point reaches a sudden stop and plateaus off. We see the same here, although due to the low number of individual trees it is not as pronounced.

```
plot(rf, main="Model Classification Error", lwd=2)
```

We calculate the confusion matrix for the Random Forest model as usual:

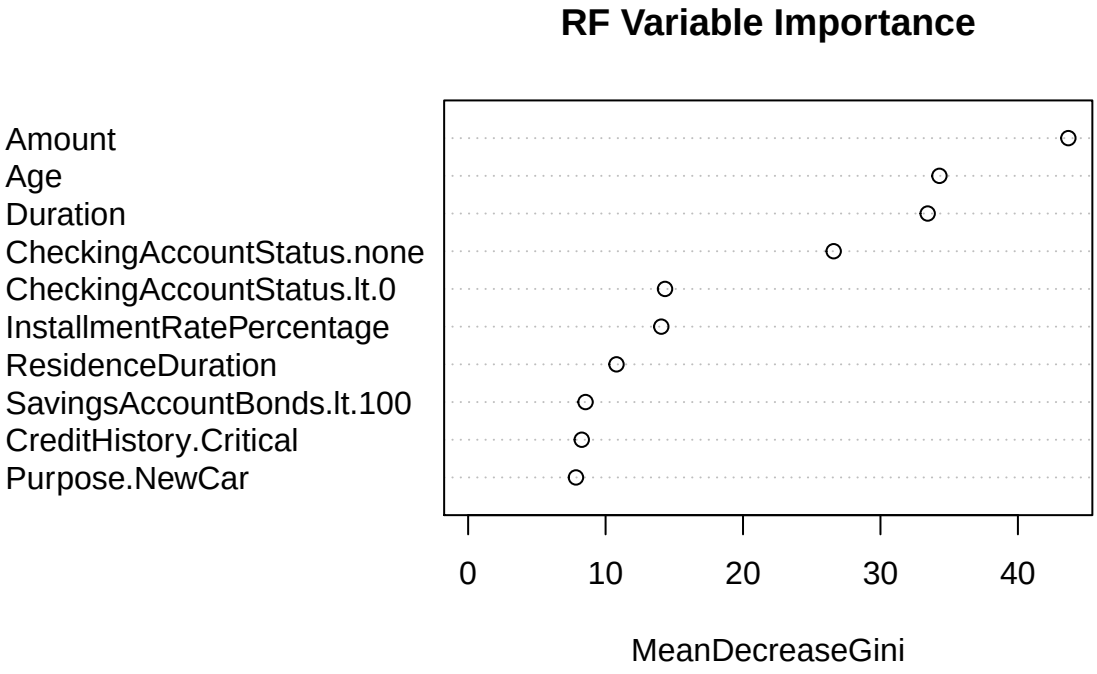
```
pred <- predict(rf, GermanCredit)
confusionMatrix(data = pred, GermanCredit$Class)
```

```
## Confusion Matrix and Statistics
##
##            Reference
## Prediction Bad Good
##       Bad  291    4
##       Good   9  696
##
##                Accuracy : 0.987
##                  95% CI : (0.9779, 0.9931)
##     No Information Rate : 0.7
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9689
##  Mcnemar's Test P-Value : 0.2673
```

```
##
##             Sensitivity : 0.9700
##             Specificity : 0.9943
##          Pos Pred Value : 0.9864
##          Neg Pred Value : 0.9872
##              Prevalence : 0.3000
##          Detection Rate : 0.2910
##    Detection Prevalence : 0.2950
##       Balanced Accuracy : 0.9821
##
##        'Positive' Class : Bad
##
```

The predictive accuracy of the model reaches a whopping 98%, which is much larger than any other of the models presented here. The Random Forest easily outperforms traditional methods, and tends to provide better results than alternative machine learning methods. This result holds across a significant number of studies. The class of Random Forest models thus provide a superior tool for classification (and regression) and have great potential for understanding and analyzing risk. We can also preview which factors were more beneficial for this classification rate, using the variable importance plot (we set it to display top 10):

```
varImpPlot(rf, n.var=10, main="RF Variable Importance")
```



We see the large and significant importance of three large determinants - amount and duration of the credit, and the age of the applicant. Again, the RM analyst needs to focus

the attention of both management and credit inspection on those and instrumentalize this knowledge in the business processes in order to successfully decrease risk.

Evaluating Classifiers: The ROC Curve

Many classification models have a tuning parameter that can crucially impinge on their quality. It is often an important part of the model-building process to fine-tune the classification model. One instrument to this end is the Receiver Operating Characteristic (or ROC) curve. It shows how the proportion between correctly predicted (true) positives and incorrectly predicted (false) positives varies with the variation of some variable like the cutoff point.

For example a logistic regression model will give numeric predictions for the likelihood of belonging to a certain class - e.g. there may be a 0.34 chance of being a true positive, or a 0.99. The prediction implies a certain classification cutoff - if the cost of mistakes is equal, then a 0.5 point is reasonable. In that instance all cases with a score below 0.5 can be classified as 0 or negative, and all above - as 1 or positive. If we decide to put the point at 0.75 then we skew classification to the negative cases, thus changing the proportion of true and false positives. High cutoffs will mean it is very difficult to classify a case as positive eliminating the false positives but at the cost of missing true ones (and vice versa). The ROC curve shows how this proportion changes with the change of the parameter in question.

Given a tuning parameter, θ, The ROC curve's coordinates are then defined as:

$$ROC(\theta) = (FPR(\theta), TPR(\theta))$$

More specifically, the x-coordinates are as follows:

$$ROC_x(\theta) = FPR(\theta) = \frac{FP(\theta)}{FP(\theta) + TN(\theta)} = \frac{FP(\theta)}{N}$$

And the y-coordinates are:

$$ROC_y(\theta) = TPR(\theta) = \frac{TP(\theta)}{FN(\theta) + TP(\theta)} = \frac{FP(\theta)}{P} = 1 - \frac{FN(\theta)}{P} = 1 - FNR(\theta)$$

The best point in the ROC curve is at (0,1), which essentially means that the classification yids 100% true positives and no false positives. The closer the actual algorithm reaches this point, the better it performs. This can also be expressed analytically with the area under the curve (AUC). If the area is equal to one, then performance is optimal. If the area is equal to 0.5 (or the ROC curve coincides with the bisector), then the classifier is worthless. In the latter case it gives results that are exactly equal to chance.

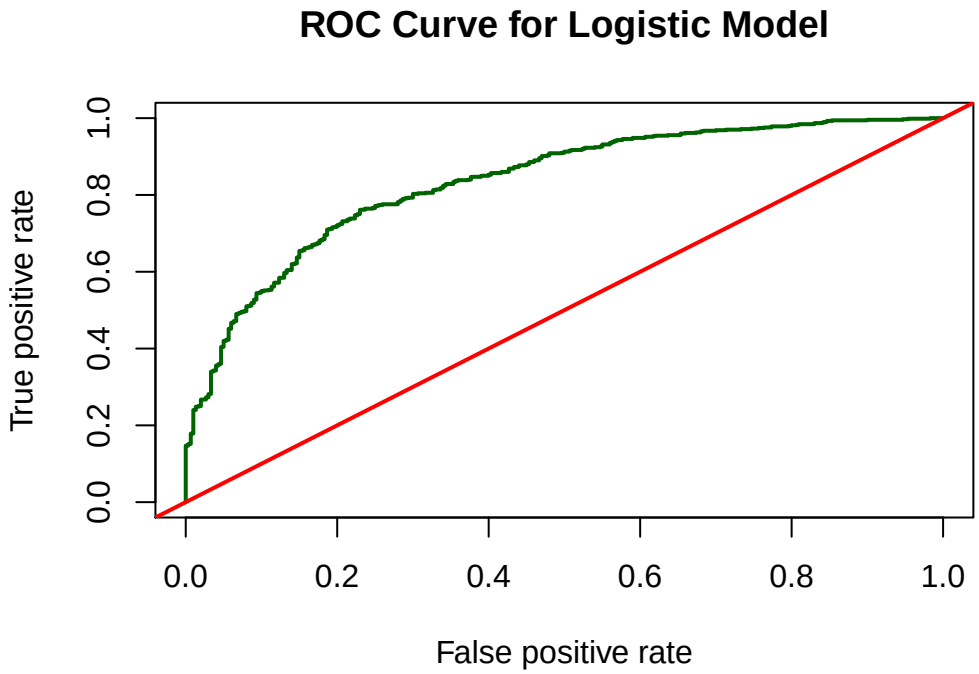
The AUC metric can also be used to compare different classification algorithms with the one with a large AUC can be judged to perform better. As a matter of illustration we can fit a

logistic regression model and construct its ROC curve. Using the German credit data we fit a model on all predictors.

```
GermanCredit$Class <- as.factor(GermanCredit$Class)
log2 <- glm(Class ~., family = "binomial", data=GermanCredit)
predlog <- predict(log2, type="response", GermanCredit)
```

The library `RROC` supports a number of functions that allow the plotting of ROC curves. The two most important of them are `prediction` and `performance` and their operation is illustrated in the following code snippet.

```
library(ROCR)
pred <- prediction(predlog, GermanCredit$Class)
perf <- performance(pred, "tpr", "fpr")
plot(perf, lwd=2, main = "ROC Curve for Logistic Model",
     col = "darkgreen") + abline(0,1, col = "red", lwd=2)
```




```
## integer(0)
```

The green line corresponds to the predictive quality of the model at different cutoffs, and the improvement over random prediction is summarized by the area between the green ROC curve and then red bisector. The larger this area, the better performing the model is. Such a tool allows the refinement of the credit risk model so that its prediction bias parallels the relative cost of true against false positives.

In the credit risk instance the positive rate may be defined as those defaulting on their loans. In this case the true positives are correctly identified debtors that will default and thus

represent avoided losses for the bank. On the other hand, the false positives are incorrectly identified people who were going to pay back but were rejected a loan (they are foregone profit). As losses are likely more painful that foregone profit, the bank may slightly skew its classification modeling to be biased to more easily classify applicants as positives by choosing an appropriate cutoff.

Big Data Implications

A recent trend in the data analytic and risk management landscape is the exponential increase of available data. We have noted that usually the quality and quantity of data has greater impact on model efficiency than the statistical method used. In that sense, the data deluge seems to be good news. The RM professional can now leverage much more information to make informed decisions on how to optimize returns and exposure.

However this era of big data poses a number of challenges on its own. Among them we should note the following:

- **Advanced data skills set** - the availability of more data calls for an advanced and varied skill set, ranging from database management tools, data processing, analytic and visualization software. The RM and data analyst have to be equipped with wide knowledge of technologies and ability to use concrete tools (e.g. SQL databases, NoSQL databases, Hadoop, Spark, R/Revolutionary Analytics, and many others).
- **Need for newer statistical methods** - in addition to data handling tools, knowledge of novel statistical approaches is an absolute must. Traditional methods often fail to scale well to large quantities of data and well-known metrics such as p-values lose their meaning. Machine learning approaches such as those presented here will tend to dominate the analytic landscape in the foreseeable future. In addition to that, parallel computation increases rapidly in importance.
- **Appreciation of possibilities and limitations** - the wealth of abundant, rich, and often real-time information raises the paradox of choice. The RM analyst needs to understand both the power of data and its limitations and how it applies to the organization she works for. This adds a new dimension to quantitative skills - the business acumen needed to apply them with maximum impact.
- **Ability to glean and communicate actionable insight** - finally the ability to communicate and convince people of the analysis and spur concrete action is vital for the success of any risk management enterprise, and this is more so as the amount of data increases.

In conclusion, we can say that the age of big data opens significant opportunities for the RM professional to bring true value to organizations but also that the challenges for that are not trivial.

A key aspect of managing risk is now understanding data and its peculiarities, mastering methods and appreciating their limitations and judiciously building a story around

organizational risk that will be compelling enough to evoke concrete action.

Project 10

Select a suitable data set for credit risk. If you wish, you can also use the classic German Credit Data, or something similar from the UCI Machine Learning Repository. Do the following:

- Split it up into two subsets - one for training and one for testing (the `createDataPartition` command from the `caret` package may be useful).
- Train at least four classification models on the training set.
- Predict new cases using the test set data.
- Compare accuracies and select best model.
- Investigate variable importance.

Discuss on the classification exercise. What are the business implications and what advise would you give to your organization based on the results.

Lecture Eleven: Black Swans and Forecasting

Forming expectations is a crucial component of the risk management process. Many organizations have to formulate concrete numbers about the future realization of a given variable of interest. This can be sales, profit, growth rates, economic developments, and many others. Such forecasting is used to aid strategic planning, to guide tactical decisions, and to aid risk reduction by forming more informed choices.

Here we look at how time series of a given variable of interest can be used to generate a forecast of its future developments. We first survey time series decomposition to understand its drivers, and then proceed to present two major classes of time series models - AR models, and VAR models. Finally, their implications and limitations are outlined. For a more detailed overview, the reader is directed to Zivot and Wang (2006).

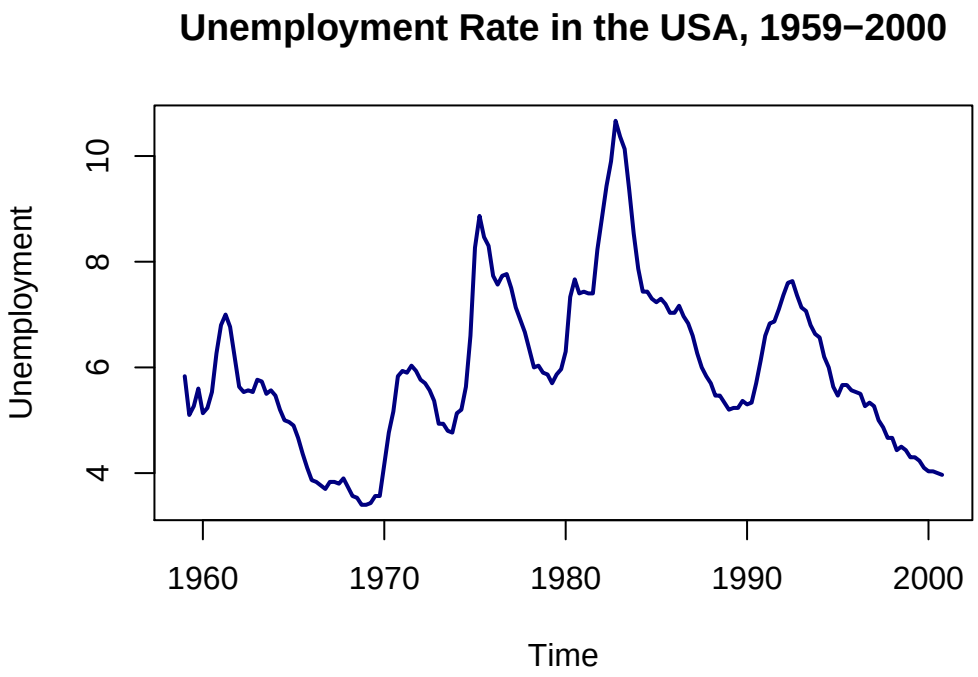
Structure of Time Series

A given time series reflects the realizations of a data-generating process over the temporal dimension. For example, this can be the data on stock market (or asset) returns, economic growth, inflation, company sales, insurance events, etc. The overall realization we observe is composed of several separate drivers. It is possible to distinguish between the following large groups:

- **Trend Component** - the trend of a time series reflects the key underlying force that drives the process. It is often the result of some economic or social fundamentals. E.g the trend in economic growth rates is due to the increases in fundamental production factors - labor, capital, and knowledge, and reflects potential output.
- **Cyclical Component** - there are some cyclical fluctuations around the trend. This can be due to either anthropogenic factor (like the business cycle) or natural causes (seasonality). For example, in the case of employment there is usually an upsurge in summer due to seasonal jobs that then disappear in autumn.
- **Random Disturbance** - there are random shocks that are not due to any systematic reason but nevertheless affect the realization of the variable. E.g. the Arab Spring is a random disturbance with a large effect on oil prices.

To better understand the structure of a time series, we load unemployment data for the US at quarterly frequency over the period Q1 1959 to Q4 2000. Its graph follows.

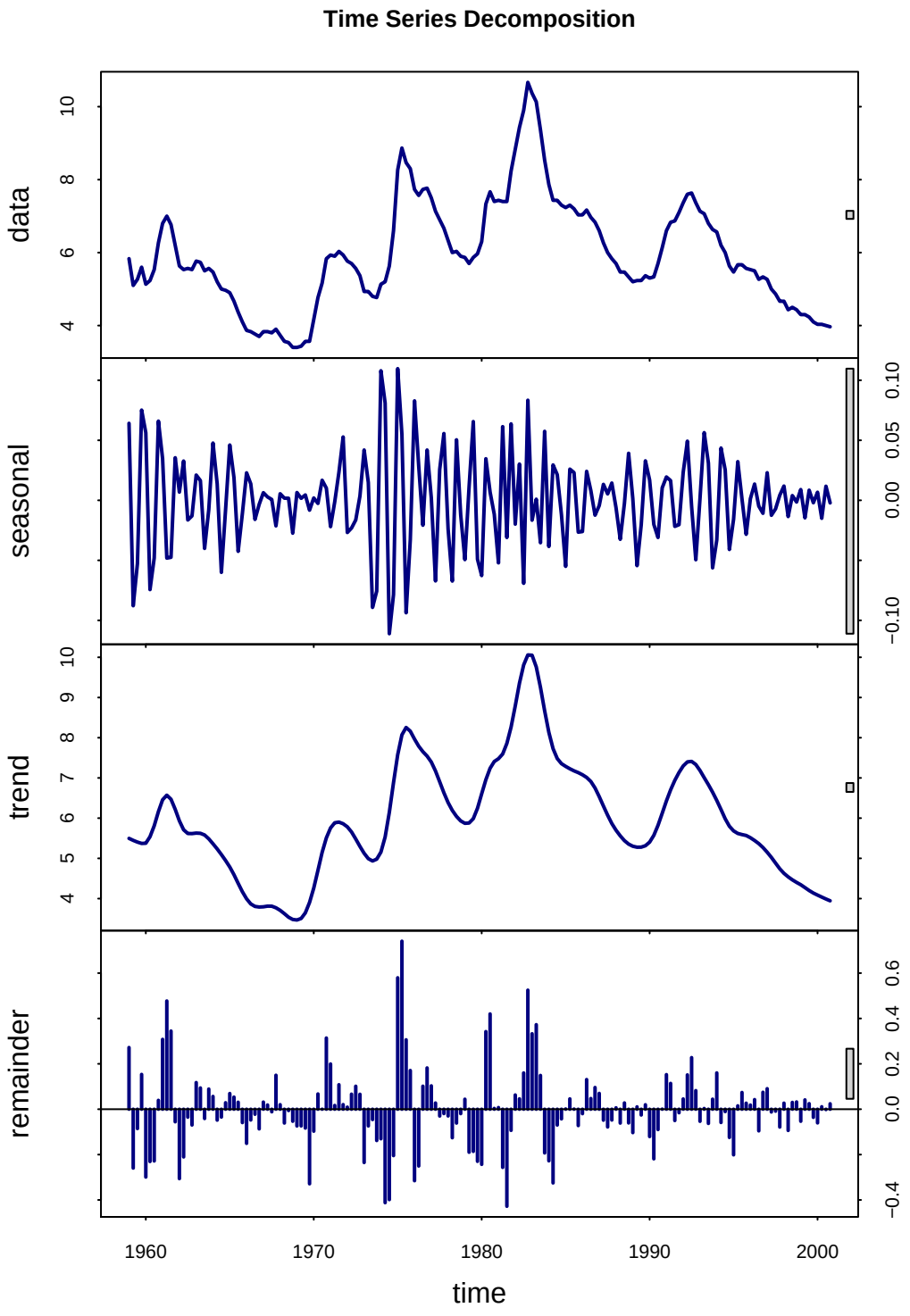
```
library(mFilter)
data(unemp)
plot(unemp, ylab="Unemployment", col="navyblue", lwd=2,
     main="Unemployment Rate in the USA, 1959-2000")
```



The R language offer many utilities for time series decomposition and we demonstrate one of them, `stl`, here. The decomposition shows a typical picture. The trend in unemployment changes with the trend growth in the economy. As overall growth slows, less workers are needed and unemployment rises. The oil crises of the 1970s and the ensuing inflation into the 1980s increases the trend unemployment, which only decreases as the US economy recovers and enters a period of sustained growth.

The seasonal component reflects the seasonality of job creation and is strongly driven by temporary jobs in tourism in the summer and winter, and temporary jobs in agriculture during summer. This seasonal pattern leads to the picture we observe. Finally the remainder in the realization of the time series is a random disturbance of very small (even empirically negligible) magnitude.

```
dec <- stl(unemp, s.window=7)
plot(dec, col="navyblue", lwd=2, main="Time Series Decomposition")
```



It is sometimes of use to distinguish between the trend and the non-trend component in time series. The trend may refer to some sort of long-run potential (e.g. potential output), while the fluctuations may be cyclical disturbances. A popular approach to do this is to use a filter on the time series. The Hodrick-Prescott filter is still a popular choice. If we assume that a

time series y_t is composed of a trend τ_t component, a cyclical component c_t and an error term η_t, then:

$$y_t = \tau_t + c_t + \eta_t$$

The trend component can be found by solving the following equation:

$$min(\sum_{t=1}^{T}(y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1}[(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2)$$

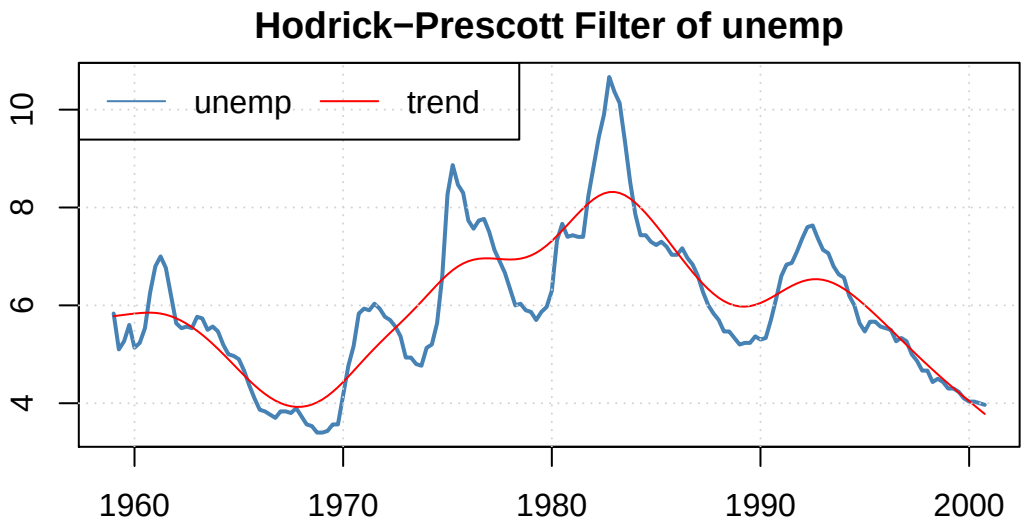
Here λ is a tuning parameter which should vary by the fourth power of the frequency of observation. More specifically, values are presented in the following table.

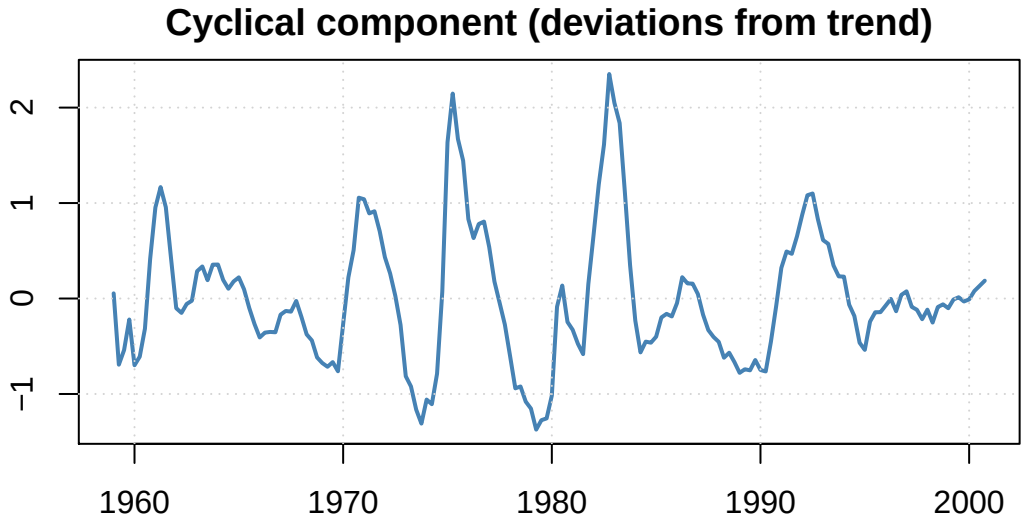
Observation Frequency	Value of λ
Monthly	129,600
Quarterly	1600
Annually	6.25

We use the unemployment data to illustrate the Hodrick-Prescott filter. In R it is implemented via the `hpfilter` command in the `mFilter` package.

In the graphs we see the typical response as the smoothed trend component is displayed next to data, and the cyclical component captures the fluctuations around it. We observe that here we lose the informational value of the purely seasonal component which may be important for some applications. The overall story remains the same, as unemployment trend tracks overall economic potential, and crises are characterized by large deviations.

```
hp <- hpfilter(unemp)
plot(hp, lwd=2)
```





Despite its popularity, we should note a few limitations of the Hodrick-Prescott filter:

- **Non-causality** - the smoothed trend depends on past *and* future observations, and as more observations become available, it will change (e.g. the HP calculates some value for time t, but as $t+1$ observation becomes available, it will be revised)
- **Instability near end of interval** - due to the non-causality, the last few points may not be perfectly accurate
- **Parametrization need** - the λ parameter guides the practical solution but its exact values are not uncontroversial and not theoretically-informed.

Despite this, the HP filter may be useful an an approximation and guide for the time series structure, that can be used for RM purposes.

Forecasting with ARIMA models

It is often the case that the RM analyst only has historical data for one variable of interest but would still like to forecast it. The assumption in this case is that the structure of the time series already contains information on the variables that influence it - i.e. past realizations of the variable and its trends capture the effects of key process drivers. This can possibly be exploited to obtain a forecast of future realizations. The simplest version of this process is to fit an auto-regressive model, whereby the variable of interest y_t is regressed on its lagged values - y_{t-1}, y_{t-2}, y_{t-3}, $\dots$ y_{t-p}, thus obtaining the so-called **AR(p)** model:

$$y_t = \theta + \sum_{i=1}^{p} \beta_i y_{t-i} + \varepsilon_t$$

Apart from the information that is contained in lags or the variable, there may be some more additional information that can be gleaned from the structure of errors. A model of this class - **MA(q)** uses q lags of the error term to capture y_t dynamics. Thus we obtain the following form:

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-q}$$

We can combine those models to obtain a fuller description of the time series, thus reaching the **ARMA(p,q)** model.

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-q}$$

If the given time series is integrated of some order d, we can make a correction for that and finally obtain the **ARIMA(p,d,q)**. This class of models is very popular and easy to use as it requires only historical data on y and the appropriate selection of the lag length for both the variable and the error terms. This can be easily done by taking recourse to some of the information criteria that are routinely calculated alongside the model in modern statistical and econometric packages. The most common ones are the AIC (Akaike Information Criterion), and BIC (Bayesian Information Criterion), and the SIC (Schwartz Information Criterion). Better models tend to produce lower values on those criteria but the RM professional should keep in mind that these are not always in agreement and thus make informed choices.

In terms of software implementation excellent ARIMA and time series facilities are common in both proprietary and open source packages. The classic R command is the `arima` one from the base package. We fit an ARIMA(1,0,1) model on the `unemp` data:

```
ar <- arima(unemp, order=c(1,0,1))
ar
```

```
##
## Call:
## arima(x = unemp, order = c(1, 0, 1))
##
## Coefficients:
##          ar1     ma1  intercept
##       0.9574  0.5963     5.7628
## s.e.  0.0210  0.0583     0.6964
##
## sigma^2 estimated as 0.07284:  log likelihood = -20.26,  aic = 48.51
```

The output gives the ar1, ma1 and intercept component allowing us to reconstruct the equation and use it for forecasting. Their standard errors show the significance of all the estimates. We thus obtain the following model:

$$unemp_t = 5.763 + 0.957 * unemp_{t-1} + 0.596 * \varepsilon_{t-1}$$

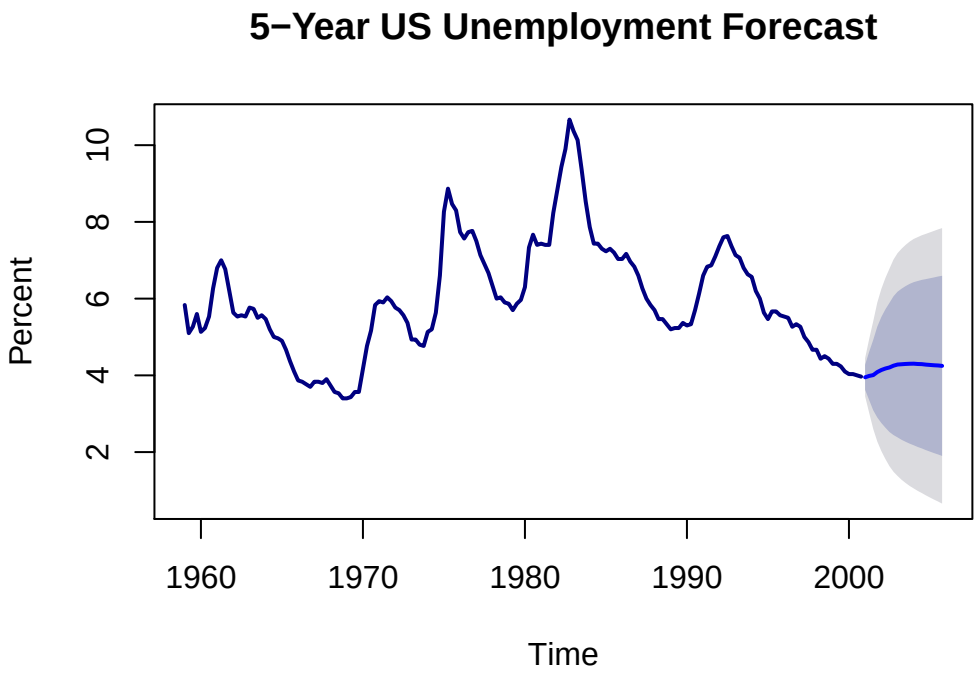
It is likely that this model is not the best performing one. The analyst can experiment with different specifications to see which fits data best and decide accordingly, or simply use the value of the AIC to choose. The general rule is that more parsimonious models are preferred. Alternatively, there are software utilities that can choose the most optimal model based on information criteria. A useful implementation is `auto.arima` from the R `forecast` package. Its application is illustrated here.

```
library(forecast)
ar <- auto.arima(unemp)
ar
```

```
## Series: unemp
## ARIMA(2,1,0)(2,0,2)[4]
##
## Coefficients:
##          ar1      ar2    sar1     sar2     sma1    sma2
##       0.6475  -0.0213  0.7243  -0.3219  -0.9832  0.2685
## s.e.  0.0802   0.0880  0.5790   0.2931   0.5959  0.4736
##
## sigma^2 estimated as 0.06494:  log likelihood=-6.43
## AIC=26.86   AICc=27.57   BIC=48.69
```

The best fitting model turns out to be an ARIMA(2,1,0) with a strong seasonal component of (2,0,2). This model seems to be better fit for the data at hand, as evidenced by the lower AIC - it has gone down from 48.5 to 26.7. This model can be used for forecasting through the `forecast` command. We estimate and plot the 20-period (or 5 years) ahead forecast.

```
fore <- forecast(ar, h=20)
plot(fore, lwd=2, col="navyblue", xlab="Time", ylab="Percent",
     main = "5-Year US Unemployment Forecast")
```



Overall, we can observe two major trends:

- **Smoothing** - the model captures the overall trend and produces a smooth forecast for unemployment that fluctuates far less than the actual realizations.
- **Increasing Risk** - as the periods ahead increase so do the 95% and the 99% confidence intervals, and by the end of the five forecasted years the 99% confidence interval says that unemployment will lie between 1% and 8%. Such precision has hardly any practical value.

Forecasting with VAR models

Vector Auto-regressive (VAR) models are an extension of classical auto-regressive time series analysis. Here the main idea is that the system under analysis is composed of interacting variables and the lags of one have important implications for the realizations of the others. For instance, in the economy the output (GDP), inflation (HICP), and unemployment strongly influence each other (through the augmented Phillips curve and the Okun law).

More formally, the VAR model is a system of equations with lagged variables that captures interdependencies between the variables. The simplest VAR model is the VAR(1) model, where current realizations depend on only one lag of each factor. If we have three variables - x, y, z, and denote with ϕ_{ij} model coefficients, then the VAR model is of the following form:

$$x_t = c_1 + \phi_{11} x_{t-1} + \phi_{12} y_{t-1} + \phi_{13} z_{t-1} + \varepsilon_1$$
$$y_t = c_1 + \phi_{21} x_{t-1} + \phi_{22} y_{t-1} + \phi_{13} z_{2-1} + \varepsilon_2$$
$$z_t = c_1 + \phi_{31} x_{t-1} + \phi_{32} y_{t-1} + \phi_{33} z_{t-1} + \varepsilon_3$$

The lag length selection is naturally one of the most important parameters in VAR estimation. This can be done either manually by investigating the fit of different VAR models to data, and their respective information criteria, or by automatic selection. While it is often useful for the analyst to spend more time understanding and appreciating data and alternative model specifications, quick analyses sometimes call for a rapid solution.

To illustrate the use of VAR we will use the `vars` packages, and thus load the `Canada` data from it.

```
library(vars)
data(Canada)
describe(Canada, skew=F)
```

```
##      vars  n   mean    sd    min    max range   se
## e       1 84 944.26  9.16 928.56 961.77 33.20 1.00
## prod    2 84 407.82  4.22 401.31 418.00 16.70 0.46
## rw      3 84 440.75 23.27 386.14 470.01 83.88 2.54
## U       4 84   9.32  1.61   6.70  12.77  6.07 0.18
```

Here we have data on labor productivity (`prod`), employment (`e`), real wage (`rw`) and unemployment (`U`). To fit a VAR model we first use the `VARselect` command from the `vars` package.

```
VARselect(Canada)
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      3      2      1      3
##
## $criteria
##                  1            2            3            4           5
## AIC(n) -6.191599834 -6.621627919 -6.709002047 -6.512701777 -6.30174681
## HQ(n)  -5.943189052 -6.174488511 -6.063134014 -5.668105118 -5.25842152
## SC(n)  -5.568879538 -5.500731387 -5.089929279 -4.395452772 -3.68632157
## FPE(n)  0.002048239  0.001337721  0.001237985  0.001534875  0.00195439
##                  6            7            8            9           10
## AIC(n) -6.194596715 -6.011720944 -6.054479536 -5.912126222 -5.867271844
## HQ(n)  -4.952542805 -4.570938409 -4.414968375 -4.073886435 -3.830303432
## SC(n)  -3.080995238 -2.399943231 -1.944525586 -1.303996035 -0.760965421
## FPE(n)  0.002278812  0.002924622  0.003073249  0.004015164  0.004961704
```

It turns out that in two of the four selection criteria (AIC and FPE), the recommended lag length is 3, one recommends 2, and one - 1. Due to the quarterly character of data it may be useful to opt for a somewhat longer lag length and thus we fit a VAR(3) model with both a constant and a trend.

```
var <- VAR(Canada, p = 3, type = "both")
var
```

```
##
## VAR Estimation Results:
## ========================
##
## Estimated coefficients for equation e:
## ======================================
## Call:
## e = e.l1 + prod.l1 + rw.l1 + U.l1 + e.l2 + prod.l2 + rw.l2 + U.l2 + e.l3 + prod.l3 +
##
##          e.l1       prod.l1         rw.l1          U.l1          e.l2
##    1.76382827    0.18518258   -0.07242428    0.12194306   -1.19012450
##       prod.l2         rw.l2          U.l2          e.l3       prod.l3
##   -0.10936038   -0.02488409   -0.03229344    0.61470608    0.02564596
##         rw.l3          U.l3         const         trend
##    0.03172493    0.35975734 -193.37043794   -0.01741357
##
##
## Estimated coefficients for equation prod:
## =========================================
## Call:
## prod = e.l1 + prod.l1 + rw.l1 + U.l1 + e.l2 + prod.l2 + rw.l2 + U.l2 + e.l3 + prod.l3
##
##         e.l1      prod.l1        rw.l1         U.l1         e.l2
##  -0.19622759   1.08138761  -0.01995026  -0.75409912  -0.15481450
##      prod.l2        rw.l2         U.l2         e.l3      prod.l3
##  -0.18078644  -0.20123648   0.74328605   0.45746821  -0.02050056
##        rw.l3         U.l3        const        trend
##   0.12123944   0.32217926 -13.21924425   0.07451666
##
##
## Estimated coefficients for equation rw:
## =======================================
## Call:
## rw = e.l1 + prod.l1 + rw.l1 + U.l1 + e.l2 + prod.l2 + rw.l2 + U.l2 + e.l3 + prod.l3 +
##
##         e.l1      prod.l1        rw.l1         U.l1         e.l2
##  -0.52468461  -0.13954289   0.86031246  -0.10820098   0.69673993
```

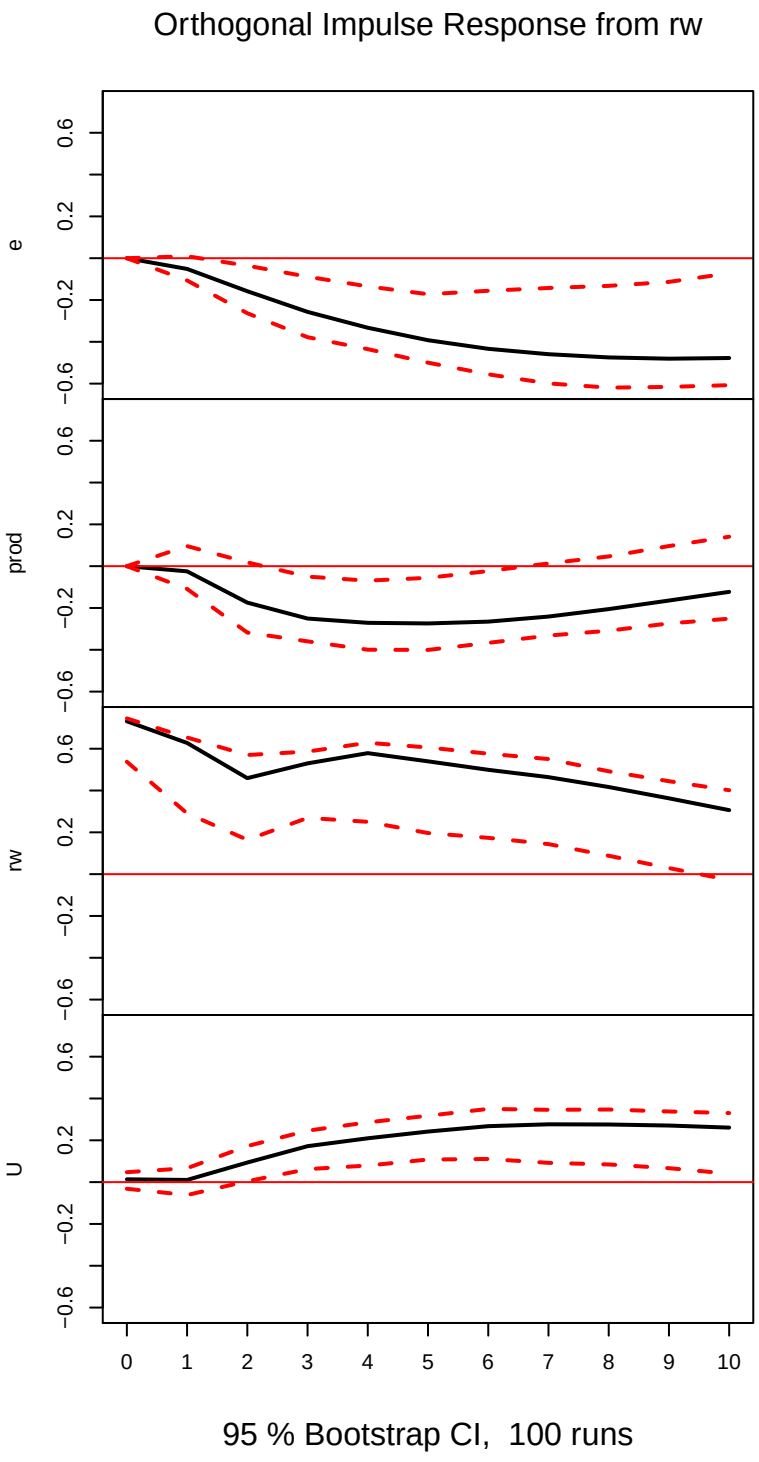
```
##      prod.l2        rw.l2         U.l2         e.l3      prod.l3
##  -0.19911318  -0.14326894  -0.38989137  -0.26039495   0.14159721
##        rw.l3         U.l3        const        trend
##   0.22136045   0.06263254 192.77720875   0.08340852
##
##
## Estimated coefficients for equation U:
## =====================================
## Call:
## U = e.l1 + prod.l1 + rw.l1 + U.l1 + e.l2 + prod.l2 + rw.l2 + U.l2 + e.l3 + prod.l3 +
##
##          e.l1       prod.l1         rw.l1          U.l1          e.l2
##  -0.630593919  -0.115838301   0.002742498   0.633745966   0.525389579
##       prod.l2         rw.l2          U.l2          e.l3       prod.l3
##   0.092189116   0.070509144  -0.102428066  -0.061912900  -0.028450268
##         rw.l3          U.l3         const         trend
##  -0.031534562   0.045531567 163.889696250   0.020203832
```

The results presented show the estimated VAR coefficients using ordinary least squares (OLS). The RM professional should keep in mind that VAR results are not always intuitive to interpret especially as lag length becomes larger. This is why such models are most often used for two tasks - Impulse Responses and forecasting. The impulse response function (command `irf`) show how one of the variables changes with a change of one unit of another variable. It is thus a measure of the effect of a shock over time.

For illustration here we present the effect of a change in the real wage (rw) in this system. The real wage increase negatively affects employment, and leads to a decrease in employed individuals. Conversely, it leads to a hike in unemployment. Productivity is initially unaffected by after half an year (2 periods) it also reacts negatively to the wage dynamics. Under the force of inertia, the real wage continues to increase but this volatility slowly dies down.

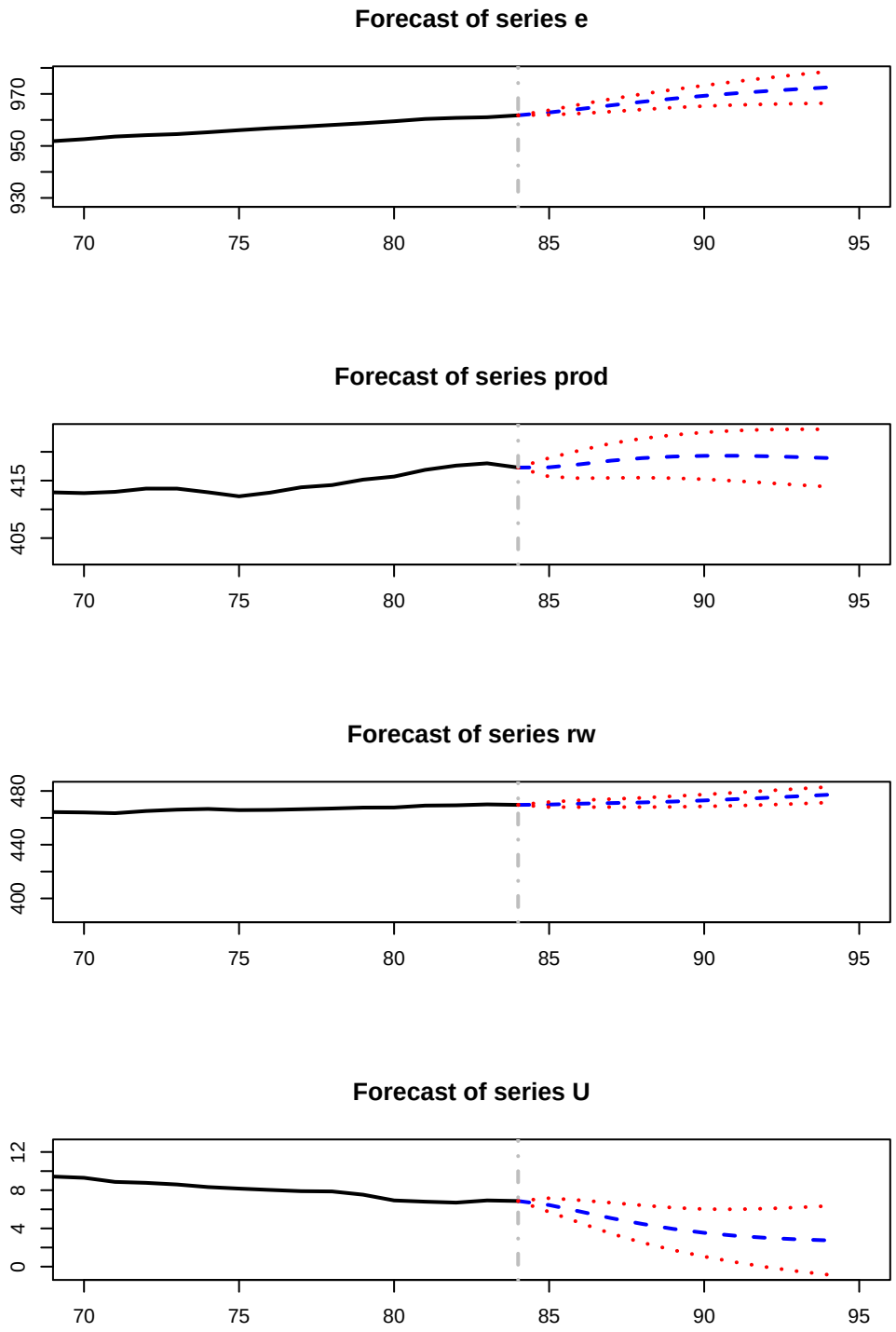
The plot presents the typical picture of initial strong effect of a shock in one variable which slowly decreases over time. Such a VAR system can be used for simulations for the benefit of risk management. The RM analyst may see the effects of a given shock on other variables he might be interested. For example, she can model a stock market index and overall economic growth and estimate how returns will be affected as the economy slows down or heats up.

```
irf <- irf(x = var, impulse="rw")
plot(irf, lwd=2)
```



Another key application for VAR modelling lies in the field of forecasting. We forecast 30 periods ahead for all variables and plot the results.

```
pred <- predict(var, n.periods=30, ci=0.99)
plot(pred, lwd=2, xlim=c(70,95))
```



Again we see the familiar pattern from the simpler auto-regressive models - the trend is

smoothed and the more forward we move into the future, the large the confidence interval for our estimates becomes. The graph "fans out", pointing at the rapid increase of uncertainty as we move away from actual data. In this data set this is even more pronounced as the variables are in log form, and their actual change will be equal to e to the power of the change we see here.

Limitations of Historical Data

Forecasting exercises are largely focused on deriving information from past data and then using it to try and predict future realizations. This is often helpful but the analyst should keep some important caveats in mind:

- **Historical data is a sample** - what we have as historical data is merely a sample of all possible realizations and does not need to include all possibilities - i.e. there is likelihood of an unexpected "black swan" event. Prior to 1929 data could not show that an event such as the Great Depression was possible, and prior to 2008 data did not indicate it could happen as often as twice in a single century. The future, therefore, does not need to repeat the past and may well surprise.
- **Parameters are not time invariant** - as complex social and economic systems change, so may the behavior of agents, thus producing very different quantitative effects as a result of similar shocks. For example, a decrease of 2% of the central banks rates now have a much more profound effect on the economy than a similar decrease 150 years ago. Under currency board, agents in Bulgaria react less strongly to fluctuations of the dollar rate as they can enjoy a fixed euro rate vis-a-vis the Bulgarian lev. Such structural breaks or regime shifts where behavior changes need to be accounted for.
- **Exploiting forecasts may ruin them** - this is also known as the famous Lucas critique. Robert Lucas claimed that as we exploit a forecast or a model people will realize that and change their behavior accordingly. For example if a risk model forecasts a high risk, organization management may take action and reduce it, thus making model predictions obsolete.
- **Increasing uncertainty** - as we have seen, the farther away we forecast, the less certain this forecast is. Intuitively, the farther into the future we look, the less we are attached to real data and parameters and assumptions start to dominate the model. Even little errors with them compound exponentially, leading to essentially meaningless numbers over the long-term.

Knowing all this, the RM professional should focus attention on short-term to mid-term actionable forecasts, where data trends and information are still meaningful. While long-run trends may be useful for strategic planning, their numeric accuracy is quite uncertain. On the other hand, short-run forecasts over the 1-3 years period can provide a relatively accurate description of trends and feed into tactical decision-making. In all likelihood, the shorter the forecasting window, the higher the accuracy.

Thus, an important part of the risk management process is to generate forecasts that give insight but do not mislead. For this purpose, the RM professional needs to pick a good model, construct it carefully using quality data and outline clear business insight. Most importantly, the model cannot show something that is not found in the data, which is why the RM analyst should guard against both the expected, and the unexpected.

Project 11

Pick a time series of interest that has risk management implications. Do the following:

- Summarize and investigate it.
- Make a decomposition of its drivers. Discuss each of them in turn. What is the business implication behind those?
- Fit an auto-regressive model of ARIMA type. Reconstruct the equation.
- Forecast the time series for the next two years. Comment on results.

How can this information be useful to an organization? What risk management recommendations can you make based on these results?

Lecture Twelve: Modelling Risk with Risky Models

A large part of the risk management process consists of trying to define risk and measure its potential impact in order to devise a risk mitigation strategy. This is greatly aided by formal models but the very presence of those induces additional considerations, known as model risk. We first review what a model risk is, what are its possible drivers, and finally conclude with a process for continuous model improvement in order to manage that. Here we follow ideas from Crouhy et al. (2006) and Tibshirani et al. (2011).

Model Risks

The famous statistician George Box once noted that "all models are wrong but some are useful". This comes to underline the fact that models are (sometimes a significant) simplification of reality and by design do not capture all the sophistication and complexity of real life. Some are so simple that their key messages are useless, while other can serve as a guide to reality. In that sense all models generate some risk which the RM professional should keep in mind.

Model risk is present whenever a model generates wrong or misleading results that may bias a decision and lead to a loss. It is generally composed of two large sub-groups of risks, namely:

- *Model Faults* - the model used is inappropriate for the situation at hand and will fail to produce meaningful results.
- *Implementation Problems* - the model is adequate but is poorly executed leading to erroneous output.

Both groups of risks may have critical implications and thus we review each of them in turn.

Model Faults

Model faults usually stem from wrongly specified models. Sometimes despite the best efforts of the analyst data is not enough to reach a conclusive decision on the most appropriate model and a judgment call has to be made. While experienced RM professionals will have an edge in this, at least some mistakes are unavoidable.

Common sources of problems include:

- **Wrong assumptions** - the analyst may make assumptions that not borne out in practice. A particularly common pitfall is to assume normality of distribution for some variable, while in fact this is not the case. This will skew probabilities and make extreme (and very risky) events seem less common that they are. Another pitfall is to assume time invariance of parameters - that a particular parameter of interest (e.g. the market beta) does not change over time. However, with structural breaks and regime

shifts, a change will likely occur. We need to keep in mind that the model only operates well within the confines of its assumptions and is bound to produce wrong results outside them.

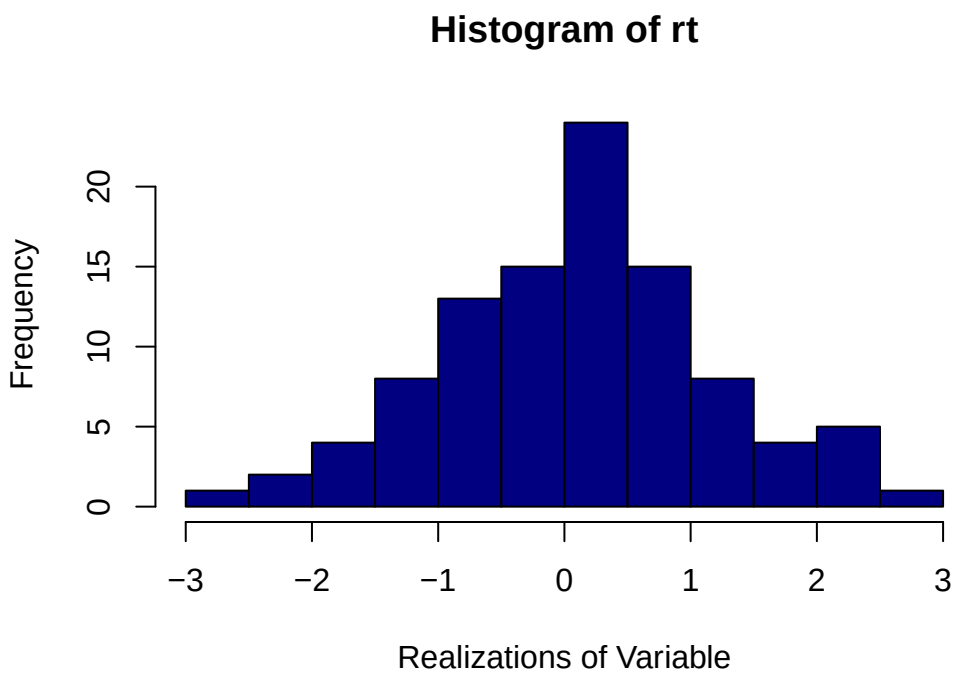
- **Wrong analytic solutions** - even if the model itself is right, its analytic solution may be wrong. This is particularly true for more complex models which need to be solved through approximations and introduction of some assumptions. The problem of wrong analytic solution can be mitigated by complementing them with alternative solutions achieved through numerical methods. Those two should converge, and if not - there is likely some issue in need of further attention.

To illustrate a possible model misspecification we will now draw 100 random numbers from a central Student T distribution with 10 degrees of freedom. Since we do know the generating process, we are pretty certain of the underlying distribution. We store those random numbers in the `rt` object.

```
set.seed(1234)
rt <- rt(100, df = 10)
```

Now imagine we did not know the distribution and need to find it instead. We plot the data on the next plot, thus obtaining:

```
hist(rt, col="navyblue", xlab="Realizations of Variable")
```



This histogram looks very much like a normal distribution. We test the data formally if it is Gaussian using the Jarque-Bera test.

```
library(tseries)
jarque.bera.test(rt)
```

```
##
##  Jarque Bera Test
##
## data:  rt
## X-squared = 0.19438, df = 2, p-value = 0.9074
```

The results are very pronounced: we fail to reject the null hypothesis of normality with a p-value of 0.91. Note that the significance level needed to reject a Gaussian distribution is $p < 0.1$, so our result does not even approach the boundary of doubt. Given the outlook of data, and the results of testing, we conclude that the best model for this data is the normal distribution, and proceed to fit its parameters.

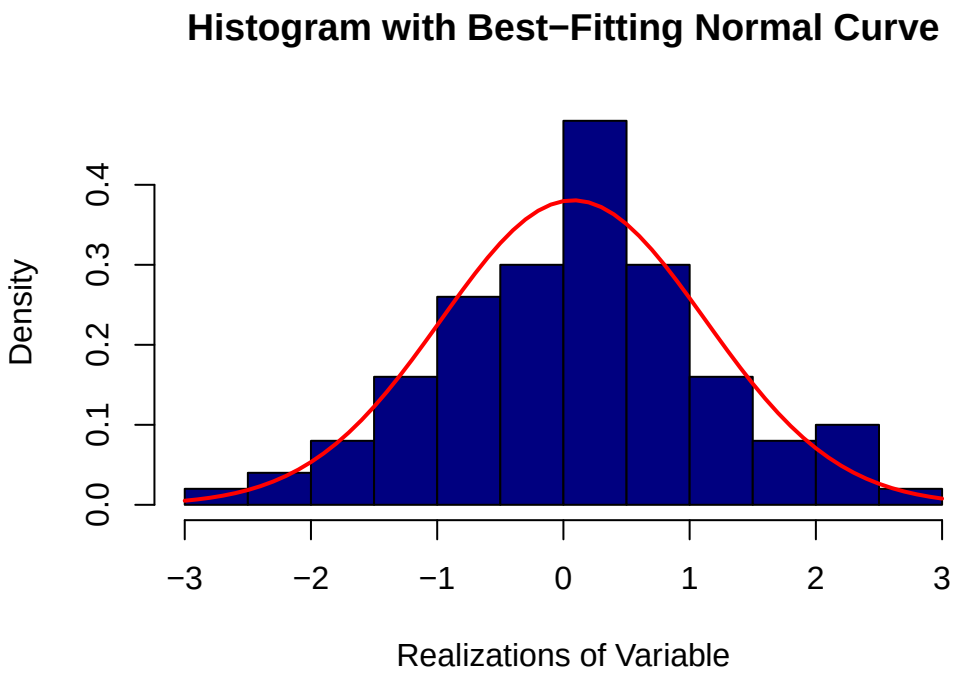
```
library(MASS)
fitdistr(x = rt, densfun = "normal")
```

```
##       mean          sd
##   0.07709422   1.04801193
##  (0.10480119) (0.07410563)
```

```
x <- seq(from = -3, to = 3, by = 0.1)
norm <- dnorm(x, mean = 0.07709422, sd=1.04801193)
```

Then we superimpose the fitted distribution and observe that it corresponds to data quite well (although not perfectly). In all probability, the RM analyst may decide that modeling data as normally distributed would be appropriate in this case.

```
hist(rt, col="navyblue", xlab="Realizations of Variable",
     prob=TRUE, main="Histogram with Best-Fitting Normal Curve")
lines(x, norm, lwd=2, col="red")
```



However we do know that in fact this is not a normal but a fat-tailed distribution. Assuming normality will underestimate extreme outcomes to the peril of the risk management recommendation. This example serves to illustrate that even clear-cut statistical conclusion may turn out wrong in the face of insufficient data or strong bias for a given approach. Practical situations are even more fraught with difficulty and enlarge the scope for both decisions and mishaps. All in all, model simplification and tractability often comes at the price of introducing model faults which at points may make the model useless. It is the RM professional's job to guard against this.

Implementation Problems

Even if the model is appropriately selected and defined, there are many possible issues that can arise during its implementation phase. Problems during the process can sometimes pose significant risk by skewing results and leading to suboptimal decisions. There are a few large types of possible risks:

- **Misspecification** - as the model is being specified, there can be an error, an addition or an omission, which could lead to meaningless results. While the business sense should be clear to the analyst, it is a mystery to the machine, which will calculate the given specification regardless of how meaningless it might be. This is especially true for large-scale simulation models which are sometimes used in the valuation of more exotic instruments such as options or derivatives.
- **Computation Issues** - during the computation phase there could be errors that

change output significantly. Of particular note is the fact that some more complex algorithms use approximations, look for local optima, or depend on random numbers to start a process. One or all of those factors can lead to significant discrepancies. Alternatively, sometimes computing power may just not be enough for the model at hand - something particularly true in the era of big data.

- **Inaccurate Data** - probably the most common implementation problem hinges on the quality of the data at hand. Data may be incomplete, inaccurate, or not representative. All of those will lead the RM professional to form an erroneous view of the characteristics of the general population, thus making even correct models to yield grossly misleading conclusions. Time spent on ensuring high data quality is extremely important as quality of data will drive model quality in a much more pronounced way than the application of sophisticated statistical methods.
- **Improper Period** - a common issue whereby the analyst uses a sub-sample of the data from an improper or not fully representative period. While the calculations will be correct for this specific period, they may fail to reflect the general population. A common critique of the VaR models during the 2007-2011 financial crisis was that they used only recent data from the last decade which did not include significant economic downturns, thus underestimating their probability.
- **Incorrect Interpretation** - finally, model results need to be interpreted with due care in a way to present a honest and understandable depiction of the main model conclusions and their implications. For instance a 95% VaR is the number that is the maximum expected loss in 95% of the cases if the market behaves normally, and not the loss in 95% of all cases. The burden of providing a correct and actionable interpretation clearly lies on the RM analyst.

Implementation problems are very often a case of human error and should therefore be viewed less in terms of technology and infrastructure, and more in terms of people and their skills.

Continous Improvement

The way to minimize model risks is to create efficient model building processes that automate manual tasks and have a specific and clear focus on continuous model improvement. As new information on model performance is generated, and new data arrives, this presents an excellent opportunity to review and improve the model in order to better fit the risk management needs of the organizations.

We will particularly note four important steps in this process:

- **Check Data** - data needs to be investigated in order to ensure consistency, accuracy and quality. As new data comes in, there might be revisions of older values (particularly true for macroeconomic data). The analyst may want to check for missing values (in R the command `is.na` can be used), and also for outliers. They may

represent true properties of the data or some typing errors. Also, the RM professional may want to check the continuity of trends.

- **Check Model** - a part of the improvement process is to re-evaluate model soundness, validity and fit. This can be done either by a review of the model itself or through using benchmarks and best practices. The analyst may also experiment by including new variables or changing specifications in order to see whether improved validity and fit to data can be obtained. Model performance indicators are crucial measures of model risks, and confidence intervals, mean squared errors, and information criteria should be carefully investigated.
- **Check Output** - an important step in the improvement process is to test results against benchmark values and business logic. Results need to be better than viable alternatives (e.g. forecasting models should outperform commercially available or naive forecasts) and to conform to business logic. If a model outputs results that are nearly impossible to materialize, then it is largely at odds with business common sense, and ultimately, needs.
- **Improve** - once the input, model, and output are reviewed, the analyst will need to prescribe improvement steps and implement them, thus building a second generation model. Those may range from the very small fine-tuning of a parameter or two, through inclusion of new variables or specifications, to a large-scale overhaul of the model and construction of a novel one. We should note that the new model is supposed to bring *sustainable improvement* over the old one, and this should be carefully monitored. After the updated model is operation, the continuous improvement cycle can begin again.

While models tend to be imperfect by construction, a process that is explicitly focused on their continuous and sustainable improvement will make them better serve the organization's needs and help them unlock more value as a key decision-support tool.

Discussion

The very presence of models may generate risks. They may lead to comfort with erroneous results, thus skewing decision-making and creating substantial loss. This is even exacerbated by the workings of of the human brain. From a psychological standpoint experts tend to exhibit over-confidence, only to be boosted by misleading model output.

This necessitates critical attitude to models, leading to continuous efforts at improvement. As there is no perfect model, the room for updates will always be there. The cycle of continuous improvement needs to focus explicitly on generating performance boosts that exceed the cost of tinkering with the model. Essentially, this means that the improvement cycle will be uneven over the model lifetime. Initially, as large performance gains are possible, the cycle will take place much more often. As it enters the phase of decreasing marginal returns to more efforts, the process will proceed at a slower pace until it eventually stops and the model is used for operational purposes. However, as the environment changes

or qualitatively new data or insight is available, then there is new potential value to be unlocked by an accelerating improvement cycle.

A further deliberation needs to be made - the utility of the model does not lie merely in the final output but also in the creative process of constructing it. During this phase the analysts and managers working on it will gain better familiarity with their organization's data, will appreciate key trends and drivers, and will become more aware of risk and how it is modeled. All this knowledge feeds into the decision-making process and has the potential to significantly magnify the benefit of the model-building exercise.

Finally, the potential for model risks also means that model results should be used as guidelines for decisions and not as final arbiters. The model output will have to be supplemented by both expert evaluation and business context in order to reach an optimal risk management decision.

Project 12

Choose one of the models that you have built as part of this course or some other models that you are interested in and have access to. Go through the following steps:

- Investigate its data and comment how data quality can impinge on model quality.
- Go through the model itself and see if any improvements are possible.
- Look into model output. Does it make sense? How does it compare to relevant benchmarks?
- Prescribe and implement improvement steps. Use the old and the updated model to do a forecasting exercise. How do they compare (in terms of significance, accuracy, and fit)?

Were results expected or not? Please comment.

Lecture Thirteen: Risk Aversion and Risk Seeking

Preferences to risk are not uniform across individuals - some can tolerate higher levels of risk, while others prefer to avoid it. In this lecture we look into more detail on how we can model risk preference in order to be able to give better risk management decisions, specifically tailored to an individual. We first review utility theory and the key difference between objective expectations, and subjective expected utility. We then proceed to derive a measure of individual risk preference - the Arrow-Pratt measure - and then conclude the discussion with a few key insights from behavioral economics. The discussion loosely follows the ideas in Mengov (2015) and the reader interested in a more in-depth view of the behavioral determinants of decisions is kindly directed to the book.

Preferences and Utility

A common way to model preference in economics and finance is the utility function which relates satisfaction or utility U to a number of goods, factors, or experiences, defined by the vector x_i, usually denoted as:

$$U = U(x_i)$$

Over discrete values of the utility function, the total utility, U, is the sum of individual utilities:

$$U(x_i) = \sum_{j=1}^{m} u_j(x_i)$$

for $i = 1, 2, ..., n$

In contrast, if utility is received continuously, we can use the integral over a certain period in order to gain understanding of total utility:

$$U(x_i) = \int_a^b u(x_i)dt$$

for $i = 1, 2, ..., n$

Under risk or uncertainty, the individual considers that the vector of outcomes x_i have to be weighted by their respective probability - the vector p_i, thus obtaining an expected utility upon which the individual can make decisions. The expected utility formula and logic is the same as the expected monetary value in Lecture 2, and the expectations defined in Lecture Four:

$$E[U] = \sum_{i=1}^{n} p_i x_i$$

or

$$E[U] = \int_a^b xf(x)dx$$

We now assume that an individual has a discrete risky prospective x with n possible outcomes x_i, each with a probability of p_i. The expected utility for this will be:

$$E[U(x)] = \sum_{i=1}^{n} u_i(p_i x_i)$$

or

$$E[U(x)] = \int_a^b u(x)f(x)dx$$

On the other hand the objective mathematical expectations will be:

$$\mu = E[x] = \sum_{i=1}^{n} p_i x_i$$

or

$$\mu = E[x] = \int_a^b xf(x)dx$$

Depending on the individual's attitude between those two quantities, we can differentiate between three different types of attitudes:

- $U(E[x]) = E[U(x)]$ - the individual is **risk-neutral** when the expected utility exactly equals the mathematical expectation. Such people are indifferent whether they take risks or not as they expect that over the long term and due to the law or large numbers, their outcomes will converge to the mathematical expectation. For example large financial institutions with many transactions can afford to be risk-neutral. Alternatively, some agents may be oblivious to risk.
- $U(E[x]) < E[U(x)]$ - the individual is **risk-averse** when the expected utility is smaller than the mathematical expectation. In this case the individual will derive less utility because of the uncertainty that have to be endured. Risk averse agents tend to prefer sure bets and are even prepared to pay to avoid risk (insurance).
- $U(E[x]) > E[U(x)]$ - the individual is **risk-seeking** when the expected utility exceeds the mathematical expectation. In those cases individuals derive satisfaction (adrenaline?) from the fact that they are exposed to risk, and in some cases they may be prepared to pay for this. For example extreme sports fans or extremely bold speculators fall in this category.

It seems that in many cases people tend to be risk averse. This gives rise to the notion of the so-called **certainty equivalents**. The certainty equivalent of a risky bet x_1 is the certain amount x_2 that gives the same utility as the utility of the expected value of x_1, or:

$$U(E[x_1]) = U(x_2)$$

Since the amount of x_2 will sometimes be lower than x_1 agents or organizations that can bear risk, can sell the certainty equivalent to individuals in exchange of the risky prospect and some premium. Essentially, this is the theory behind insurance.

The Arrow-Pratt Measure

A way to quantity the individual's risk attitude given his preferences (or utility function, $U(p_i, x_i)$) is the Arrow-Pratt measure. We derive it from first principles for the case of continuous events. Assume an individual will have wealth w_f (a proxy for consumption) composed of a riskless component w_0 and and risk component x with expected value of μ, thus:

$$w_f = w_0 + x$$

and

$$\mu = E[x]$$

The mathematical expectations of total wealth is thus:

$$E[w_f] = E[w_0 + x] = w_0 + E[x]$$

The expected utility is thus:

$$U(w_f) = \int_a^b u(w_0 + x) f(x) dx$$

If we denote the certainty equivalent as w^*, then the total utility from the certainty equivalent is $U(w^*)$. The natural question is what is the conditions under which the individual will be indifferent between the certainty equivalent, and the risky prospect, or:

$$U(w^*) = \int_a^b u(w_0 + x) f(x) dx$$

The price for this trade can be defined as the difference between the certainty equivalent and the risky prospect, or:

$$p_a = w^* - w_f$$

A risk-neutral individual will have an asking price exactly equal to the mathematical expectation of the risky component of w_f, or $p_a = E[x]$. A risk-averse or risk-seeking individual will generally ask for a different price. From this we can defined the risk premium η as being:

$$\eta = \mu - p_a$$

Positive risk premiums signal risk aversion, while negative one - risk seeking behavior. We substitute the price in the utility integral in order to obtain the following:

$$U(w_0 + p_a) = \int_a^b u(w_0 + x) f(x) dx$$

We can approximate both $U(w_0 + p_a)$ and $U(w_0 + x)$ using Taylor series approximations, reaching the following:

$$U(w_0 + p_a) \approx U(w_0 + \mu) + (p_a - \mu) U'(w_0 + \mu)$$

and

$$U(w_0 + x) \approx U'(w_0 + \mu) + (x - \mu) U'(w_0 + \mu) + \frac{(x - \mu)^2}{2!} U''(w_0 + \mu)$$

The second approximation is more precise (including the term with the second derivative) as the potential difference of $(w_0 + x)$ from the point of approximation $(w_0 + \mu)$ is probably larger. We substitute those approximations in the utility integral, and also use the definition of variance:

$$\int_a^b (x - \mu)^2 f(x) dx = \sigma^2$$

We finally obtain an expression for the risk premium:

$$p_a - \mu = \eta \approx -\frac{\sigma^2}{2} \frac{U''(w_0 + \mu)}{U'(w_0 + \mu)}$$

We see how the risk premium is directly proportional to the variance (squared standard deviation). The larger the dispersion of outcomes the more the risk-averse individual will be willing to pay in order to avoid risk. The opposite is true for the risk-seeker. An important thing to note for this measure is the presence of the first and second derivative of the utility function. This means that the metric is strictly personalized to an individual and its understanding can provide for tailor-made risk recommendations that take into account every agent's risk tolerance.

Constant Relative Risk Aversion

It is often useful to formally model risk preference directly in the utility function. There are a number of useful mathematical formulations for this but a particularly popular approach is to use a utility function of the Constant Relative Risk Aversion class. It is characterized by a

constant rate of substitution between two goods and is also easy to use and mathematically tractable. Total utility is denoted as $U(x)$ while x is some factor contributing to it. With θ we denote the risk aversion of the given individual. The CRRA utility function is of the following form:

$$u(x) = \frac{1}{1-\theta}x^{1-\theta}$$

The higher the values of the theta parameter, the more risk averse the individual is. Lower values signify less risk aversion, and negative values capture risk-seeking behavior. We should note that the function is not defined for $\theta = 1$. As the parameter approaches unity, the value of the function approaches $log(x)$, which is a useful way to complete its definition.

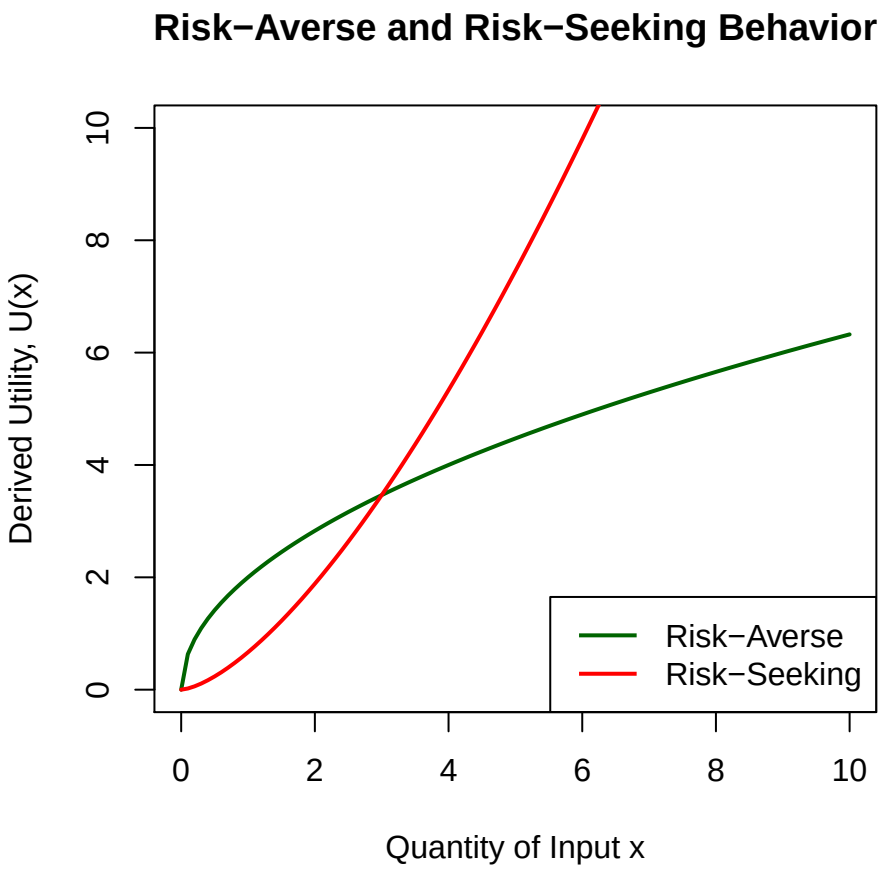
We present graphically the two types of behavior, setting:

- $\theta = 0.5$ for risk aversion
- $\theta = -0.5$ for risk seeking

```
x <- seq (0, 10, by = 0.1)
u1 <- 1/(1-0.5)*x^(1-0.5)
u2 <- 1/(1-(-0.5))*x^(1-(-0.5))
plot(x, u1, type="l", ylim = c(0, 10), col="darkgreen", lwd=2,
     xlab="Quantity of Input x", ylab = "Derived Utility, U(x)",
     main = "Risk-Averse and Risk-Seeking Behavior") + lines(x, u2,
      col="red", lwd=2)
```

```
## integer(0)
```

```
legend("bottomright", col = c("darkgreen","red"),
       legend=c("Risk-Averse", "Risk-Seeking"), lwd=2)
```



As we can see from the graph the risk-aversion utility function is concave and this form leads to the fact that the expected utility from a given uncertain amount of x is less than the utility from the value equaling the mathematical expectation, or $U(E[x]) < E[U(x)]$. On the other hand risk-seeking behavior is convex and thus $U(E[x]) > E[U(x)]$. While such CRRA functions are admittedly a simplified way of relating utility and risk preferences, they may be useful for modeling purposes.

Insights from Behavioral Economics

An important result from behavioral economics is that when making decision under risk agents usually do not use the objective mathematical probabilities p_i but rather a set of subjective probabilities $\pi(p_i)$ that are calculated by the brain. The difference between the two is somewhat significant which leads to skewing risky decisions.

Another key insight is that people derive asymmetric utility from positive and negative events. For example the gain of 10$ will add less utility than the loss of 10$ would subtract. This asymmetry has key RM implications. So far we have assumed risk as both upsides and downsides, and have treated them more or less equally. However, if we know that individuals

will lose significantly more utility from a loss than they would gain from an equal profit, then this slightly skews the job of the RM professional to avoid the downside.

Precise mathematical modeling of those consistent traits of human behavior is not entirely uncontroversial but the intellectual legacy of the Prospect Theory and its associated theories has been hugely influential. A popular possibility to relate objective and subjective probability is given through the following formula:

$$\pi(p) = \frac{p^{\delta}}{(p^{\delta} + (1-p)^{\delta})^{\frac{1}{\delta}}}$$

Laboratory experiment allow us to calibrate the value of the parameter δ for the cases of gains (δ_{π^+}) and losses (δ_{π^-}). Those are shown in the following table:

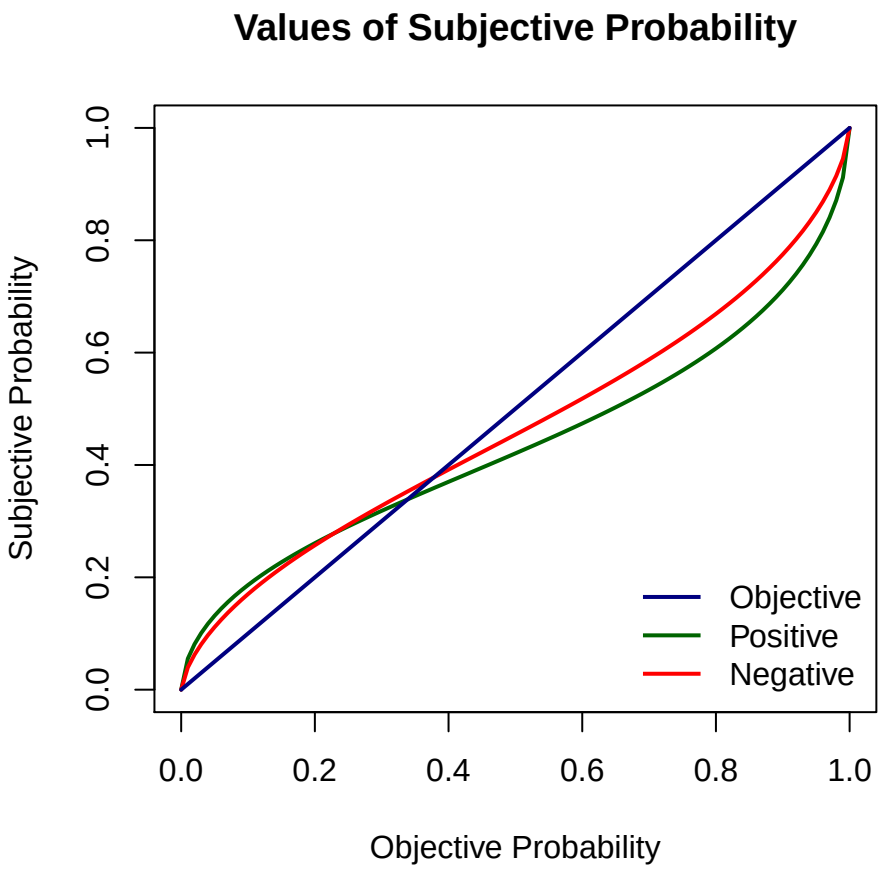
Probability	Parameter Value
$\pi^+(p)$	$\delta_{\pi^+} = 0.61$
$\pi^-(p)$	$\delta_{\pi^-} = 0.69$

Using those values we can construct the subjective probability and see its deviations from objective probability. This is done in the following plot.

```
p <- seq(from = 0, to = 1, by = 0.01)
deltapos <- 0.61; deltaneg <- 0.69;
pipos <- (p^deltapos)/(p^deltapos + (1 - p)^deltapos)^(1/deltapos)
pineg <- (p^deltaneg)/(p^deltaneg + (1 - p)^deltaneg)^(1/deltaneg)
plot(p, pipos, type= "l", lwd=2, col="darkgreen",
     xlab="Objective Probability",  ylab="Subjective Probability",
     main="Values of Subjective Probability") + lines(p, pineg,
     lwd=2, col="red") + lines(p, p, lwd=2, col="navyblue")
```

```
## integer(0)
```

```
legend(x="bottomright", legend=c("Objective", "Positive", "Negative"),
       col=c("navyblue", "darkgreen", "red"), lwd=2, bty="n")
```



Those insights from the economic laboratory can lead us to redefine the value of expected utility and the risk preferences stemming from it. We can say that total utility now depends both on objective outcomes x_i and their likelihood p_i as well as how the individuals perceive this likelihood, or $\pi(p_i)$. We thus obtain the following expression:

$$U(x_i, p_i) = \sum_{i=1}^{m} \pi^{+}(p_i)u(x_i) + \sum_{k=m+1}^{n} \pi^{-}(p_k)u(x_k)$$

Total expected utility in this expression is the sum of the expected utility derived from positive (gain) and negative (loss) experience weighted by how individuals *perceive* the likelihood for those events. The objective probability is thus modified by the individual's subjective perception of it, π.

This understanding of utility leads to four large behavioral archetypes of risk preference of the individual agent:

- **Risk aversion under large probability of gains** - this is a classical case, underlying traditional utility theory. This can explain the reticent behavior of small individual investor when they refuse to take risk by investing in equity and prefer the relative certainty of a bank deposit.

- **Risk aversion under small probability of loss** - this behavior is driven by an overestimation of small probabilities of negative events, especially those under 20%. Such overestimation makes people purchase insurance which can sometimes be sold at a larger than the fair premium due to this type of risk preference.
- **Risk seeking under small probability of gains** - such behavior is again driven by an overestimation of small probabilities but for positive events. Examples of this include the propensity of some individuals to participate in lottery games which often have small or even negative expected payoff.
- **Risk seeking under large probability of loss** - the underestimation of probabilities for negative outcomes, especially as they surpass 80% may lead agents to endeavor in excessively risky behavior. Examples of this include war and peacetime heroes.

Overall, human behavior may deviate from the axioms of strict economic rationality but it often does so in a relatively predictable way. It is therefore of crucial importance for the RM professional to understand and help correct biases in risk-taking behavior as he (or she) struggles to find the optimal trade-off between risk and reward.

Concluding Discussion

Individuals have both different preferences and different attitudes to risk. There are many models that capture basic characteristics of human behavior but it seems that classical axiomatic models tend to lose in descriptive and predictive power to more modern explanations advanced by the fields of behavioral economics, psychology and neuroeconomics.

Our brains are hardwired in such a way that we are systematically unable to correctly appreciate probabilities, thus giving rise to behavior that is excessively risky or not risky enough. As for business implications, this means that under a lot of different circumstances individuals are miscalculating risk and foregoing profit that they should take, and exposing themselves to unnecessary loss.

Therefore the risk management professionals who appreciate both the objective likelihood of risks, as well as the subjective perceptions of them can add significant value by gently guiding individuals and organizations to an optimal level of risk taking and return.

Project 13

Interview at least 5 individuals and try to gain data about their risk preferences. Attempt to construct a utility function using any way you deem appropriate and parametrize it.

- Investigate it analytically and graphically.
- What business insight can you construct from this exercise?

Lecture Fourteen: Concluding Comments

The Risk Management lecture cycle provided insight into both the philosophy and the practical application of some common tools for understanding and controlling the unexpected. We presented frameworks and instruments that aid the RM process and underscore the culture of intelligent appreciation of risk. A few larger points transcend specific topics and reappear throughout the discussion. Those key take-aways are some of the main (hopefully useful) conclusions from the course.

- **Risk has both positive and negative implications** and excessive fear of exposure prevents the agent from reaping the upsides and possible rewards in a risky world.
- **Risk needs to be actively managed and controlled** irrespective of whether data for that is currently available. Pretending that risk does exist has larger negative influence than investing efforts in mitigating it.
- **Risk does not disappear** – the best we can hope for is to pool it and let those who can bear it best to do so (e.g. insurers). Note that even very rare events happen and the world may seem less random than we expect.
- **Risk can be quantified using historical data** but those numbers should be interpreted carefully – we want to measure risk in the future and the past is not always a good guide.
- **Rational actors diversify** – they employ a wide range of strategies and those that succeed hedge against those that fail.
- **Risk can be expressed through very sophisticated mathematical models.** These are useful but rough guidelines. Usually they are ridden with assumptions and results vary dramatically over little deviations of parameter values.
- **VaR, ES, and other risk management models may give a false sense of precision and comfort** while we live in a risky world. The way to overcome this is to stay conservative and provide for larger buffers.
- **Sometimes we may not observe all relevant data** - due to data limitations the RM analyst does not always have empirical data and has to resort to simulations to gather it. Those are ridden with assumptions and provide at best a very approximate guide to reality.
- **Classification problems are prevalent but unfair** as they need to distinguish between classes of outcomes with different resulting utility. It may make business sense to skew or calibrate the model in order to optimize business results rather than statistical accuracy.
- **Models are imperfect simplifications of reality by construction.** Their errors increase at a dramatic pace and should not be substituted for common sense.
- **Risk management needs to be driven by business considerations** - only meaningful business risks are to be analyzed and the cost of the process should never

exceed its benefit.

- **Individuals have skewed risk perceptions** - they tend to overweigh small probabilities, and underweigh larger ones, which leads to predictable decision errors.
- **The key of risk management is not the elimination of risk but the ability to take intelligent risk-adjusted decisions** that are consonant with organizational needs, specific context and individual preferences.

All those key insights point to the important role of the risk manager as an enabler of people and organizations to take well-considered risks in search of adequate returns.

References

Basel Committee on Banking Supervision. (2011). *Basel Capital Accord III*. Switzerland: The Basel Committee on Banking Supervision

Basel Committee on Banking Supervision. (2004). *Basel Capital Accord II*. Switzerland: The Basel Committee on Banking Supervision

Crouhy, M., Galai, D., Mark, R. (2006) *The Essentials of Risk Management*. NY: McGraw-Hill.

Gerunov, A. (2016). Modeling Economic Choice under Radical Uncertainty: Machine Learning Approaches. *MPRA Working Paper #69199*. University of Munich Personal RePEc Archive.

Hastie, T., Tibshirani, R., & Friedman, J. (2011). *The Elements of Statistical Learning*. NY: Springer.

Knight, F. H. (1921). *Risk, Uncertainty, and Profit* NY: Hart, Schaffner and Marx.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91.

Markowitz, H. (1968). *Portfolio selection: efficient diversification of investments (Vol. 16)*. Yale University Press.

Mengov, G. (2015). *Decision Science: A Human-Oriented Perspective (Vol. 89)*. Berlin: Springer.

Pelsser, A. (2000). *Efficient methods for valuing interest rate derivatives*. Springer Science & Business Media.

Project Management Institute / PMI. (2013). *Project Management Body of Knowledge, 5th Edition*. US: PMI.

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robert, C., & Casella, G. (2009). *Introducing Monte Carlo Methods with R*. Springer Science & Business Media.

Sharpe, W. F. (1994). The Sharpe ratio. *The Journal of Portfolio Management*, 21(1), 49-58.

Zivot, E. & Wang, J. (2006). *Modeling Financial Time Series with S-PLUS*. NY: Springer.