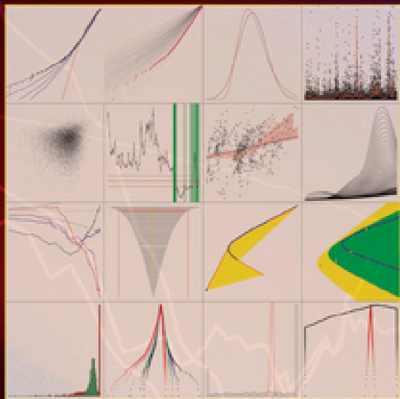
Wiley Series in Probability and Statistics



JUSSI KLEMELÄ

NONPARAMETRIC FINANCE

WILEY

Nonparametric Finance

WILEY SERIES IN PROBABILITY AND STATISTICS
Established by *Walter A. Shewhart and Samuel S. Wilks*

The ***Wiley Series in Probability and Statistics*** is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches.

This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

A complete list of titles in this series can be found at
http://www.wiley.com/go/wsps

Nonparametric Finance

Jussi Klemelä

WILEY

This edition first published 2018

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Klemelä, Jussi, 1965- author.
Title: Nonparametric finance / by Jussi Sakari Klemelä.
Description: Hoboken, NJ : John Wiley & Sons, Inc., [2018] | Series: Wiley series in probability and statistics | Includes bibliographical references and index. |
Identifiers: LCCN 2017038242 (print) | LCCN 2017046064 (ebook) | ISBN 9781119409113 (pdf) | ISBN 9781119409120 (epub) | ISBN 9781119409106 (cloth)
Subjects: LCSH: Finance—Statistical methods. | Finance—Mathematical models.
Classification: LCC HG176.5 (ebook) | LCC HG176.5 .K553 2018 (print) | DDC 332.01/51954–dc23
LC record available at https://lccn.loc.gov/2017038242

Cover design by Wiley
Cover image: (Background) © Michel Leynaud/Gettyimages;
(Image grid) Courtesy of Jussi Klemelä

Set in 10/12pt WarnockPro by SPi Global, Chennai, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents

Preface

We study applications of nonparametric function estimation into risk management, portfolio management, and option pricing.

The methods of nonparametric function estimation have not been commonly used in risk management. The scarcity of data in the tails of a distribution makes it difficult to utilize the methods of nonparametric function estimation. However, it has turned out that some semiparametric methods are able to improve purely parametric methods.

Academic research has paid less attention to portfolio selection, as compared to the attention that has been paid to risk management and option pricing. We study applications of nonparametric prediction methods to portfolio selection. The use of nonparametric function estimation to reach practical financial decisions is an important part of machine learning.

Option pricing might be the most widely studied part of quantitative finance in academic research. In fact, the birth of modern quantitative finance is often dated to the 1973 publication of the Black–Scholes option pricing formula. Option pricing has been dominated by parametric methods, and it is especially interesting to provide some insights of nonparametric function estimation into option pricing.

The book is suitable for mathematicians and statisticians who would like to know about applications of mathematics and statistics into finance. In addition, the book is suitable for graduate students, researchers, and practitioners of quantitative finance who would like to study some underlying mathematics of finance, and would like to learn new methods. Some parts of the book require fluency in mathematics.

Klemelä (2014) is a book that contains risk management (volatility prediction and quantile estimation) and it describes methods of nonparametric regression, which can be applied in portfolio selection. In this book, we cover those topics and also include a part about option pricing.

The chapters are rather independent studies of well-defined topics. It is possible to read the individual chapters without a detailed study of the previous material.

The research in the book is reproducible, because we provide R-code of the computations. It is my hope, that this makes it easier for students to utilize the book, and makes it easier for instructors to adapt the material into their teaching.

The web page of the book is available in http://jussiklemela.com/statfina/.

Helsinki, Finland *Jussi Klemelä*
June 2017

1

Introduction

Nonparametric function estimation has many useful applications in quantitative finance. We study four areas of quantitative finance: statistical finance, risk management, portfolio management, and pricing of securities.[1]

A main theme of the book is to study quantitative finance starting only with few modeling assumptions. For example, we study the performance of nonparametric prediction in portfolio selection, and we study the performance of nonparametric quadratic hedging in option pricing, without constructing detailed models for the markets. We use some classical parametric methods, such as Black–Scholes pricing, as benchmarks to provide comparisons with nonparametric methods.

A second theme of the book is to put emphasis on the study of economic significance instead of statistical significance. For example, studying economic significance in portfolio selection could mean that we study whether prediction methods are able to produce portfolios with large Sharpe ratios. In contrast, studying statistical significance in portfolio selection could mean that we study whether asset returns are predictable in the sense of the mean squared prediction error. Studying economic significance in option pricing could mean that we study whether hedging methods are able to well approximate the payoff of the option. In contrast, studying statistical significance in option pricing could mean that we study the goodness-of-fit of our underlying model for asset prices. Studying statistical significance can be important for understanding the underlying reasons for economic significance. However, the study of economic significance is of primary importance, and the study of statistical significance is of secondary importance.

1 The quantitative finance section of preprint archive "arxiv.org" contains four additional sections: computational finance, general finance, mathematical finance, and trading and market microstructure. We cover some topics of computational finance that are useful in derivative pricing, such as lattice methods and Monte Carlo methods. In addition, we cover some topics of mathematical finance, such as the fundamental theorems of asset pricing.

Nonparametric Finance, First Edition. Jussi Klemelä.

A third theme of the book is the connections between the various parts of quantitative finance.

1) There are connections between risk management and portfolio selection: In portfolio selection, it is important to consider not only the expected returns but also the riskiness of the assets. In fact, the distinction between risk management and portfolio selection is not clear-cut.
2) There are connections between risk management and option pricing: The prices of options are largely influenced by the riskiness of the underlying assets.
3) There are connections between portfolio management and option pricing: Options are important assets to be included in a portfolio. In addition, multiperiod portfolio selection and option hedging can both be casted in the same mathematical framework.

Volatility prediction is useful in risk management, option pricing, and portfolio selection. Thus, volatility prediction is a constant topic throughout the book.

1.1 Statistical Finance

Statistical finance makes statistical analysis of financial and economic data.

Chapter 2 contains a description of the basic financial instruments, and it contains a description of the data sets that are analyzed in the book.

Chapter 3 studies univariate data analysis. We study univariate financial time series, but ignore the time series properties of data. A decomposition of a univariate distribution into the central part and into the tail parts is an important theme of the chapter.

1) We use different estimators for the central part and for the tails. Nonparametric density estimation is efficient at the center of a univariate distribution, but in the tails of the distribution the scarcity of data makes nonparametric estimation difficult. When we combine a nonparametric estimator for the central part and a parametric estimator for the tails then we obtain a semiparametric estimator for the distribution.
2) We use different visualization methods for the central part and for the tails. We apply two basic visualization tools: (1) kernel density estimates and (2) tail plots. Kernel density estimates can be used to visualize and to estimate the central part of the distribution. Tail plots are an empirical distribution based tool, and they can be used to visualize the tails of the distribution.

Chapter 4 studies multivariate data analysis. Multivariate data analysis considers simultaneously several time series, but the time series properties are ignored, and thus the analysis can be called cross-sectional. A basic concept is the copula, which makes it possible to compose a multivariate distribution into the part that describes the dependence and into the parts that describe the marginal distributions. We can estimate the marginal distributions using nonparametric methods, but to estimate dependence for a high-dimensional distribution it can be useful to apply parametric models. Combining nonparametric estimators of marginals and a parametric estimator of the copula leads to a semiparametric estimator of the distribution. Note that there is an analogy between the decomposition of a multivariate distribution into the copula and the marginals, and between the decomposition of a univariate distribution into the tails and the central area.

Chapter 5 studies time series analysis. Time series analysis adds the elements of dependence and time variation into the univariate and multivariate data analysis. Completely nonparametric time series modeling tends to become quite multidimensional, because dependence over k consecutive time points leads to the estimation of a k-dimensional distribution. However, a rather convenient method for time series analysis is obtained by taking as a starting point a univariate or a multivariate parametric model, and estimating the parameter using time localized smoothing. For example, we can apply time localized least squares or time localized maximum likelihood.

Chapter 6 studies prediction. Prediction is a central topic in time series analysis. The previous observations are used to predict the future observations. A distinction is made between moving average type of predictors and state space type of predictors. Both types of predictors can arise from parametric time series modeling: moving average and GARCH (1, 1) models lead to moving average predictors, and autoregressive models lead to state space predictors. It is easy to construct nonparametric moving average predictors, and nonparametric regression analysis leads to nonparametric state space predictors.

1.2 Risk Management

Risk management studies measurement and management of financial risks. We concentrate on the market risk, which means the risk of unfavorable moves of asset prices.[2]

Chapter 7 studies volatility prediction. Prediction of volatility means in our terminology that the square of the return of a financial asset is predicted. The volatility prediction is extremely useful in almost every part of quantitative

2 Other relevant types of risk are credit risk, liquidity risk, and operational risk. Credit risk means the risk of the default of a debtor and the risks resulting from downgrading the rating of a debtor. Liquidity risk means the risk from additional cost of liquidating a position when buyers are rare. Operational risk means the risk caused by natural disasters, failures of the physical plant and equipment of a firm, failures in electronic trading, clearing or wire transfers, trading and legal liability losses, internal and external theft and fraud, inappropriate contractual negotiations, criminal mismanagement, lawsuits, bad advice, and safety issues.

finance: we can apply volatility prediction in quantile estimation, and volatility prediction is an essential tool in option pricing and in portfolio selection. In addition, volatility prediction is needed when trading with variance products. We concentrate on the following three methods:

1) GARCH models are a classical and successful method to produce volatility predictions.
2) Exponentially weighted moving averages of squared returns lead to volatility predictions that are as good as GARCH $(1, 1)$ predictions.
3) Nonparametric state space smoothing leads to improvements of GARCH $(1, 1)$ predictions. We apply kernel regression with two explanatory variables: a moving average of squared returns and a moving average of returns. The response variable is a future squared return. A moving average of squared returns is in itself a good volatility predictor, but including a kernel regression on top of moving averages improves the predictions. In particular, we can take the leverage effect into account. The leverage effect means that when past returns have been low, then the future volatility tends to be higher, as compared to the future volatility when the past returns have been high.

Chapter 8 studies estimation of quantiles. The term *value-at-risk* is used to denote upper quantiles of a loss distribution of a financial asset. Value-at-risk at level $0.5 < p < 1$ has a direct interpretation in risk management: it is such value that the probability of losing more has a smaller probability than $1 - p$. We concentrate on the following three main classes of quantile estimators:

1) The empirical quantile estimator is a quantile of the empirical distribution. The empirical quantile estimator has many variants, since it can be used in conditional quantile estimation and it can be modified by kernel smoothing. In addition, empirical quantiles can be combined with volatility based and excess distribution based methods, since empirical quantiles can be used to estimate the quantiles of the residuals.
2) Volatility based quantile estimators apply a location-scale model. A volatility estimator leads directly to a quantile estimator, since estimation of the location is less important. The performance of volatility based quantile estimators depends on the choice of the base distribution, whose location and scale is estimated. However, in a time series setting the use of the empirical quantiles of the residuals provides a method that bypasses the problem of the choice of the base distribution.
3) Excess distribution based quantile estimators model the tail parametrically. These estimators ignore the central part of the distribution and model only the tail part parametrically. The tail part of the distribution is called the excess distribution. Extreme value theory can be used to justify the choice of the generalized Pareto distribution as the model for the excess distribution. Empirical work has confirmed that the generalized Pareto distribution

provides a good fit in many cases. In a time series setting the estimation can be improved if the parameters of the excess distribution are taken to be time changing. In addition, in a time series setting we can make the estimation more robust to the choice of the parametric model by applying the empirical quantiles of the residuals. In this case, the definition of a residual is more involved than in the case of volatility based quantile estimators.

1.3 Portfolio Management

Portfolio management studies optimal security selection and capital allocation. In addition, portfolio management studies performance measurement.

Chapter 9 discusses some basic concepts of portfolio theory.

1) A major issue is to introduce concepts for the comparison of wealth distributions and return distributions. The comparison can be made by the Markowitz mean–variance criterion or by the expected utility. We need to define what it means that a return distribution is better than another return distribution. This is needed both in portfolio selection and in performance measurement.
2) A second major issue is the distinction between the one period portfolio selection and multiperiod portfolio selection. We concentrate on the one period portfolio selection, but it is instructive to discuss the differences between the approaches.

Chapter 10 studies performance measurement.

1) The basic performance measures that we discuss are the Sharpe ratio, certainty equivalent, and the alpha of an asset.
2) Graphical tools are extremely helpful in performance measurement. The performance measures are sensitive to the time period over which the performance is measured. The graphical tools address the issue of the sensitivity of the time period to the performance measures. The graphical tools help to detect periods of good performance and the periods of bad performance, and thus they give clues for searching explanations for good and bad performance.

Chapter 11 studies Markowitz portfolio theory. Markowitz portfolios are such portfolios that minimize the variance of the portfolio return, under a minimal requirement for the expected return of the portfolio. Markowitz portfolios can be utilized in dynamic portfolio selection by predicting the future returns, future squared returns, and future products of returns of two assets, as will be done in Chapter 12.

Chapter 12 studies dynamic portfolio selection. Dynamic portfolio selection means in our terminology such trading where the weights of the portfolio are rebalanced at the beginning of each period using the available information. Dynamic portfolio selection utilizes the fact that the expected returns, the expected squared returns (variances), and the expected products of returns (covariances) change in time. The classical insight of efficient markets has to be modified to take into account the predictability of future returns and squared returns.

1) First, we discuss how prediction can be used in portfolio selection. Time series regression can be applied in portfolio selection both when we use the maximization of the expected utility and when we use mean–variance preferences. In the case of the maximization of the expected utility, we predict the future utility transformed returns with time series regression. In the case of mean–variance preferences we predict, the future returns, squared returns, and products of returns.
2) The Markowitz criterion can be seen as decomposing the expected utility into the first two moments. The decomposition has the advantage that different methods can be used to predict the returns, squared returns, and products of returns. The main issue is to study the different types of predictability of the mean and the variance. In fact, most of the predictability comes from the variance part, whereas the expectation part has a much weaker predictability.
   a) We need to use different prediction horizons for the prediction of the returns and for the prediction of the squared returns. For the prediction of the returns we need to use a prediction horizon of 1 year or more. For the prediction of squared returns we can use a prediction horizon of 1 month or less.
   b) We need to use different prediction methods for the prediction of the returns and for the prediction of the squared returns. For the prediction of the returns, it is useful to apply such explanatory variables as dividend yield and term spread. For the prediction of the squared returns we can apply GARCH predictors or exponentially weighted moving averages.

1.4 Pricing of Securities

Pricing of securities considers valuation and hedging of financial securities and their derivatives.

Chapter 13 studies principles of asset pricing. We start the chapter by a heuristic introduction to pricing of securities, and discuss such concepts as absolute pricing, relative pricing using arbitrage, and relative pricing using "statistical arbitrage."[3]

3 The term *statistical arbitrage* refers often to pairs trading and to the application of mean reversion. We use term *statistical arbitrage* more generally, to refer to cases where two payoffs are close to each other with high probability. Thus, also term *probabilistic arbitrage* could be used.

1) The first main topic is to state and prove the first fundamental theorem of asset pricing in discrete time models, and to state the second fundamental theorem of asset pricing. These theorems provide the foundations on which we build the development of statistical methods of asset pricing. We give a constructive proof of the first fundamental theorem of asset pricing, instead of using tools of abstract functional analysis. The constructive proof of the first fundamental theorem of asset pricing turns out to be useful, because the method can be applied in practise to price options in incomplete models. The construction uses the Esscher martingale measure, and it is a special case of using utility functions to price derivatives.
2) The second main topic is to discuss evaluation of pricing and hedging methods. The basic evaluation method will be to measure the hedging error. The hedging error is the difference between the payoff of the derivative and the terminal value of the hedging portfolio. By measuring the hedging error, we simultaneously measure the modeling error and the estimation error. Minimizing the hedging error has economic significance, whereas modeling error and estimation error are underlying statistical concepts. Thus, emphasizing the hedging error is an example of emphasizing economic significance instead of statistical significance.

Chapter 14 studies pricing by arbitrage. The principle of arbitrage-free pricing combines two different topics: pricing of futures and pricing of options in complete models, like binary models and the Black–Scholes model.

1) A main topic is pricing in multiperiod binary models. First, these models introduce the idea of backward induction, which is an important numerical tool to value options in the Black–Scholes model, and which is an important tool in quadratic hedging. Second, these models lead asymptotically to the Black–Scholes prices.
2) A second main topic is to study the properties of Black–Scholes hedging. We illustrate how hedging frequency, strike price, expected return, and volatility influence the hedging error. These illustrations give insight into hedging methods in general, and not only into Black–Scholes hedging.
3) A third main topic is to study how Black–Scholes pricing and hedging performs with various volatility predictors. Black–Scholes pricing and hedging provides a benchmark, against which we can measure the performance of other pricing methods. Black–Scholes pricing and hedging assumes that the stock prices have a log-normal distribution with a constant volatility. However, when we combine Black–Scholes pricing and hedging with a time changing GARCH (1, 1) volatility, then we obtain a method that is hard to beat.

Chapter 15 gives an overview of several pricing methods in incomplete models. Binary models and the Black–Scholes model are complete models, but we are interested in option pricing when the model makes only few restrictions on the underlying distribution of the stock prices. Chapter 16 is devoted to quadratic hedging, and in Chapter 15 we discuss pricing by utility maximization, pricing by absolutely continuous changes of measures, pricing in GARCH models, pricing by a nonparametric method, pricing by estimation of the risk neutral density, and pricing by quantile hedging.

1) A main topic is to introduce two general approaches for pricing derivatives in incomplete models: the method of utility functions and the method of an absolutely continuous change of measure (Girsanov's theorem). For some Gaussian processes and for some utility functions these methods coincide. The method of utility functions can be applied to construct a nonparametric method of pricing options, whereas Girsanov's theorem can be applied in the case of some processes with Gaussian innovations, such as some GARCH processes.
2) A second main topic is to discuss pricing in GARCH models. GARCH $(1,1)$ model gives a reasonable fit to the distribution of stock prices. Girsanov's theorem can be used to find a natural pricing function when it is assumed that the stock returns follow a GARCH $(1,1)$ process. Heston–Nandi modification of the standard GARCH $(1,1)$ model leads to a computationally attractive pricing method. Heston–Nandi model has been rather popular, and it can be considered as a discrete time version of continuous time stochastic volatility models.

Chapter 16 studies quadratic hedging. In quadratic hedging the price and the hedging coefficients are determined so that the mean squared hedging error is minimized. The hedging error means the difference between the terminal value of the hedging portfolio and the value of the option at the expiration.

1) A main aim of the chapter is to derive recursive formulas for quadratically optimal prices and hedging coefficients. It is important to cover both the global and the local quadratic hedging. Local quadratic hedging leads to formulas that are easier to implement than the formulas of global quadratic hedging. Quadratic hedging has some analogies with linear least squares regression, but quadratic hedging is a version of sequential regression, which is done in a time series setting. In addition, quadratic hedging does not assume a linear model, but we are searching the best linear approximation in the sense of the mean squared error.
2) A second main aim of the chapter is to implement quadratic hedging. This will be done only for local quadratic hedging. We implement local quadratic hedging nonparametrically, without assuming any model for the underlying distribution of the stock prices. Although quadratic hedging finds an optimal linear approximation for the payoff of the option, the quadratically optimal price and hedging coefficients have a nonlinear dependence on volatility, and thus nonparametric approach may lead to a better fit for these nonlinear functions than a parametric modeling.

Chapter 17 studies option strategies. Option strategies provide a large number of return distributions to choose from, so that it is possible to create a portfolio that is tailored to the expectations and the risk profile of each investor. We discuss such option strategies as vertical spreads, strangles, straddles, butterflies, condors, and calendar spreads. Options can be combined with stocks to create covered calls and protective put. Options can be combined with bonds to create capital guarantee products. We give insight into these option strategies by estimating the return distributions of the strategies.

Chapter 18 describes interest rate derivatives. The market of interest rate derivatives is even larger than the market of equity derivatives. Interest rate forwards include forward zero-coupon bonds, forward rate agreements, and swaps. Interest rate options include caps and floors.

Part I

Statistical Finance

2

Financial Instruments

The basic assets which are traded in financial markets include stocks and bonds. A large part of financial markets consists of trading with derivative assets, like futures and options, whose prices are derived from the prices of the basic assets. Stock indexes can be considered as derivative assets, since the price of a stock index is a linear combination of the prices of the underlying stocks. A stock index is a more simple derivative asset than an option, whose terminal price is a nonlinear function of the price of the underlying stock.

In addition, we describe in this section the data sets which are used throughout the book to illustrate the methods.

2.1 Stocks

Stocks are securities representing an ownership in a corporation. The owner of a stock has a limited liability. The limited liability implies that the price of a stock is always nonnegative, so that the price S_t of a stock at time t satisfies

$$0 \le S_t < \infty.$$

Stock issuing companies have a variety of legal forms depending on the country of domicile of the company.[1] Common stock typically gives voting rights in company decisions, whereas preferred stock does not typically give voting rights, but the owners of preferred stocks are entitled to receive a certain amount of dividend payments before the owners of common stock can receive any dividends.

1 Statistical data of stock prices is usually available only for the stocks that are publicly traded in a stock exchange. In UK the companies whose stocks are publicly traded are called public limited companies (PLC), and in Germany they are called Aktiengesellschaften (AG). The companies whose owners have a limited liability but whose stocks are not publicly traded are called private companies limited by shares (Ltd), and Gesellschaft mit beschränkter Haftung (GmbH).

Nonparametric Finance, First Edition. Jussi Klemelä.

2.1.1 Stock Indexes

We define a stock index, give examples of the uses of stock indexes, and give examples of popular stock indexes.

2.1.1.1 Definition of a Stock Index

The price of a stock index is a weighted sum of stock prices. The value I_t of a stock index at time t is calculated by formula

$$I_t = C \sum_{i=1}^{d} n_i S_t^i, \tag{2.1}$$

where C is a constant, d is the number of stocks in the index, n_i is the number of shares of stock i, and S_t^i is a suitably adjusted price of stock i at time t, where $i = 1, \dots, d$. Note that $n_i S_t^i$ is the market capitalization of stock i. The definition of a stock index involves three parameters: constant C, numbers n_i, and values S_t^i:

1) The constant C can be chosen, for example, to make the value of the index equal to 100 at a given past day. When the constitution of the index is changed, then the constant C is changed, to keep the index equal to 100 at the chosen day.
2) The numbers n_i can equal the total number of shares of stock i, but they can also be equal to the number of freely floating stocks. Float market capitalization excludes stocks which are not freely floating (cannot be bought in the open market).
3) The values S_t^i are calculated differently depending on whether the index is a price return index or a total return index. Price return indexes are calculated without regard to cash dividends but total return indexes are calculated by reinvesting cash dividends. The adjusted closing price of a stock is the closing price of a stock which is adjusted to cash dividends, stock dividends, stock splits, and also to more complex corporate actions, such as rights offerings. The calculation of the adjusted closing price is often made by data providers.

2.1.1.2 Uses of Stock Indexes

Stock indexes can be used to summarize information about stock markets. Stock indexes can also be used as a proxy for the market index when testing and applying finance theories. The market index is the stock index which sums the values of all companies worldwide. Stock indexes are traded in futures markets and in exchanges as exchange traded funds (ETF). Furthermore, investment banks provide financial instruments whose values depend on stock indexes.

2.1.1.3 Examples of Stock Indexes

Dow Jones Industrial Average Dow Jones Industrial Average is an index where the prices are not weighted by the number of shares, and thus Dow Jones Industrial Average is an exception of the rule (2.1). Dow Jones Industrial Average is just a sum of the prices of the components, multiplied by a constant.

S&P 500 S&P 500 was created at March 4, 1957. It was calculated back until 1928 and the basis value was taken to be 10 from 1941 until 1943. The S&P 500 index is a price return index, but there exists also total return versions (dividends are invested back) and net total return versions (dividends minus taxes are invested back) of the S&P 500 index. The S&P 500 is a market value weighted index: prices of stocks are weighted according to the market capitalizations of the companies. Since 2005 the index is float weighted, so that the market capitalization is calculated using only stocks that are available for public trading.

Nasdaq-100 Nasdaq-100 is calculated since January 31, 1985. The basis value was at that day 250. Nasdaq-100 is a price index, so that the dividends are not included in the value of the index. Nasdaq-100 is a different index than Nasdaq Composite, which is based on 3000 companies. Nasdaq-100 is calculated using the 100 largest companies in Nasdaq Composite. Nasdaq-100 is a market value weighted index, but the influence of the largest companies is capped (the weight of any single company is not allowed to be larger than 24%).

DAX 30 DAX 30 (Deutscher AktienindeX) was created at July 1, 1988. The basis value is 1000 at December 31, 1987. DAX 30 is a performance index (dividends are reinvested in calculating the value of the index). DAX 30 stock index is a market value weighted index of 30 largest German companies. Market value is calculated using only free floating stocks (stocks that are not owned by an owner which has more than 5% of stocks). The largeness of a company is measured by taking into account both the free floating market value and the transaction volume (total value of the stocks that are exchanged in a given time period). The weight of any single company is not allowed to be larger than 10%.

2.1.2 Stock Prices and Returns

Statistical analysis of stock markets is usually done from time series of returns. Before defining a return time series we describe the initial price data in its raw form, as it is evolving in a stock exchange, and we describe some methods of sampling of prices.

2.1.2.1 Initial Price Data

During the opening hours of an exchange the stocks are changing hands at irregular time points. The stock exchange receives bid prices with volumes (numbers of stocks one is willing to buy with the given bid price) from buyers, and ask prices with volumes from the sellers. The exchange has an algorithm which allocates the stocks from the sellers to the buyers. The allocation happens when there are bid prices and ask prices that meet each other (ask prices that are smaller or equal to bid prices). The algorithms of stock allocation take into account the arrival times of the orders, the volumes of the orders, and the types of the orders.

The most common order types are the market order and the limit order. A market order expresses the intention to buy the stock at the lowest ask price, or the intention to sell the stock at the highest bid price. A limit order expresses the intention to buy the stock at the lowest ask price, under the condition that the ask price is lower than the given limit price, or the intention to sell the stock at the highest bid price, under the condition that the bid price is higher than the given limit price.

2.1.2.2 Sampling of Prices

The price changes at irregular time intervals in a stock exchange, but for the purpose of a statistical analysis we typically sample price at equispaced intervals.

To obtain a time series of daily prices, we can pick the closing price of each trading day. The closing price can be considered as the consensus reached between the sellers and the buyers about the fair price, taking into account all information gathered during the day. An alternative method would choose the opening price.

However, depending on the purpose of the analysis, we can sample data once in a second, once in 10 days, or once in a month, for example. Note that when the sampling interval is longer (monthly, quarterly, or yearly), the number of observations in a return time series will be smaller, and thus the statistical conclusions may be more vague. Note also, that the distribution of the returns may vary depending on the sampling frequency.

It is not obvious how to define equispaced sampling, since we can measure the time as the physical time, trading time, or effective trading time:

1) The physical time is the usual time in calendar days. Assume that we want to sample data once in 20 days. If we use the physical time, then we calculate all calendar days.
2) The trading time or market time takes into account only the time when markets are open. For example, when we want to sample data once in 20 days and we use trading time, then we calculate only the trading days (not all calendar days). However, information is accumulating also during the weekends (and during the night), which would be an argument in favor of physical time.

3) The effective trading time takes into account that the market activity is not uniform during market hours. To define the sampling interval, we could take into account the number of transactions, or the volume of the transactions. The effective trading time is interesting especially when we gather intraday data, but it can be used also in the case of longer sampling intervals, to correct for diminishing market activity during summer or at the end of year.[2]

Sampling daily closing prices can be interpreted as using the trading time, because weekends and holidays are ignored in the daily sampling. Since there is roughly the same number of trading days in every week and every month, we can interpret sampling the weekly and monthly closing prices both as using the physical time and using the trading time. Discussion about scales in finance is provided by Mantegna and Stanley (2000).

2.1.2.3 Stock Returns

Let us consider a time series $S_0, \dots, S_T$ of stock prices, sampled at equispaced time points. We can calculate gross returns, net returns, or logarithmic returns.

1) Gross returns (price relatives) are defined by

$$\frac{S_{t+1}}{S_t},$$

2) net returns (relative price differences) are defined by

$$\frac{S_{t+1} - S_t}{S_t},$$

3) logarithmic returns (continuously compounded returns) are defined by

$$\log\left(\frac{S_{t+1}}{S_t}\right),$$

where $t = 0, \dots, T-1$.

Gross returns are positive numbers like 1.02 (when the stock rose 2%) or 0.98 (when the stock fell 2%). Value zero for a gross return means bankruptcy. The gross returns have a concrete interpretation: starting with wealth W_t and buying a stock with price S_t leads to the wealth $W_{t+1} = W_t \times S_{t+1}/S_t$.

Net returns are obtained from gross returns by subtracting one, and thus net returns are numbers larger than -1. Net returns are numbers like 0.02 (when the stock rose 2%) or -0.02 (when the stock fell 2%). Value -1 for a net return means bankruptcy.

---

2 Let V_u be the number or the volume of the transactions at time u. After sampling time t_i is chosen, we can determine the next sampling time t_{i+1} by

$$t_{i+1} = \min\left\{t : \sum\{V_u : t_i \le u \le t\} \ge C\right\},$$

where $C > 0$ is a constant.

Logarithmic returns are obtained from gross returns by taking the logarithm.[3] A logarithmic return can take any real value, but typically logarithmic returns are close to net returns, because $\log(x) \approx x - 1$ when $x \approx 1$. Value $-\infty$ for a logarithmic return means bankruptcy. The logarithmic function is an example of a utility function, as discussed in Section 9.2.2. We will consider taking the logarithm as an application of a utility function, and apply mainly gross returns. However, there are some reasons for the use of logarithmic returns. First, we can derive approximate distributions for the stock price by applying limit theorems for the sum of the logarithmic returns, which makes the study of logarithmic returns interesting. Indeed, we can write

$$S_T = S_0 \exp\left\{ \sum_{t=0}^{T-1} \log\left(\frac{S_{t+1}}{S_t}\right) \right\}. \tag{2.2}$$

See (3.49) for a more detailed derivation of the log-normal model for stock prices. Second, taking logarithms of returns transforms the original time series of prices to a stationary time series, as explained in the connection of Figure 5.1.

For a statistical modeling we need typically a stationary time series. Stationarity is defined in Section 5.1. For example, autoregressive moving average processes (ARMA) and generalized autoregressive conditional heteroskedasticity (GARCH) models, defined in Section 5.3, are stationary time series models. The original time series of stock prices is not a stationary time series, but it can be argued that a return time series is close to stationarity.[4]

Note that we can write, analogously to (2.2),

$$S_T = S_0 + \sum_{t=0}^{T-1} (S_{t+1} - S_t).$$

Thus, we can derive approximate distributions for the stock price by applying limit theorems for the sum of the price differences. See (3.46) for a more detailed derivation of the normal model for stock prices. The time series of price differences is not a stationary time series, as discussed in the connection of Figure 5.2. However, for short time periods a time series of price differences can be approximately stationary. Thus, modeling price differences instead of returns can be reasonable.

3 We take the logarithm to be the natural logarithm, with e (Euler's number or Napier's constant) as the basis. The logarithmic functions with other bases could be used as well.

4 Time series $\{Y_t\}$ is called strictly stationary, if $(Y_1, \ldots, Y_t)$ and $(Y_{1+k}, \ldots, Y_{t+k})$ are identically distributed for all $t, k \in \{0, \pm 1, \pm 2, \ldots\}$. Stationarity means, roughly speaking, that every subperiod of the time series has similar statistical characteristics. For example, consider a stock whose price is 1\$, which then rises to have a price of 100\$. The change of 1\$ is very large at the beginning of the period but moderate at the end of the period. Thus, the time series of prices is not stationary.

2.2 Fixed Income Instruments

One unit of currency today is better than one unit of currency tomorrow. Fixed income research studies how much one should pay today, in order to receive a cash payment at a future day.

Fixed income instruments are described in more detail in Chapter 18. Here we give an overview of zero-coupon bonds, coupon paying bonds, interest rates, and of calculation of bond returns.

2.2.1 Bonds

Bonds include zero-coupon bonds and coupon bearing bonds.

1) A zero-coupon bond, or a pure discount bond, is a certificate which gives the owner a nominal amount P (principal) at the future maturity time T. Typically we take $P = 1$.
2) Coupon bearing bonds make regular payments (coupons) before the final payment at the maturity. A coupon bond can be defined as a series of payments $P_1, \dots, P_n$ at times $T_1, \dots, T_n$. The terminal payment contains the principal and the final coupon payment.[5]

A zero-coupon bond is a more basic instrument than a coupon bond, because a coupon bond can be defined as a portfolio of zero-coupon bonds. Let $C(t_0, T_n)$ be the price of a coupon bond which starts at t_0 and makes payments $P_1, \dots, P_n$ at times $T_1 < \cdots < T_n$, where $T_1 > t_0$. It holds that

$$C(t_0, T_n) = \sum_{i=1}^{n} P_i Z(t_0, T_i),$$

where $Z(t_0, T_i)$ are the prices of zero-coupon bonds starting at t_0 with maturity T_i, and with principal $P = 1$.

The cash flow generated by a bond is determined when the bond is issued. The bond can be traded before its maturity and its price can fluctuate before the maturity. For example, the price of a zero-coupon bond with the nominal amount P is equal to P at the maturity, but its price fluctuates until the maturity is reached. The price fluctuates as a function of interest rate fluctuation. Thus, bonds bear interest rate risk if they are not kept until maturity. If the bonds are kept until maturity they bear the inflation risk and the risk of the default of the issuer.

Bonds can be divided by the issuer. The main classes are government bonds, municipal bonds, and corporate bonds. Credit rating services give credit ratings

5 For example, a 5 year 4% semi-annual coupon bond with 1000\$ face value makes ten 20\$ payments every 6 months and the final payment of 1000\$. Thus $P_i = 20\$$ for $i = 1, \dots, n-1$ and the last payment is $P_n = 1020\$$, where $n = 10$.

to the bond issuers. Credit ratings help the investors to evaluate the probability of the payment default. Credit rating services include Standard & Poor's and Moody's.

US Treasury securities are backed by the US government. US Treasury securities include Treasury bills, Treasury notes, and Treasury bonds.

1) Treasury bills are zero-coupon bonds with original time to maturity of 1 year or less.[6]
2) Treasury notes are coupon bonds with original time to maturity between 2 and 10 years.
3) Treasury bonds are coupon bonds with original time to maturity of more than 10 years.

Widely traded German government bonds include Bundesschatzanweisungen (Schätze), which are 2 year notes, Bundesobligationen (Bobls), which are 5 year notes, and Bundesanleihen (Bunds and Buxl), which are 10 and 30 year bonds.

There are many types of fixed income securities. Callable bonds are such bonds that allow the bond issuer to purchase the bond back from the bondholders. The callable bonds make it possible for the issuer to retire old high-rate bonds and issue new low-rate bonds. Floating rate bonds (floaters) are such bonds whose rates are adjusted periodically to match inflation rates. Treasury STRIPS are such fixed income securities where the principal and the interest component of US Treasury securities are traded as separate zero coupon securities. The acronym STRIPS means separate trading of registered interest and principal securities.

2.2.2 Interest Rates

Interest rates are the basis for many financial contracts. We can separate between the government rates and the interbank rates. The government rates are deduced from the bonds issued by the governments and the interbank rates are obtained from the rates at which deposits are exchanged between banks.

Libor (London interbank offered rate) and Euribor (Euro interbank offered rate) are important interbank rates. Eonia (Euro overnight index average) is an overnight interest rate within the eurozone, but unlike the Euribor and Libor does not include term loans. Eonia is similar to the federal funds rate in the US. Sonia (Sterling overnight index average) is the reference rate for overnight unsecured transactions in the Sterling market.

Euribor and Libor are comparable base rates. Euribor rates are trimmed averages of interbank interest rates at which a collection of European banks are

6 The Treasury issues bills with times to maturity of 13 weeks, 26 weeks, and 52 weeks (3-month bills, 6-month bills, and 1-year bills). 13-week bills and 26-week bills are auctioned once a week and 52-week bills are auctioned once a month.

prepared to lend to one another. Libor rates are trimmed averages of interbank interest rates at which a collection of banks on the London money market are prepared to lend to one another. Euribor and Libor rates come in different maturities. In contrast to Euribor rates, the Libor rates come in different currencies. Euribor and Libor rates are not based on actual transactions, whereas Eonia is based on actual transactions. A study published in May 2008 in The Wall Street Journal suggested that the banks may have understated the borrowing costs. This led to reform proposals concerning the calculation of the Libor rates.

The Eonia rate is the rate at which banks provide unsecured loans to each other with a duration of 1 day within the Euro area. The Eonia rate is a volume weighted average of transactions on a given day and it is computed by the European Central Bank by the close of the real-time gross settlement on each business day. Eonia can be considered as the 1 day Euribor rate or as the Euro version of overnight index swaps (OIS). The Eonia panel consists of over 50 mostly European banks. The banks are chosen to the panel based on their premium credit rating and the high volume of their money market transactions conducted within the Eurozone. Banks on the Eonia panel are the same banks included in the Euribor panel.

Euribor rates are used as a reference rate for euro-denominated forward rate agreements, short term interest rate futures contracts, and interest rate swaps. Libor rates are used for Sterling and US dollar-denominated instruments.

2.2.2.1 Definitions of Interest Rates

The different definitions of interest rate are discussed in detail in Chapter 18. As an example we can consider a loan where the interest is paid at the end of a given period, and the interest is quoted in annual rate. Rate conventions determine how the quoted annual rate relates to the actual payment. Maybe the most common convention is to pay $P \times rT/360$, where P is the principal, r is the annual rate, and T is the number of calendar days of the deposit or loan. Note that loan rates are either rates that apply to a loan starting now until a given expiry, or forward rates, that are rates applying to a loan starting in the future for a given period of time.

Rates are quoted in percents but they are compared in basis points, where a basis point is 0.01%, that is, 1% is 100 basis points.

2.2.2.2 The Risk Free Rate

The risk free rate is different depending on the investment horizon. For one day horizon the risk free rate could be the Eonia rate or the rate of a bank account, and for 1 month horizon the risk free rate could be the rate of 1 month government bond.

2.2.3 Bond Prices and Returns

A 10 year zero-coupon bond has the time to maturity of 10 years at the emission, after 1 year the time to maturity is 9 years, after 2 years the time to maturity is 8 years, and so on. The price of the zero-coupon bond is fluctuating according to the fluctuation of the interest rates, until the price equals the nominal value at the maturity. Thus, the price of the 10 year zero-coupon bond gives information about the 10 year interest rate at the emission, after 1 year the price of the bond gives information about the 9 year interest rate, after 2 years the price of the bond gives information about the 8 year interest rate, and so on.

Information of the bond markets is given by data providers in terms of the yields. The yield of a zero-coupon bond is defined as

$$Y(t,T) = -\frac{1}{T-t} \log Z(t,T), \tag{2.3}$$

where $T-t$ is the time to maturity in fractions of a year, and $Z(t,T)$ is the bond price with $Z(T,T)=1$. The price of a bond can be written in terms the yield as

$$Z(t,T) = \exp\{-(T-t)Y(t,T)\}.$$

See Section 18.1.2 for a discussion of the yield of a zero-coupon bond.

Let $s < t \le T$, where T is the expiration day of the zero-coupon bond. The prices are $Z(s,T)$ and $Z(t,T)$. The return of a bond trader is equal to

$$\begin{aligned}\frac{Z(t,T)}{Z(s,T)} &= \frac{\exp\{-(T-t)Y(t,T)\}}{\exp\{-(T-s)Y(s,T)\}}\\ &= \exp\{(T-s)[Y(s,T)-Y(t,T)] + (t-s)Y(t,T)\},\end{aligned} \tag{2.4}$$

where we used the fact $T-t = T-s-(t-s)$.

Data providers give a time series $Y_0,\ldots,Y_n$ of yields of a τ year bond, where

$$Y_i = -\frac{1}{\tau} \log Z(t_i, t_i+\tau),$$

where $t_0 < \cdots < t_n$ are the time points of sampling. How to obtain a time series $R_0,\ldots,R_n$ of the returns of a bond investor? Let us denote $t_i = s$, $t_{i+1} = t$, and $T-s=\tau$. Then $Y(s,T) = Y_i$. Let us make approximation

$$Y(t,T) = Y(t_{i+1}, t_i+\tau) \approx Y(t_{i+1}, t_{i+1}+\tau) = Y_{i+1}.$$

Then (2.4) implies

$$R_i \approx \exp\{\tau(Y_i - Y_{i+1}) + (t_{i+1}-t_i)Y_{i+1}\}, \tag{2.5}$$

where $t_{i+1}-t_i$ is the length of the sampling interval in fractions of a year. For example, with monthly sampling $t_{i+1}-t_i = 1/12$.

2.3 Derivatives

Derivatives are financial assets whose payoff is defined in terms of more basic assets. We describe first forwards and futures, and after that we describe options. For many assets trading with derivatives is more active than trading with the basic assets. For example, exchange rates and commodities are traded more actively in the future markets than in the spot markets.

Over-the-counter (OTC) derivatives are traded directly between two counterparties. Exchange traded derivatives are traded in an exchange, which acts as an intermediary party between the traders.

2.3.1 Forwards and Futures

First we define forwards and futures. After that we give examples of some actively traded futures. Forwards are derivatives traded over the counter whereas futures contracts are traded on exchanges. The underlyings of a forward or a futures contract can be stocks (single-stock futures), commodities, currencies, interest rates, or stock indexes, for example.

2.3.1.1 Forwards

A forward is a contract written at time t_0, with a commitment to accept delivery of (or to deliver) the specified number of units of the underlying asset at a future date T, at forward price F_{t_0}, which is determined at t_0.

At time t_0 nothing changes hands, all exchanges will take place at time T. A long position is a commitment to accept the delivery at time T. A short position is a commitment to deliver the contracted amount. The current price of the underlying is called the spot price.

2.3.1.2 Futures

A futures contract can be considered as a special case of a forward contract. An instrument is called a futures contract if the trading is done in a futures exchange, where the forward commitment is made through a homogenized contract so that the size of the underlying asset, the quality of the underlying asset, and the expiration date are preset. In addition, futures exchanges require a daily mark-to-market of the positions.

A futures exchange acts as an intermediary between the participants of a futures contract. The existence of the intermediary minimizes the risk of the default of the participants of the contract. When a participant enters a futures contract the exchange requires to put up an initial amount of liquid assets into the margin account. Marking to market means that the daily futures price is settled daily so that the exchange will draw money out of one party's margin account and put it into the others so that the daily loss or profit is taken into account. If the margin account goes below a certain value, then a margin call is made and the account owner must add money to the margin account. In contrast to futures contracts, forward contracts may not require any marking to market until the expiration day.

A futures contract can be settled with cash or with the delivery of the underlying. For example, if the underlying of the futures contract is a stock index, then the futures contract is usually settled with cash. A futures contract can be closed before the expiration day by entering the opposite direction futures contract.

On the delivery date, the amount exchanged is not the specified price on the contract but the spot value (i.e., the original value agreed upon, since any gain or loss has already been previously settled by marking to market).

The situation where the price of a commodity for future delivery is higher than the spot price, or where a far future delivery price is higher than a nearer future delivery, is known as contango. The reverse, where the price of a commodity for future delivery is lower than the spot price, or where a far future delivery price is lower than a nearer future delivery, is known as backwardation.

2.3.2 Options

We describe calls and puts, applications of options, and some exotic options.

2.3.2.1 Calls and Puts

The buyer of a call option receives the right to buy the underlying instrument and the buyer of a put option receives the right to sell the underlying instrument.

An European call option gives the right to buy an asset at the given expiration time T at the given strike price K. An European put option gives the right to sell an asset at the given expiration time T at the given strike price K. Let us denote with C_t the price of an European call option at time t and with S_t the price of the asset. The value C_T of the European call option at the expiration time T is equal to

$$C_T = \max\{S_T - K, 0\}.$$

Let us denote with P_t the price of a put option at time t. The value of the European put option at the expiration time T is equal to

$$P_T = \max\{K - S_T, 0\}.$$

American options have a different mode concerning the right to exercise the option than the European options. American call and put options can be exercised at any time before the expiration date, whereas European options can be exercised only at the expiration day. Thus an American option is more expensive than the corresponding European option. When we use the term “option” without a further qualification, then we refer to an European option.

The following terminology is used to describe options.

- A call option is out of the money if $S_t < K$. A call option is at the money if $S_t = K$. A call option is in the money if $S_t > K$. A call option is deep out of the money (deep in the money) if $S_t \ll K$ ($S_t \gg K$).
  The moneyness of a call option is defined as S_t/K. The moneyness of a put option is defined as K/S_t.[7]
- Before the expiration time T the price of a call option satisfies

$$C_t > (S_t - K)_+;$$

see (14.10). The difference $C_t - (S_t - K)_+$ is called the time value of the option. The value $(S_t - K)_+$ is called the intrinsic value. Thus,

$$C_t = \text{time value} + \text{intrinsic value}. \tag{2.6}$$

2.3.2.2 Applications of Options

Options can serve at least the following purposes:

1) Options can be used to create a large number of different payoffs. Some payoffs applied in option trading are described in Chapter 17. For example, buying a call and a put with the same strike price and the same expiration creates a straddle position which profits from large positive or negative movements of the underlying.
2) Options can provide insurance. With options it is possible to create a payoff which cuts the losses that could occur without using of the options. Buying a put option gives an insurance in the case one has to sell in a future time an asset one possesses. Buying a call option gives an insurance in the case when one has to buy in a future time an asset one does not possess. Examples of providing insurance with options include the following:
   - Buying a put option on a stock gives an insurance policy for an investor. If an investor owns a stock, buying a put option will cut the future possible losses.
   - Buying a put option on an exchange rate gives an insurance policy for a company receiving payments on a foreign currency in future.
3) Call options can be used to give a compensation to managers, since the payoff of a call option is positive only when the stock price is larger than the strike price.
4) Options make leveraging possible, since option trading requires a small initial capital as compared to stock trading.[8]

---

7 Sometimes moneyness is defined by $S_t/(Ke^{-r(T-t)})$ and $Ke^{-r(T-t)}/S_t$, where $T - t$ is the time to expiration in fractions of year and r is the annualized short term interest rate.

8 Suppose that the stock price is $S_t = 100$, the strike price is $K = 105$, and the call price is $C_t = 5$. If the stock price rises to $S_T = 110$ at the expiration time of the call option, then the owner of the stock has the return of 10% but the owner of the call option has the return of $(110 - 105)/5 = 100\%$.

2.3.2.3 Exotic Options

We say that an option is exotic if it is not an European or an American call or put option.

Bermudan Options There exists three basic modes concerning the right to exercise the option: European, American, and Bermudan. A Bermudan option can be exercised at some times or time periods before the expiration. whereas European options can be exercised only at the expiration date, and American options can be exercised at any time before the expiration.

Asian Options The value of an Asian call option at the expiration is

$$C_T = \max\{0, M_T - K\},$$

where

$$M_T = \frac{1}{n}\sum_{i=1}^{n} S_{t_i}$$

with $t_1 < \cdots < t_n \le T$ being a collection of predetermined time points. Asian options are more resistant to manipulation than European options: The value of an European option at the expiration depends on the value of the underlying asset at one time point (the expiration date), whereas the value of an Asian option depends on the values of the underlying asset at several time points.

Barrier Options Barrier options disappear if the underlying either exceeds, or goes under the barrier. Alternatively, a barrier option could have value only if it has exceeded, or went under the barrier. Knock-in options come into existence if some barrier is hit and knock-out options cease to exist if some barrier is hit. One speaks of up-and-out, down-and-out, up-and-in, down-and-in options. For example, a knock-out option on stock S_t, written at time 0, with expiration time T, has the payoff

$$B_T = \begin{cases} 0, & \text{when } \max_{0\le t\le T} S_t \ge H, \\ \max\{0, S_T - K\}, & \text{otherwise,} \end{cases}$$

where $K > 0$ is the strike price, and $H > K$ is the barrier. Barrier options are cheaper than the corresponding European options, which makes them useful.

Multiasset Options Multiasset options involve many underlying assets and many strike prices. We give some examples of multiasset options.

1) A call can be generalized to a multiasset option with payoff

$$\max\left\{S_T^1 - K_1, S_T^2 - K_2, 0\right\},$$

where S^1 and S^2 are the underlying assets and $K_1, K_2 > 0$ are strike prices. A payoff can have elements of a call and a put:

$$\max\left\{K_1 - S_T^1, S_T^2 - K_2, 0\right\}.$$

2) The payoff of an option on a linear combination can be written as

$$f\left(\sum_{i=1}^{d} w_i S^i\right),$$

where $f : \mathbf{R} \to \mathbf{R}$ is a payoff function, $S^1, \dots, S^d$ are assets, and $w_1, \dots, w_d$ are weights. For example, an option on a linear combination can be an option on an index or an option on a spread.

3) Outperformance options are calls on the maximum and puts on the minimum. We have that

$$\max\left\{S^1, S^2\right\} = S^1 + \max\left\{S^2 - S^1, 0\right\}.$$

Thus, the payoff of an outperformance option can be written as a payoff of a linear combination of the underlying and an option on the spread between the underlyings.

4) The payoff of a univariate digital option is $I_{[K,\infty)}(S_T)$, where $K > 0$ is the strike price. The option pays one unit at the maturity time if the value of the underlying exceeds the strike price. The bivariate digital option pays one unit if both of the underlyings exceed the respective strike prices. The payoff is

$$I_{[K_1,\infty)\times[K_2,\infty)}\left(S_T^1, S_T^2\right).$$

5) The payoff of an option written on a basket can be written as

$$G\left(\psi\left(S_T^1, \dots, S_T^N\right)\right),$$

where G is a univariate function and ψ is a multivariate function. For example, $G(x) = (x - K)_+$ and $\psi(S^1, \dots, S^d) = \min(S^1, \dots, S^d)$, or $\psi(S^1, \dots, S^d) = \sum_{i=1}^d w_i S^i$.

2.4 Data Sets

We describe the data sets which are used to illustrate the methods throughout the book. Some additional data are described in Section 6.3.

2.4.1 Daily S&P 500 Data

The daily S&P 500 data consists of the daily closing prices starting at January 4, 1950 and ending at April 2, 2014, which gives 16,046 daily observations.[9]

Figure 2.1 shows (a) the daily closing prices S_t and (b) the returns $R_t = S_t/S_{t-1}$ of S&P 500.

9 The data is obtained from Yahoo (http://finance.yahoo.com/) with ticker ^GSPC.

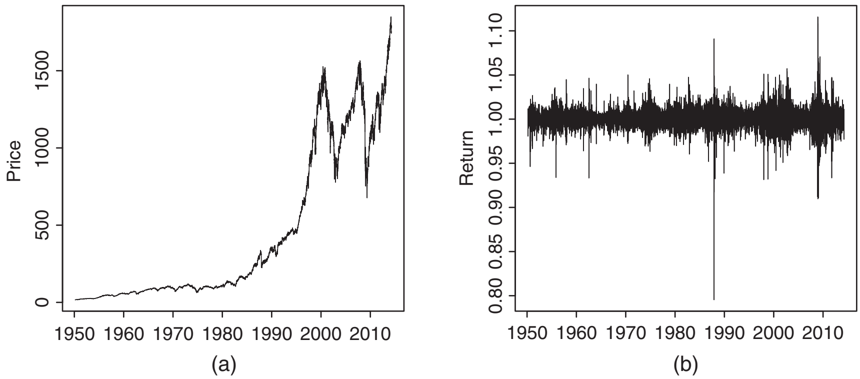


Figure 2.1 *S&P 500 index*. (a) Daily closing prices of S&P 500 and (b) daily returns.

2.4.2 Daily S&P 500 and Nasdaq-100 Data

The S&P 500 and Nasdaq-100 data consists of the daily closing prices starting at October 1, 1985 and ending at May 21, 2014, which gives 7221 daily observations.[10]

Figure 2.2 shows (a) the normalized prices and (b) a scatter plot of the returns of S&P 500 and Nasdaq-100. S&P 500 prices is shown with black and the Nasdaq-100 prices is shown with red. The prices are normalized so that they start with value one for both indexes. (Note that the normalized price is the cumulative wealth when the initial wealth is one.)

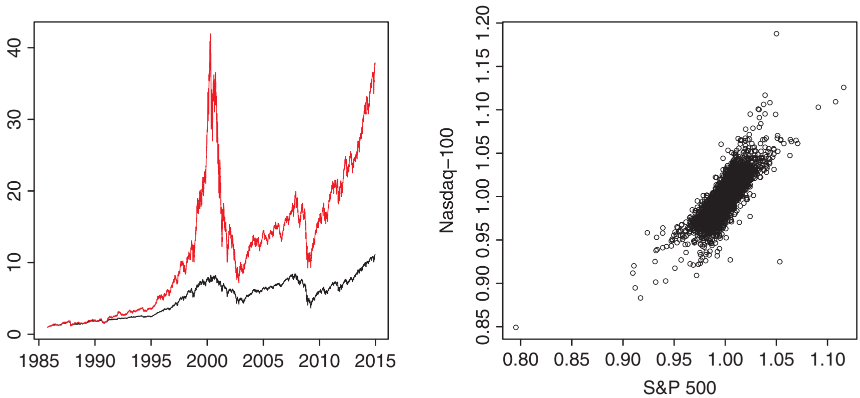
2.4.3 Monthly S&P 500, Bond, and Bill Data

The data consists of the monthly returns of S&P 500 index, monthly returns of US Treasury 10 year bond, and monthly rates of US Treasury 1 month bill. The data starts at May 1953 and ends at December 2013, which gives 728 monthly observations.[11] The 10 year bond returns are calculated from the yields as in (2.5).

Figure 2.3 shows (a) cumulative wealth and (b) a scatter plot of returns of S&P 500 and 10 year bond. The cumulative wealth is $W_t = \prod_{i=1}^{t} R_i$, where R_i are the gross returns. The cumulative wealth of S&P 500 is shown with black, 10 year bond with red, and 1 month bill with blue. Figure 2.4 shows (a) the treasury bill rates (blue) and (b) the yields of 10 year Treasury bond (red).

10 The data is obtained from Yahoo (http://finance.yahoo.com/) with tickers ^GSPC and ^NDX.

11 The data is obtained from http://www.hec.unil.ch/agoyal/ (Amit Goyal).

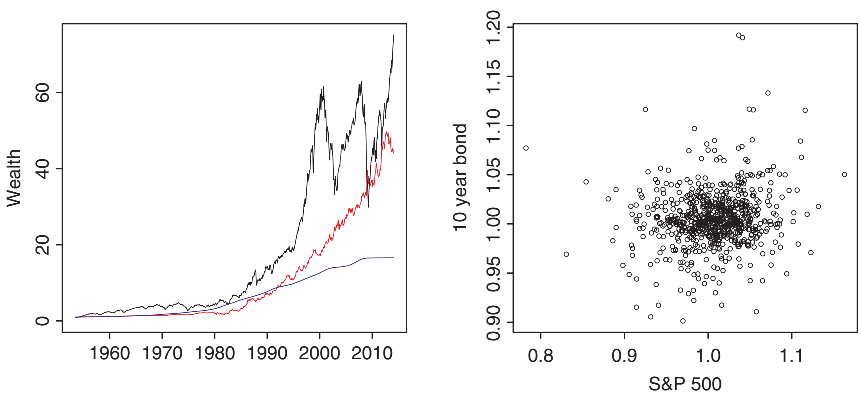


(a) (b)

Figure 2.2 *S&P 500 and Nasdaq-100 indexes*. (a) The prices of S&P 500 (black) and Nasdaq-100 (red). The prices are normalized to start at value one. (b) A scatter plot of the daily returns of S&P 500 and Nasdaq-100.

2.4.4 Daily US Treasury 10 Year Bond Data

The US Treasury 10 year bond data consists of the daily yields starting at January 2, 1962 and ending at March 3, 2014, which gives 13,006 daily observations.[12] We have described the US 10 year Treasury bonds in Section 2.2.1.



(a) (b)

Figure 2.3 *S&P 500, US Treasury 10 year bond, and 1 month bill*. (a) The cumulative wealth of S&P 500 (black), 10 year bond (red), and 1 month bill (blue). The cumulative wealths are normalized to start at value one. (b) A scatter plot of monthly returns of S&P 500 and 10 year bond.

12 The data is obtained from Federal Reserve Bank of St. Louis with ticker DGS10, see the web site http://research.stlouisfed.org/. There were 13,590 days when the market is open but the data was missing in 584 days.

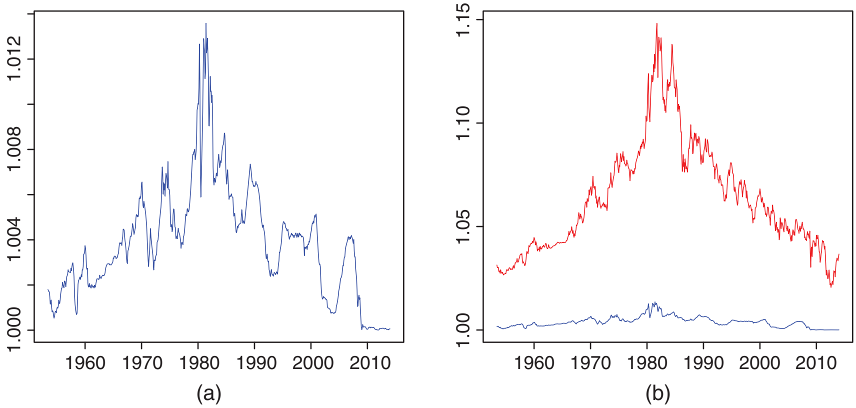


Figure 2.4 *US Treasury bill rates and 10 year bond yields*. (a) Treasury bill rates (blue). (b) Yields of 10 year Treasury bond (red).

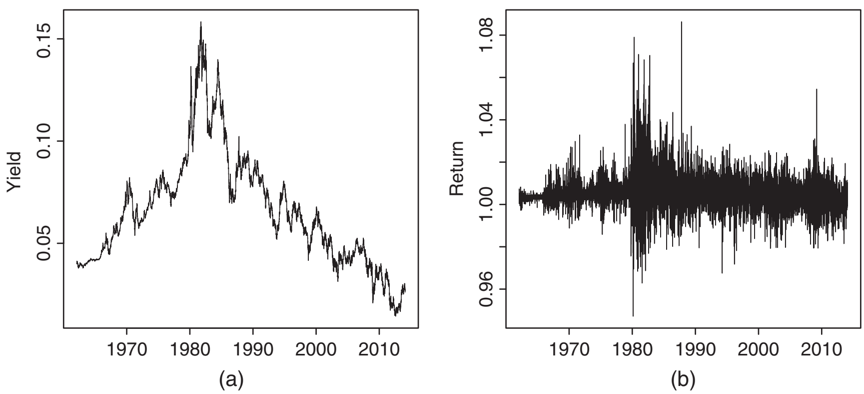


Figure 2.5 *10 year US Treasury bond*. (a) Daily yields of the 10 year US Treasury bond and (b) daily returns of the bond.

Figure 2.5 shows (a) the daily yields and (b) the daily returns of the US 10 year Treasury bond. The 10 year bond returns are calculated from the yields as in (2.5).

2.4.5 Daily S&P 500 Components Data

The S&P 500 components data consists of daily closing prices of 312 stocks, which were components of S&P 500 at May 23, 2014. The data starts September 30, 1991 and ends at May 23, 2014. There are 5707 daily observations.

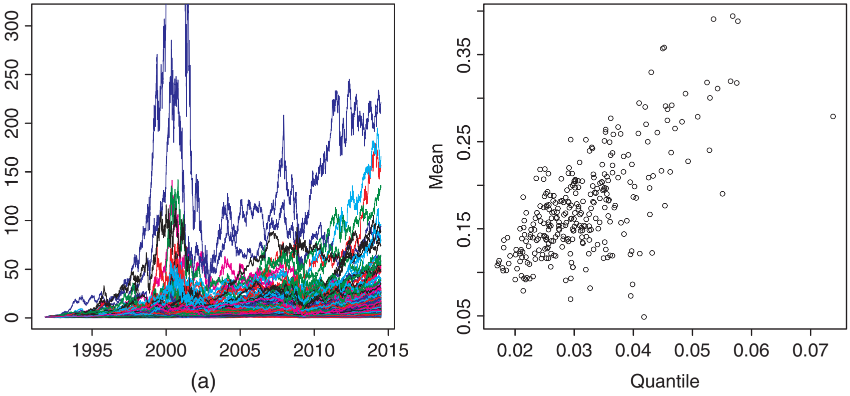


Figure 2.6 *S&P 500 components*. (a) Time series of the normalized prices of the components. (b) A scatter plot of (q_i, μ_i), where q_i are the 95% empirical quantiles of the negative returns, and μ_i are the annualized sample means of the returns.

Figure 2.6(a) shows the normalized prices of the stocks. The prices are normalized to have value one at the beginning. Panel (b) shows a scatter plot of points (q_i, μ_i), where q_i are the 95% empirical quantiles of the negative returns of the ith stock, and μ_i are the annualized sample means of the returns of the ith stock.[13]

13 That is, q_i satisfies approximately $P(R_t^i \leq -q_i) = 0.05$, where $R_t^i = S_t/S_{t-1} - 1$ is the net return of the ith stock, and μ_i is approximately $250 \times ER_t^i$.

3

Univariate Data Analysis

Univariate data analysis studies univariate financial time series, but ignoring the time series properties of data. Univariate data analysis studies also cross-sectional data. For example, returns at a fixed time point of a collection of stocks is a cross-sectional univariate data set.

A univariate series of observations can be described using such statistics as sample mean, median, variance, quantiles, and expected shortfalls. These are covered in Section 3.1.

The graphical methods are explained in Section 3.2. Univariate graphical tools include tail plots, regression plots of the tails, histograms, and kernel density estimators. We use often tail plots to visualize the tail parts of the distribution, and kernel density estimates to visualize the central part of the distribution. The kernel density estimator is not only a visualization tool but also a tool for estimation.

We define univariate parametric models like normal, log-normal, and Student models in Section 3.3. These are parametric models, which are alternatives to the use of the kernel density estimator.

For a univariate financial time series it is of interest to study the tail properties of the distribution. This is done in Section 3.4. Typically the distribution of a financial time series has heavier tails than the normal distributions. The estimation of the tails is done using the concept of the excess distribution. The excess distribution is modeled with exponential, Pareto, gamma, generalized Pareto, and Weibull distributions. The fitting of distributions can be done with a version of maximum likelihood. These results prepare us to quantile estimation, which is considered in Chapter 8.

Central limit theorems provide tools to construct confidence intervals and confidence regions. The limit theorems for maxima provide insight into the estimation of the tails of a distribution. Limit theorems are covered in Section 3.5.

Section 3.6 summarizes the univariate stylized facts.

Nonparametric Finance, First Edition. Jussi Klemelä.

3.1 Univariate Statistics

We define mean, median, and mode to characterize the center of a distribution. The spread of a distribution can be measured by variance, other centered moments, lower and upper partial moments, lower and upper conditional moments, quantiles (value-at-risk), expected shortfall, shortfall, and absolute shortfall.

We define both population and sample versions of the statistics. In addition, we define both unconditional and conditional versions of the statistics.

3.1.1 The Center of a Distribution

The center of a distribution can be defined using the mean, the median, or the mode. The center of a distribution is an unknown quantity that has to be estimated using the sample mean, the sample median, or the sample mode. The conditional versions of theses quantities take into account the available information. For example, if we know that it is winter, then the expected temperature is lower than the expected temperature when we know that it is summer.

3.1.1.1 The Mean and the Conditional Mean

The population mean is called the expectation. The population mean can be estimated by the arithmetic mean. The conditional mean is estimated using regression analysis.

The Population Mean The population mean (expectation) of random variable $Y \in \mathbf{R}$, whose distribution is continuous, is defined as

$$EY = \int_{-\infty}^{\infty} y f_Y(y)\, dy, \tag{3.1}$$

where $f_Y : \mathbf{R} \to \mathbf{R}$ is the density function of Y.[1] Let $X \in \mathbf{R}^d$ be an explanatory random variable (random vector). The conditional expectation of Y given $X = x$ can be defined by

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y\, f_{Y \mid X=x}(y)\, dy,$$

where $f_{Y \mid X=x}(y) : \mathbf{R} \to \mathbf{R}$ is the conditional density.[2]

1 The density function $f_Y : \mathbf{R} \to \mathbf{R}$ is a function which satisfies (1) $f_Y(y) \geq 0$ for almost all $y \in \mathbf{R}$, and (2) $P(Y \in A) = \int_A f_Y(y)\, dy$ for measurable $A \subset \mathbf{R}$. Thus, we can express all probabilities as integrals of f_Y.

2 The conditional density is defined as

$$f_{Y \mid X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}, \quad y \in \mathbf{R}, \quad x \in \mathbf{R}^d, \tag{3.2}$$

The population mean of random variable $Y \in \mathbf{R}$, whose distribution is discrete with the possible values $y_1, \dots, y_N$, is defined as

$$EY = \sum_{i=1}^{N} y_i P(Y = y_i). \tag{3.3}$$

The conditional expectation can be defined as

$$E(Y \mid X = x) = \sum_{i=1}^{N} y_i P(Y = y_i \mid X = x).$$

The Sample Mean Given a sample $Y_1, \dots, Y_T$ from the distribution of Y, the mean EY can be estimated with the sample mean (the arithmetic mean):

$$\bar{Y} = \frac{1}{T} \sum_{t=1}^{T} Y_t. \tag{3.4}$$

Regression analysis studies the estimation of the conditional expectation. In regression analysis, we observe values $X_1, \dots, X_T$ of the explanatory random variable (random vector), in addition to observing values $Y_1, \dots, Y_T$ of the response variable. Besides linear regression there exist various nonparametric methods for the estimation of the conditional expectation. For example, in kernel regression the arithmetic mean in (3.4) is replaced by a weighted mean

$$\hat{f}(x) = \sum_{t=1}^{T} p_t(x) Y_t,$$

where $p_t(x)$ is a weight that is large when X_t is close to x and small when X_t is far away from x. Now $\hat{f}(x)$ is an estimate of the conditional mean $f(x) = E(Y \mid X = x)$, for $x \in \mathbf{R}^d$. Kernel regression and other regression methods are described in Section 6.1.2.

The Annualized Mean The return of a portfolio is typically estimated using the arithmetic mean and it is expressed as the annualized mean return. Let $S_{t_0}, \dots, S_{t_n}$ be observed stock prices, sampled at equidistant time points. Let $R_{t_i} = (S_{t_i} - S_{t_{i-1}})/S_{t_{i-1}}$, $i = 1, \dots, n$, be the net returns. Let the sampling interval be $\Delta t = t_i - t_{i-1}$. The annualized mean return is

$$\frac{1}{\Delta t} \frac{1}{n} \sum_{i=1}^{n} R_{t_i}. \tag{3.5}$$

---

when $f_X(x) > 0$, where $f_{X,Y} : \mathbf{R}^{d+1} \to \mathbf{R}$ is the joint density of (X, Y), and $f_X : \mathbf{R}^d \to \mathbf{R}$ is the density of X:

$$f_X(x) = \int_{\mathbf{R}} f_{X,Y}(x, y)\, dy, \quad x \in \mathbf{R}^d.$$

If $f_X(x) = 0$, then $f_{Y \mid X=x}(y) = 0$.

For the monthly returns $\Delta t = 1/12$. For the daily returns $\Delta t = 1/250$, because there are about 250 trading days in a year. Sampling of prices and several definitions of returns are discussed in Section 2.1.2.

The Geometric Mean Let $S_0, \dots, S_T$ be the observed stock prices and let $R_t = S_t/S_{t-1}$, $t = 1, \dots, T$, be the gross returns. The geometric mean is defined as

$$\left(\prod_{t=1}^{T} R_t\right)^{1/T}.$$

The logarithm of the geometric mean is equal to the arithmetic mean of the logarithmic returns:

$$\frac{1}{T}\sum_{t=1}^{T} \log R_t.$$

Note that $W_t = \prod_{i=1}^{t} R_i$ is the cumulative wealth at time t when we start with wealth 1. Thus,

$$\frac{1}{T}\log W_T = \frac{1}{T}\sum_{t=1}^{T} \log R_t.$$

3.1.1.2 The Median and the Conditional Median

The median can be defined in the case of a continuous distribution function of a random variable $Y \in \mathbf{R}$ as the number median$(Y) \in \mathbf{R}$ satisfying

$$P(Y \le \text{median}(Y)) = 0.5.$$

Thus, the median is the point that divides the probability mass into two equal parts. Let us define the distribution function $F : \mathbf{R} \to \mathbf{R}$ by

$$F(y) = P(Y \le y).$$

When F is continuous, then

$$\text{median}(Y) = F^{-1}(0.5).$$

In general, covering also the case of discrete distributions, we can define the median uniquely as the generalized inverse of the distribution function:

$$\text{median}(Y) = \inf\{y : F(y) \ge 0.5\}. \tag{3.6}$$

The conditional median is defined using the conditional distribution function

$$F_{Y|X=x}(y) = P(Y \le y \,|\, X = x),$$

where X is a random vector taking values in $\mathbf{R}^d$. Now we can define

$$\text{median}(Y \,|\, X = x) = \inf\{y : F_{Y|X=x}(y) \ge 0.5\}, \tag{3.7}$$

where $x \in \mathbf{R}^d$.

The sample median of observations $Y_1, \dots, Y_T \in \mathbf{R}$ can be defined as the observation that has as many smaller observations as larger observations:

$$\text{median}(Y_1, \dots, Y_T) = Y_{([T/2]+1)}, \tag{3.8}$$

where $Y_{(1)} \le \cdots \le Y_{(T)}$ is the ordered sample and $[x]$ is the largest integer smaller or equal to x. The sample median is a special case of an empirical quantile. Empirical quantiles are defined in (8.21)–(8.23).

3.1.1.3 The Mode and the Conditional Mode

The mode is defined as an argument maximizing the density function of the distribution of a random variable:

$$\text{mode}(Y) = \underset{y \in \mathbf{R}}{\text{argmax}}\, f_Y(y), \tag{3.9}$$

where $f_Y : \mathbf{R} \to \mathbf{R}$ is the density function of the distribution of Y. The density f_Y can have several local maxima, and the use of the mode seems to be interesting only in cases where the density function is unimodal (has one local maximum). The conditional mode is defined as an argument maximizing the conditional density:

$$\text{mode}(Y \,|\, X = x) = \underset{y \in \mathbf{R}}{\text{argmax}}\, f_{Y|X=x}(y).$$

A mode can be estimated by finding a maximizer of a density estimate:

$$\widehat{\text{mode}}(Y) = \underset{y \in \mathbf{R}}{\text{argmax}}\, \hat{f}_Y(y),$$

where $\hat{f}_Y : \mathbf{R} \to \mathbf{R}$ is an estimator of the density function f_Y. Histograms and kernel density estimators are defined in Section 3.2.2.

3.1.2 The Variance and Moments

Variance and higher order moments characterize the dispersion of a univariate distribution. To take into account only the left or the right tail we define upper and lower partial moments and upper and lower conditional moments.

3.1.2.1 The Variance and the Conditional Variance

The variance of random variable Y is defined by

$$\text{Var}(Y) = E(Y - EY)^2 = EY^2 - (EY)^2. \tag{3.10}$$

The standard deviation of Y is the square root of the variance of Y. The conditional variance of random variable Y is equal to

$$\text{Var}(Y \,|\, X = x) = E\{[Y - E(Y \,|\, X = x)]^2 \,|\, X = x\} \tag{3.11}$$

$$= E(Y^2 \,|\, X = x) - [E(Y \,|\, X = x)]^2. \tag{3.12}$$

The conditional standard deviation of Y is the square root of the conditional variance.

The Sample Variance The sample variance is defined by

$$\widehat{\mathrm{Var}}(Y) = \frac{1}{T}\sum_{i=1}^{T}(Y_i - \bar{Y})^2 = \frac{1}{T}\sum_{i=1}^{T} Y_i^2 - \bar{Y}^2, \tag{3.13}$$

where $Y_1, \dots, Y_T$ is a sample of random variables having identical distribution with Y, and $\bar{Y}$ is the sample mean.[3]

The Annualized Variance The sample variance and the standard deviation of portfolio returns are typically annualized, analogously to the annualized sample mean in (3.5). Let $S_{t_0}, \dots, S_{t_n}$ be the observed stock prices, sampled at equidistant time points. Let $R_{t_i} = (S_{t_i} - S_{t_{i-1}})/S_{t_{i-1}}$, $i = 1, \dots, n$, be the net returns. Let the sampling interval be $\Delta t = t_i - t_{i-1}$. The annualized sample variance of the returns is

$$\frac{1}{\Delta t}\,\frac{1}{n}\sum_{i=1}^{n}(R_{t_i} - \bar{R})^2,$$

where $\bar{R} = n^{-1}\sum_{i=1}^{n} R_{t_i}$. For the monthly returns $\Delta t = 1/12$. For the daily returns $\Delta t = 1/250$, because there are about 250 trading days in a year. Sampling of prices and several definitions of returns are discussed in Section 2.1.2.

3.1.2.2 The Upper and Lower Partial Moments

The definition of the variance of random variable $Y \in \mathbf{R}$ can be generalized to other centered moments

$$E|Y - EY|^k,$$

for $k = 1, 2, \dots$. The variance is obtained when $k = 2$. The centered moments take a contribution both from the left and the right tail of the distribution. The lower partial moments take a contribution only from the left tail and the upper partial moments take a contribution only from the right tail. For example, if we are interested only in the distribution of the losses, then we use the lower partial moments of the return distribution, and if we are interested only in the distribution of the gains, then we use the upper partial moments. The upper partial moment is defined as

$$\mathrm{UPM}_{\tau,k}(Y) = E(Y - \tau)_+^k = E[(Y - \tau)^k I_{[\tau,\infty)}(Y)], \tag{3.14}$$

where $k = 0, 1, 2, \dots$, $(x)_+ = \max\{x, 0\}$, and $\tau \in \mathbf{R}$. The lower partial moment is defined as

$$\mathrm{LPM}_{\tau,k}(Y) = E(\tau - Y)_+^k = E[(\tau - Y)^k I_{(-\infty,\tau]}(Y)]. \tag{3.15}$$

3 The sample variance is often defined as $(T-1)^{-1}\sum_{i=1}^{T}(Y_i - \bar{Y})^2$, because this is an unbiased estimator of the population variance. For large and moderate T it does not matter whether the divisor is T or $T - 1$.

When Y has density f_Y, we can write

$$\mathrm{UPM}_{\tau,k}(Y) = \int_\tau^\infty (y-\tau)^k f_Y(y)\,dy, \quad \mathrm{LPM}_{\tau,k}(Y) = \int_{-\infty}^\tau (\tau-y)^k f_Y(y)\,dy.$$

For example, when $k = 0$, then

$$\mathrm{UPM}_{\tau,0}(Y) = P(Y \geq \tau), \quad \mathrm{LPM}_{\tau,0}(Y) = P(Y \leq \tau),$$

so that the upper partial moment is equal to the probability that Y is greater or equal to τ, and the lower partial moment is equal to the probability that Y is smaller or equal to τ. For $k = 2$ and $\tau = EY$ the partial moments are called the upper and lower semivariance of Y. For example, the lower semivariance is defined as

$$E[(Y - EY)^2 I_{(-\infty,EY]}(Y)]. \tag{3.16}$$

The square root of the lower semivariance can be used to replace the standard deviation in the definition of the Sharpe ratio, or in the Markowitz criterion.

The sample centered moments are

$$\frac{1}{T}\sum_{t=1}^{T} |Y_i - \bar{Y}|^k,$$

where $\bar{Y}$ is the sample mean. The sample upper and the sample lower partial moments are

$$\widehat{\mathrm{UPM}}_{\tau,k}(Y) = \frac{1}{T}\sum_{i=1}^{T}(Y_i-\tau)_+^k, \quad \widehat{\mathrm{LPM}}_{\tau,k}(Y) = \frac{1}{T}\sum_{i=1}^{T}(\tau - Y_i)_+^k. \tag{3.17}$$

For example, when $k = 0$ we have

$$\widehat{\mathrm{LPM}}_{\tau,0}(Y) = \frac{N(\tau)}{T},$$

where

$$N(\tau) = \#\{Y_i \,:\, i = 1,\ldots,T, \quad Y_i \leq \tau\}. \tag{3.18}$$

3.1.2.3 The Upper and Lower Conditional Moments

The upper conditional moments are the moments conditioned on the right tail of the distribution and the lower conditional moments are the moments conditioned on the left tail of the distribution. The upper conditional moment is defined as

$$\mathrm{UCM}_{\tau,k}(Y) = E[(Y-\tau)^k \mid Y - \tau \geq 0]$$

and the lower conditional moment is defined as

$$\mathrm{LCM}_{\tau,k}(Y) = E[(\tau - Y)^k \mid \tau - Y \geq 0], \tag{3.19}$$

where $k = 0, 1, 2, \ldots$ and $\tau \in \mathbf{R}$ is a target rate.

The sample lower conditional moment is

$$\widehat{\mathrm{LCM}}_{\tau,k}(Y) = \frac{1}{N(\tau)} \sum_{i=1}^{T} (\tau - Y_i)_+^k, \tag{3.20}$$

where $N(\tau)$ is defined in (3.18). Note that in (3.17) the sample size is the denominator but in (3.20) we have divided with the number of observations in the left tail.

We can condition also on an external variable X and define conditional on X versions of both upper and lower moments, and upper and lower conditional moments.

3.1.3 The Quantiles and the Expected Shortfalls

The quantiles are applied under the name value-at-risk in risk management to characterize the probability of a tail event. The expected shortfall is a related measure for a tail risk.

3.1.3.1 The Quantiles and the Conditional Quantiles

The pth quantile is defined as

$$Q_p(Y) = \inf\{y : F(y) \geq p\}, \tag{3.21}$$

where $0 < p < 1$ and $F(y) = P(Y \leq y)$ is the distribution function of Y. The value-at-risk is defined in (8.3) as a quantile of a loss distribution. For $p = 1/2$, $Q_p(Y)$ is equal to median(Y), defined in (3.6). In the case of a continuous distribution function, we have

$$F(Q_p(Y)) = p$$

and thus it holds that

$$Q_p(Y) = F^{-1}(p),$$

where F^{-1} is the inverse of F. The pth conditional quantile is defined replacing the distribution function of Y with the conditional distribution function of Y:

$$Q_p(Y \mid X = x) = \inf\{y : F_{Y|X=x}(y) \geq p\}, \quad x \in \mathbf{R}^d, \tag{3.22}$$

where $0 < p < 1$ and $F_{Y|X=x}(y) = P(Y \leq y \mid X = x)$ is the conditional distribution function of Y.

The empirical quantile is defined as

$$\hat{Q}_p = Y_{(\lceil pT \rceil)}, \tag{3.23}$$

where $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(T)}$ is the ordered sample and $\lceil x \rceil$ is the smallest integer $\ge x$. We give equivalent definitions of the empirical quantile in Section 8.4.1. Chapter 8 discusses various estimators of quantiles and conditional quantiles.

3.1.3.2 The Expected Shortfalls

The expected shortfall is a measure of risk that aggregates all quantiles in the right tail (or in the left tail). When Y has a continuous distribution function, then the expected shortfall for the right tail is

$$\mathrm{ES}_p(Y) = E(Y \mid Y \ge Q_p(Y)) = \frac{1}{1-p}\, E(Y I_{[Q_p(Y),\infty)}(Y)), \tag{3.24}$$

where $0 < p < 1$. Thus, the pth expected shortfall is the conditional expectation under the condition that the random variable is larger than the pth quantile. The term "tail conditional value-at-risk" is sometimes used to denote the expected shortfall. In the general case, when the distribution of Y is not necessarily continuous, the expected shortfall for the right tail is defined as

$$\mathrm{ES}_p(Y) = \frac{1}{1-p} \int_p^1 Q_u(Y)\, du, \quad 0 < p < 1. \tag{3.25}$$

The equality of (3.24) and (3.25) for the continuous distributions is proved in McNeil *et al.* (2005, lemma 2.16). In fact, denoting $q_p = Q_p(Y)$,

$$\begin{aligned} E[Y I_{(q_p,\infty)}(Y)] &= E[F^{-1}(U) I_{[q_p,\infty)}(F^{-1}(U))] \\ &= E[F^{-1}(U) I_{[p,1)}(U)] \\ &= \int_p^1 F^{-1}(u)\, du, \end{aligned}$$

where $U \sim \mathrm{Uniform}([0,1])$ and we use the fact that $F^{-1}(U) \sim Y$.[4] Finally, note that $P(Y \ge Q_p(Y)) = 1 - p$ for continuous distributions.

The expected shortfall for the left tail is

$$\mathrm{ES}_p(Y) = \frac{1}{p} \int_0^p Q_u(Y)\, du, \quad 0 < p < 1.$$

When Y has a continuous distribution function, then the expected shortfall for the left tail is

$$\mathrm{ES}_p(Y) = E(Y \mid Y \le Q_p(Y)) = \frac{1}{p}\, E(Y I_{(-\infty,Q_p(Y)]}(Y)). \tag{3.26}$$

This expression shows that in the case of a continuous distribution function, $p\mathrm{ES}_p(Y)$ is equal to the expectation that is taken only over the left tail, when the left tail is defined as the region that is on the left side of the pth quantile of

4 We have that $P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$.

the distribution. Note that the expected shortfall for the left tail is related to the lower conditional moment of order $k = 1$ and target rate $\tau = Q_p(Y)$:

$$\begin{aligned} ES_p(Y) &= Q_p(Y) - E(Q_p(Y) - Y \mid Y \le Q_p(Y)) \\ &= Q_p(Y) - \mathrm{LCM}_{Q_p(Y),1}, \end{aligned}$$

where the lower conditional moment $\mathrm{LCM}_{Q_p(Y),1}$ is defined in (3.19).[5]

The expected shortfall for the right tail, as defined in (3.24), can be estimated from the data $Y_1, \dots, Y_T$ by

$$\widehat{ES}_p(Y) = \frac{1}{T - m + 1} \sum_{i=m}^{T} Y_{(i)}, \tag{3.27}$$

where $Y_{(1)} \le \cdots \le Y_{(T)}$ and $m = \lceil pT \rceil$, with, for example, $p = 0.95$ or $p = 0.99$. When the expected shortfall is for the left tail, as defined by (3.26), then we define the estimator as

$$\widehat{ES}_p(Y) = \frac{1}{m} \sum_{i=1}^{m} Y_{(i)}, \tag{3.28}$$

where $m = \lceil pT \rceil$ with, for example, $p = 0.05$ or $p = 0.01$.

3.2 Univariate Graphical Tools

We consider sequence $Y_1, \dots, Y_T \in \mathbf{R}$ of real numbers, and assume that the sequence is a sample from a probability distribution. We want to visualize the sequence in order to discover properties of the underlying distribution. We divide the graphical tools to those that are based on the empirical distribution function and the empirical quantiles, and to those that are based on

5 Analogously to the definition of lower partial moments in (3.15), we can define the absolute shortfall as

$$AS_p(Y) = E(YI_{(-\infty, Q_p(Y)]}(Y)).$$

The absolute shortfall for the left tail is related to the lower partial moment of order $k = 1$ and target rate $\tau = Q_p(Y)$:

$$\begin{aligned} AS_p(Y) &= pQ_p(Y) - E((Q_p(Y) - Y)I_{(-\infty, Q_p(Y)]}(Y)) \\ &= pQ_p(Y) - \mathrm{LPM}_{Q_p(Y),1}. \end{aligned}$$

The absolute shortfall is estimated from observations $Y_1, \dots, Y_T$ by

$$\widehat{AS}_p(Y) = \frac{1}{T} \sum_{i=1}^{m} Y_{(i)},$$

where $Y_{(1)} \le \cdots \le Y_{(T)}$ is the ordered sample and $m = \lceil pT \rceil$. Here, we divide by T, but in the estimator (3.28) of the expected shortfall we divide by m.

the estimation of the underlying density function. The distribution function and quantiles based tools give more insight about the tails of the distribution, and the density based tools give more information about the center of the distribution.

A two-variate data can be visualized using a scatter plot. For a univariate data there is no such obvious method available. Thus, visualizing two-variate data may seem easier than visualizing univariate data. However, we can consider many of the tools to visualize univariate data to be scatter plots of points

$$(Y_i, \mathrm{level}(Y_i)), \quad i = 1, \dots, T, \tag{3.29}$$

where level : $\{Y_1, \dots, Y_T\} \to \mathbf{R}$ is a mapping that attaches a real value to each data point $Y_i \in \mathbf{R}$. Thus, in a sense we visualize univariate data by transforming it into a two-dimensional data.

3.2.1 Empirical Distribution Function Based Tools

The distribution function of the distribution of random variable $Y \in \mathbf{R}$ is

$$F(x) = P(Y \le x), \quad x \in \mathbf{R}.$$

The empirical distribution function can be considered as a starting point for several visualizations: tail plots, regression plots of tails, and empirical quantile functions. We use often tail plots. Regression plots of tails have two types: (1) plots that look linear for an exponential tail and (2) plots that look linear for a Pareto tail.

3.2.1.1 The Empirical Distribution Function

The empirical distribution function $\hat{F}$, based on data $Y_1, \dots, Y_T$, is defined as

$$\hat{F}(x) = \frac{1}{T} \#\{Y_i : Y_i \le x, \quad i = 1, \dots, T\}, \tag{3.30}$$

where $x \in \mathbf{R}$, and $\#A$ means the cardinality of set A. Note that the empirical distribution function is defined in (8.20) using the indicator function. An empirical distribution function is a piecewise constant function. Plotting a graph of an empirical distribution function is for large samples practically the same as plotting the points

$$(Y_{(i)}, i/T), \quad i = 1, \dots, T, \tag{3.31}$$

where $Y_{(1)} \le \cdots \le Y_{(T)}$ are the ordered observations. Thus, the empirical distribution function fits the scheme of transforming univariate data to two-dimensional data as in (3.29).

Figure 3.1 shows empirical distribution functions of S&P 500 net returns (red) and 10-year bond net returns (blue). The monthly data of S&P 500 and US Treasury 10-year bond returns is described in Section 2.4.3. Panel (a) plots the points (3.31) and panel (b) zooms to the lower left corner, showing the

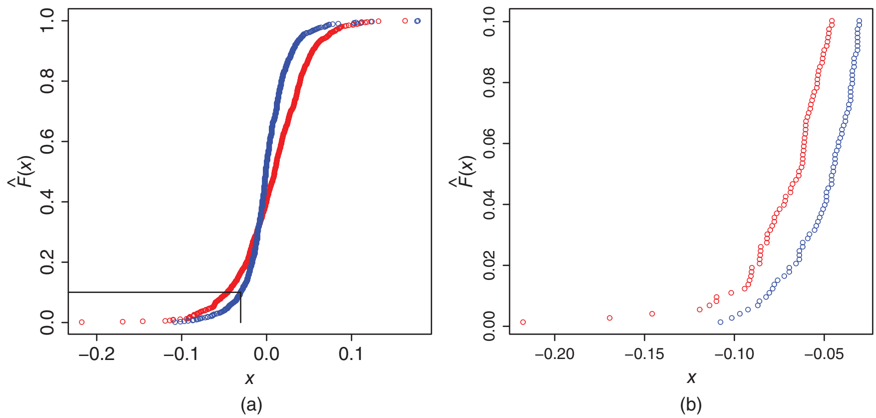


Figure 3.1 *Empirical distribution functions*. (a) Empirical distribution functions of S&P 500 returns (red) and 10-year bond returns (blue); (b) zooming at the lower left corner.

empirical distribution function for the $[T/10]$ smallest observations; the empirical distribution function is shown on the range $x \in (-\infty, \hat{q}_p)$, where $\hat{q}_p$ is the pth empirical quantile for $p = 0.1$. Neither of the estimated return distributions dominates the other: The S&P 500 distribution function is higher at the left tail but lower at the right tail. That is, S&P 500 is more risky than the 10-year bond. Note that Section 9.2.3 discusses stochastic dominance: a first return distribution dominates stochastically a second return distribution when the first distribution function takes smaller values everywhere than the second distribution function.

3.2.1.2 The Tail Plots

The left and right tail plots can be used to visualize the heaviness of the tails of the underlying distribution. A smooth tail plot can be used to visualize simultaneously a large number of samples. The tail plots are almost the same as the empirical distribution function, but there are couple of differences:

1) In tail plots we divide the data into the left tail and the right tail, and we visualize separately the two tails.
2) In tail plots the y-axis shows the number of observations and a logarithmic scale is used for the y-axis.

Tail plots have been applied in Mandelbrot (1963), Bouchaud and Potters (2003), and Sornette (2003).

The Left and the Right Tail Plots The observations in the left tail are

$$\mathcal{L} = \{Y_i : Y_i < u,\ i = 1, \ldots, T\},$$

where $u = \hat{q}_p$ is the pth empirical quantile for $0 < p < 1/2$. For the left tail plot we choose the level

$$\mathrm{level}(Y_i) = \#\{Y_j : Y_j \le Y_i, Y_j \in \mathcal{L}\}, \quad Y_i \in \mathcal{L}. \tag{3.32}$$

Thus, the smallest observation has level one, the second smallest observation has level two, and so on. Note that $\mathrm{level}(Y_i)$ is often called the rank of Y_i. The left tail plot is the two-dimensional scatter plot of the points $(Y_i, \mathrm{level}(Y_i))$, $Y_i \in \mathcal{L}$, when the logarithmic scale is used for the y-axis.

The observations in the right tail are

$$\mathcal{R} = \{Y_i : Y_i > u,\ i = 1, \dots, T\},$$

where $u = \hat{q}_p$ is the pth empirical quantile for $1/2 < p < 1$. We choose the level of Y_i as the number of observations larger or equal to Y_i:

$$\mathrm{level}(Y_i) = \#\{Y_j : Y_j \ge Y_i, Y_j \in \mathcal{R}\}, \quad Y_i \in \mathcal{R}. \tag{3.33}$$

Thus, the largest observation has level one, the second largest observation has level two, and so on. The right tail plot is the two-dimensional scatter plot of the points $(Y_i, \mathrm{level}(Y_i))$, $Y_i \in \mathcal{R}$, when the logarithmic scale is used for the y-axis.

The left tail plot can be considered as an estimator of the function

$$L(x) = TF(x), \tag{3.34}$$

where F is the underlying distribution function and $x \in (-\infty, u]$. Indeed, for the level in (3.32) we have that $\mathrm{level}(Y_i) = T\hat{F}(Y_i)$. The right tail plot can be considered as an estimator of the function

$$R(x) = T(1 - F(x)), \tag{3.35}$$

where $x \in [u, \infty)$. For the level in (3.33) we have that $\mathrm{level}(Y_i) \approx T(1 - \hat{F}(Y_i))$.

Figure 3.2 shows the left and right tail plots for the daily S&P 500 data, described in Section 2.4.1. Panel (a) shows the left tail plot and panel (b) shows the right tail plot. The black circles show the data points. The y-axis is logarithmic. The colored curves show the population versions (3.34) and (3.35) for the Gaussian distribution (red) and for the Student distributions with degrees of freedom $\nu = 3, 4, 5, 6$ (blue).[6] We can see that for the left tail Student's distribution with degrees of freedom $\nu = 3$ gives the best fit, but for the right tail degrees of freedom $\nu = 4$ gives the best fit.

6 The Gaussian curve in the left tail plot shows the function $x \mapsto T\Phi((x - \hat{\mu})/\hat{\sigma})$, where Φ is the distribution function of the standard Gaussian distribution, $\hat{\mu}$ is the sample mean, and $\hat{\sigma}$ is the sample standard deviation. In the right tail plot the function is $x \mapsto T(1 - \Phi((x - \hat{\mu})/\hat{\sigma}))$. The Student curves in the left tail plot are $x \mapsto TF_\nu((x - \hat{\mu})/\hat{\sigma})$, where F_ν is the distribution function of the Student distribution with degrees of freedom ν, and $\hat{\sigma} = \hat{s}/\sqrt{\nu/(\nu - 2)}$, where $\hat{s}$ is the sample standard deviation. The Student distributions are defined in (3.53). Note that when $Y \sim N(\mu, \sigma^2)$, then the distribution function of Y is $\Phi((x - \mu)/\sigma)$. When $Y \sim t(\nu, \mu, \sigma^2)$, then the distribution function of Y is $F_\nu((x - \mu)/\sigma)$.

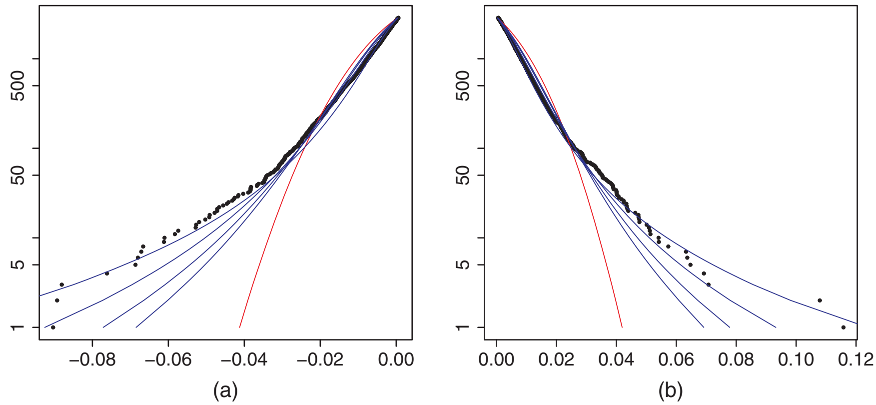


Figure 3.2 *Left and right tail plots*. (a) The left tail plot for S&P 500 returns; (b) the right tail plot. The red curve shows the theoretical Gaussian curve and the blue curves show the Student curves for the degrees of freedom $\nu = 3$–6.

A left tail plot and a right tail plot can be combined into one figure, at least when both the left and the right tails are defined by taking the threshold to be the sample median $u = \hat{q}_{0.5}$ (see Figures 14.24(a) and 14.25(a)).

Smooth Tail Plots Figure 3.3 shows smooth tail plots for the S&P 500 components data, described in Section 2.4.5. Panel (a) shows left tail plots and panel (b) shows right tail plots. The gray scale image visualizes with one picture all tail plots of the stocks in the S&P 500 components data. The red points show the tail plots of S&P 500 index, which is also shown in Figure 3.2. Note that the x-axes have the ranges $[-0.1, 0]$ and $[0, 0.1]$, so that the extreme observations are not shown. Note that instead of the logarithmic scale of y-values $1, \dots, [T/2]$, we have used values $\log(1), \log(2), \dots, \log([T/2])$ on the y-axis. We can see that the index has lighter tails than most of the individual stocks.

In a smooth tail plot we make an image that simultaneously shows several tail plots. Let us have m stocks and T returns for each stock. We draw a separate left or right tail plot for each stock. Plotting these tail plots in the same figure would cause overlapping, and we would see only a black image. That is why we use smoothing. We divide the x-axis to 300 grid points, say. The y-axis has $[T/2]$ grid points. Thus, we have $300 \times [T/2]$ pixels. For each x-value we compute the value of a univariate kernel density estimator at that x-value. Each kernel estimator is constructed using m observations. This is done for each $[T/2]$ rows, so that we evaluate $[T/2]$ estimates at 300 points. See Section 3.2.2 about kernel density estimation. We choose the smoothing parameter using the normal reference rule and use the standard Gaussian kernel. The values of the density estimate are raised to the power of 21 before applying the gray scale.

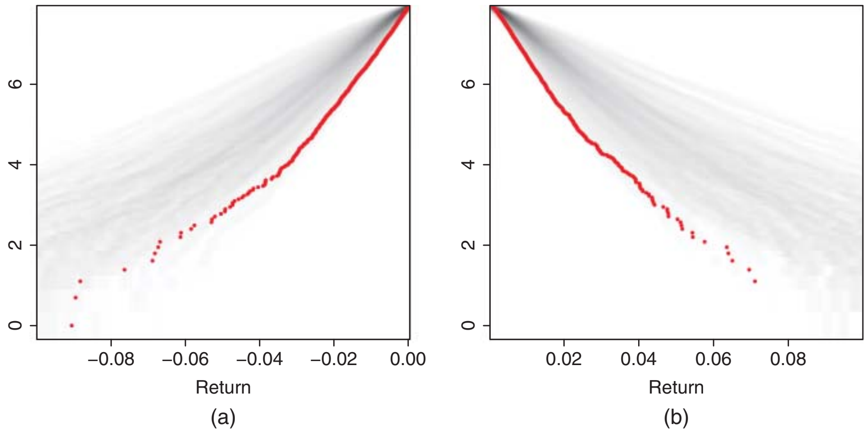


Figure 3.3 *Smooth tail plots*. The gray scale images show smooth tail plots of a collection of stocks in the S&P 500 index. The red points show the tail plots of the S&P 500 index. (a) A smooth left tail plot; (b) a smooth right tail plot.

3.2.1.3 Regression Plots of Tails

Regression plots are related to the empirical distribution function, just like tail plots, but now the data is transformed so that it lies on $[0, \infty)$, both in the case of the left tail and in the case of the right tail. We use the term "regression plot" because these plots suggest fitting linear regression curves to the data. We distinguish the plot for which exponential tails looks linear and the plot for which Pareto tails look linear.

Plots which Look Linear for an Exponential Tail Let the original observations be $Y_1, \dots, Y_T$. Let $u \in \mathbf{R}$ be a threshold. We choose u to be an empirical quantile $\hat{q}_p$ for some $p \in (0, 1)$: $\hat{q}_p = Y_{(m)}$ for $m = [pT]$, where $Y_{(1)} < \cdots < Y_{(T)}$ are the ordered observations. Let $\mathcal{T}_l$ be the left tail and $\mathcal{T}_r$ be the right tail, transformed so that the observations lie on $[0, \infty)$:

$$\mathcal{T}_l = \{u - Y_i \, : \, Y_i \le u\}, \quad \mathcal{T}_r = \{Y_i - u \, : \, Y_i \ge u\}.$$

For the left tail $u = \hat{q}_p$ for $p \in (0, 1/2)$ and for the right tail $u = \hat{q}_p$ for $p \in (1/2, 1)$. Let us denote by $\mathcal{T}$ either the left tail or the right tail. Denote

$$n = \#\mathcal{T}.$$

Let

$$\hat{F}(z) = \frac{1}{n+1} \#\{Z_i \in \mathcal{T} \, : \, Z_i \le z\}$$

be the empirical distribution function, based on data $\mathcal{T}$. Note that in the usual definition of the empirical distribution function we divide by n, but now we divide by $n + 1$ because we need that $\hat{F}(Z_i) < 1$, in order to take the logarithm of $1 - \hat{F}(Z_i)$. Denote

$$\mathcal{T} = \{Z_1, \dots, Z_n\}.$$

Assume that the data is ordered:

$$Z_1 < \cdots < Z_n.$$

We have that

$$\hat{F}(Z_i) = \frac{i}{n+1}.$$

The regression plot that is linear for exponential tails is a scatter plot of the points[7]

$$\{(Z_i, \log(1 - \hat{F}(Z_i))) : Z_i \in \mathcal{T}\}. \tag{3.36}$$

Figure 3.4 shows scatter plots of points in (3.36). We use the S&P 500 daily data, described in Section 2.4.1. Panel (a) plots data in the left tail with $p = 10\%$ (black), $p = 5\%$ (red), and $p = 1\%$ (blue). Panel (b) plots data in the right tail with $p = 90\%$ (black), $p = 95\%$ (red), and $p = 99\%$ (blue).

The data looks linear for exponential tails and convex for Pareto tails. The exponential distribution function is $F(x) = 1 - \exp\{-x/\beta\}$ for $x \geq 0$, where $\beta > 0$. The exponential distribution function satisfies

$$\log(1 - F(x)) = -\frac{x}{\beta}\, I_{(0,\infty)}(x).$$

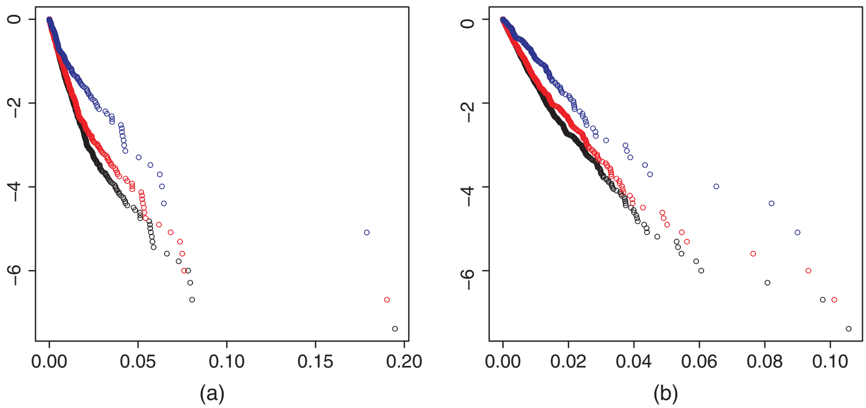


Figure 3.4 *Regression plots which are linear for exponential tails: S&P 500 daily returns.* (a) Left tail with $p = 10\%$ (black), $p = 5\%$ (red), and $p = 1\%$ (blue); (b) right tail with $p = 90\%$ (black), $p = 95\%$ (red), and $p = 99\%$ (blue).

7 Denote $p_i = \hat{F}(Z_i)$ and $q_{p_i} = Z_i$. Then we can write (3.36) as a plot of points $(q_{p_i}, \log(1 - p_i))$. The plot of points $(-\log(1 - p_i), q_{p_i})$ is called the return level plot; see Coles (2004, pp. 49, 81).

Plotting the curve

$$x \mapsto -\frac{x}{\beta} \tag{3.37}$$

for $x \geq 0$ and for various values of $\beta > 0$ shows how well the exponential distributions fit the tail. The Pareto distribution function for the support $[0, \infty)$ is $F(x) = 1 - (u/(x+u))^{\alpha}$ for $x \geq 0$, where $\alpha > 0$; see (3.74). The Pareto distribution function satisfies

$$\log(1 - F(x)) = -\alpha \log\left(\frac{x+u}{u}\right) I_{(0,\infty)}(x).$$

Plotting the curve

$$x \mapsto -\alpha \log\left(\frac{x+u}{u}\right) \tag{3.38}$$

for $x \geq 0$ and for various values of $\alpha > 0$ shows how well the Pareto distributions fit the tail.[8]

Figure 3.5 shows how parametric models are fitted to the left tail, defined by the pth empirical quantile with $p = 5\%$. We use the S&P 500 daily data, as described in Section 2.4.1. Panel (a) shows fitting of exponential tails: we show functions (3.37) for three values of parameter β. Panel (a) shows fitting of Pareto tails: we show functions (3.38) for three values of parameter α.

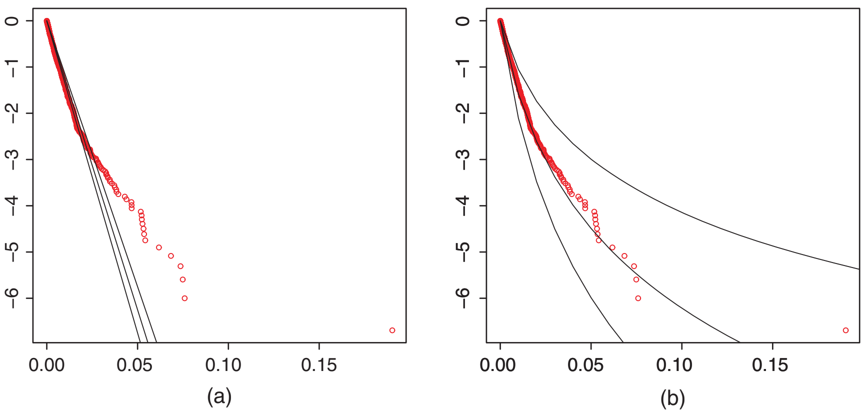


Figure 3.5 *Fitting of parametric families for data that is linear for exponential tails.* The data points are from left tail of S&P 500 daily returns, defined by the pth empirical quantile with $p = 0.05$. (a) Fitting of exponential distributions; (b) fitting of Pareto distributions.

8 The generalized Pareto distribution is defined in (3.83). The distribution function F satisfies

$$\log(1 - F(x)) = \begin{cases} -\frac{1}{\xi} \log\left(1 + \frac{\xi x}{\beta}\right) I_{[0,\infty)}(x), & \xi > 0, \\ -\frac{x}{\beta} I_{[0,\infty)}(x), & \xi = 0, \end{cases}$$

where $\beta > 0$.

The middle values of the parameters are the maximum likelihood estimates, defined in Section 3.4.2.

Plots which Look Linear for a Pareto Tail Let

$$\mathcal{T}_l = \{Y_i/u : Y_i \leq u\}, \quad \mathcal{T}_r = \{Y_i/u : Y_i \geq u\}.$$

For the right tail we assume that $u > 0$ and for the left tail we assume that $u < 0$. Let us denote by $\mathcal{T}$ either the left tail or the right tail. Denote

$$\mathcal{T} = \{Z_1, \dots, Z_n\}.$$

Assume that the data is ordered: $Z_1 < \cdots < Z_n$. The regression plot that is linear for Pareto tails is a scatter plots of the points

$$\left\{(\log Z_i, \log(1 - \hat{F}(Z_i))) : Z_i \in \mathcal{T}\right\}. \tag{3.39}$$

Figure 3.6 shows scatter plots of points in (3.39). We use the S&P 500 daily data, described in Section 2.4.1. Panel (a) plots data in the left tail with $p = 10\%$ (black), $p = 5\%$ (red), and $p = 1\%$ (blue). Panel (b) plots data in the right tail with $p = 90\%$ (black), $p = 95\%$ (red), and $p = 99\%$ (blue).

The data looks linear for Pareto tails and concave for exponential tails. The exponential distribution function for the support $[u, \infty)$ is $F(x) = 1 - \exp\{-(x-u)/\beta\}$ for $x \geq u$, where $\beta > 0$. The exponential distribution function satisfies

$$\log(1 - F(x)) = -\frac{x-u}{\beta}\, I_{(u,\infty)}(x).$$

Plotting the curve

$$x \mapsto -\frac{|u|}{\beta}\,(e^x - 1)$$

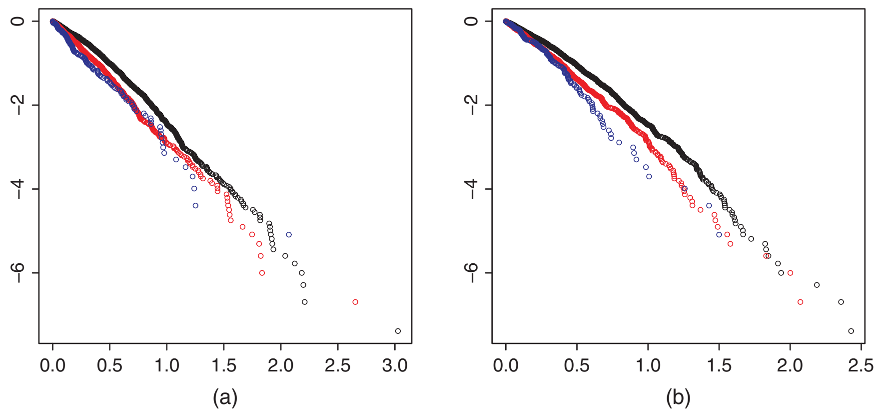


Figure 3.6 *Regression plots which are linear for Pareto tails: S&P 500 daily returns.* (a) Left tail with $p = 10\%$ (black), $p = 5\%$ (red), and $p = 1\%$ (blue); (b) right tail with $p = 90\%$ (black), $p = 95\%$ (red), and $p = 99\%$ (blue).

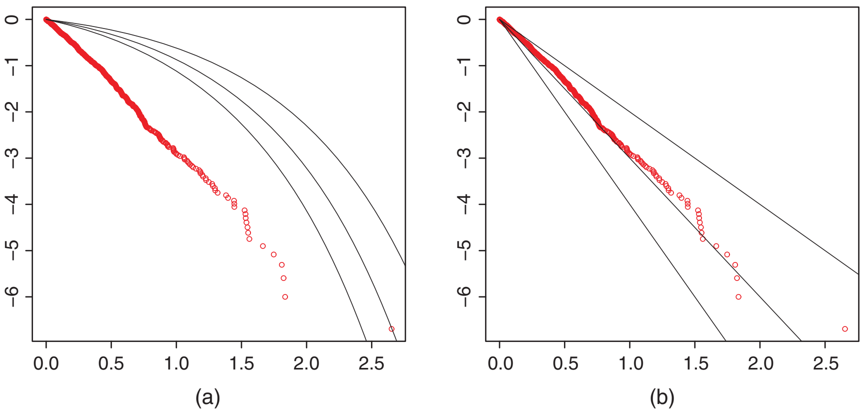


Figure 3.7 *Fitting of parametric families for data that is linear for Pareto tails*. The data points are from left tail of S&P 500 daily returns, defined by the pth empirical quantile with $p = 0.05$. (a) Fitting of exponential distributions; (b) fitting of Pareto distributions.

for $x \geq 0$ and for various values of $\beta > 0$ shows how well the exponential distributions fit the tail. The Pareto distribution function for the support $[u, \infty)$ is $F(x) = 1 - (u/x)^{\alpha}$ for $x \geq u$, where $\alpha > 0$. The Pareto distribution function satisfies

$$\log(1 - F(x)) = -\alpha \log\left(\frac{x}{u}\right) I_{(u,\infty)}(x).$$

Plotting the curve

$$x \mapsto -\alpha x$$

for $x \geq 0$ and for various values of $\alpha > 0$ shows how well the Pareto distributions fit the tail.

Figure 3.7 shows how parametric models are fitted to the left tail, defined by the pth empirical quantile with $p = 5\%$. We use the S&P 500 daily data, described in Section 2.4.1. Panel (a) shows fitting of exponential tails: we show functions (3.37) for three values of parameter β. Panel (a) shows fitting of Pareto tails: we show functions (3.38) for three values of parameter α. The middle values of the parameters are the maximum likelihood estimates, defined in Section 3.4.2.

3.2.1.4 The Empirical Quantile Function

The pth quantile of the distribution of the random variable $Y \in \mathbf{R}$ is defined in (3.21) as

$$Q_p = \inf\{y : F(y) \geq p\},$$

where $0 < p < 1$ and $F(y) = P(Y \leq y)$ is the distribution function of Y. The empirical quantile can be defined as

$$\hat{Q}_p = \inf\{y : \hat{F}(y) \geq p\},$$

where $\hat{F}$ is the empirical distribution function, as defined in (3.30); see (8.21). Section 8.4.1 contains equivalent definitions of the empirical quantile.

The quantile function is

$$p \mapsto Q_p, \quad p \in (0, 1).$$

For continuous distributions the quantile function is the same as the inverse of the distribution function. The empirical quantile function is

$$p \mapsto \hat{Q}_p, \quad p \in (0, 1), \tag{3.40}$$

where $\hat{Q}_p$ is the empirical quantile. A quantile function can be used to compare return distributions. A first return distribution dominates a second return distribution when the first quantile function takes higher values everywhere than the second quantile function. See Section 9.2.3 about stochastic dominance.

Plotting a graph of the empirical quantile function is close to plotting the points

$$\left(i/T, Y_{(i)}\right), \quad i = 1, \ldots, T, \tag{3.41}$$

where $Y_{(1)} < \cdots < Y_{(T)}$ are the ordered observations.

Figure 3.8 shows empirical quantile functions of S&P 500 returns (red) and 10-year bond returns (blue). The monthly data of S&P 500 and US Treasury 10-year bond returns is described in Section 2.4.3. Panel (a) plots the points (3.41) and panel (b) zooms at the lower left corner, showing the empirical quantile on the range $p \in (0, 0.1)$. Neither of the estimated return distributions dominates the other: The S&P 500 returns have a higher median and higher upper quantiles, but they have smaller lower quantiles. That is, S&P 500 is more risky than 10-year bond.

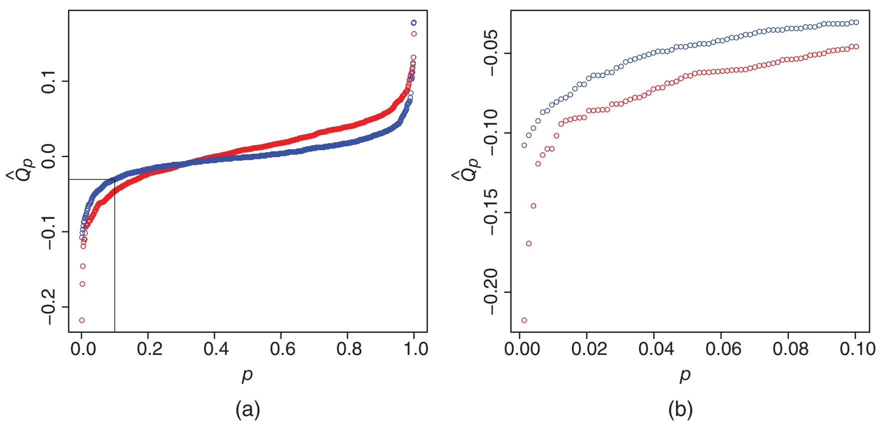


Figure 3.8 *Empirical quantile functions.* (a) Empirical quantile functions of S&P 500 returns (red) and 10-year bond returns (blue); (b) zooming to the lower left corner.

3.2.2 Density Estimation Based Tools

We describe both histograms and kernel density estimators.

3.2.2.1 The Histogram

A histogram estimator of the density of $X \in \mathbf{R}^d$, based on identically distributed observations $X_1, \dots, X_T$, is defined as

$$\hat{f}(x) = \sum_{i=1}^{M} \frac{n_i/T}{\text{volume}(R_i)} I_{R_i}(x), \quad x \in \mathbf{R}^d, \tag{3.42}$$

where $\{R_1, \dots, R_M\}$ is a partition on $\mathbf{R}^d$ and

$$n_i = \#\{i : X_i \in R, \quad i = 1, \dots, T\}$$

is the number of observations in R_i. The partition is a collection of sets $R_1, \dots, R_M$ that are (almost surely) disjoint and they cover the space of the observed values $X_1, \dots, X_T$.[9]

Figure 3.9(a) shows a histogram estimate using S&P 500 returns. We use the S&P 500 monthly data, described in Section 2.4.3. The histogram is

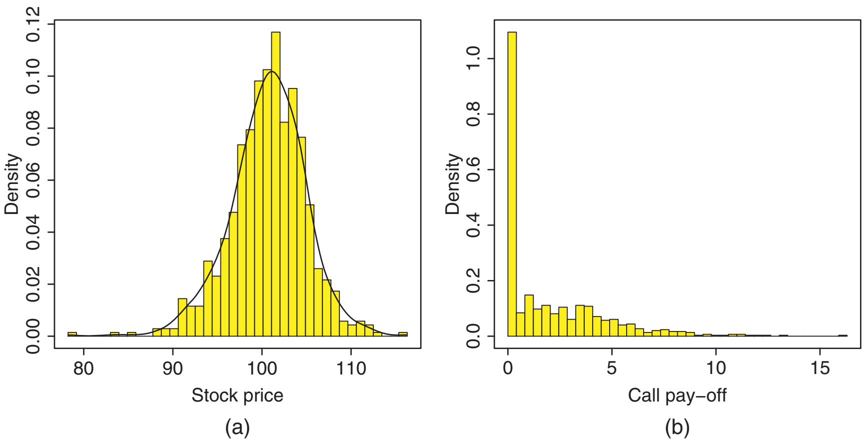


Figure 3.9 *Histogram estimates.* (a) A histogram of historically simulated S&P 500 prices. A graph of kernel density estimate is included. (b) A histogram of historically simulated call option pay-offs.

---

9 In the univariate case the partition to the intervals of equal length can be defined by

$$R_i = [a_i, b_i], \quad a_i = X_{(1)} + \delta(i-1), \quad b_i = a_i + \delta, \quad \delta = (X_{(T)} - X_{(1)})/M,$$

where $X_{(1)} = \min\{X_i\}$ and $X_{(T)} = \max\{X_i\}$. Then the histogram can be written as

$$\hat{f}(x) = \frac{1}{T\delta} \sum_{i=1}^{M} n_i \, I_{R_i}(x), \quad x \in \mathbf{R}.$$

constructed from the data $100 \times R_t$, $t = 1, \dots, T$, where R_t are the monthly gross returns. Panel (b) shows a histogram constructed from the historically simulated pay-offs of the call option with the strike price 100. The histogram is constructed from the data $\max\{100R_t - 100, 0\}$, $t = 1, \dots, T$. Panel (a) includes a graph of a kernel density estimate, defined in (3.43). The histogram in panel (b) illustrates that a histogram is convenient to visualize the density of data that is not from a continuous distribution; for this data the value 0 has a probability about 0.5.

3.2.2.2 The Kernel Density Estimator

The kernel density estimator $\hat{f}(x)$ of the density function $f : \mathbf{R}^d \to \mathbf{R}$ of random vector $X \in \mathbf{R}^d$, based on identically distributed data $X_1, \dots, X_T \in \mathbf{R}^d$, is defined by

$$\hat{f}(x) = \frac{1}{T} \sum_{i=1}^{T} K_h(x - X_i), \quad x \in \mathbf{R}^d, \tag{3.43}$$

where $K : \mathbf{R}^d \to \mathbf{R}$ is the kernel function, $K_h(x) = K(x/h)/h^d$, and $h > 0$ is the smoothing parameter.[10]

We can also take the vector smoothing parameter $h = (h_1, \dots, h_d)$ and $K_h(x) = K(x_1/h_1, \dots, x_d/h_d)/\prod_{i=1}^d h_i$. The smoothing parameter of the kernel density estimator can be chosen using the normal reference rule:

$$h_i = \left(\frac{4}{d+2}\right)^{1/(d+4)} T^{-1/(d+4)} \hat{\sigma}_i, \tag{3.44}$$

for $i = 1, \dots, d$, where $\hat{\sigma}_i$ is the sample standard deviation for the ith variable; see Silverman (1986, p. 45). Alternatively, the sample variances of the marginal distributions can be normalized to one, so that $\hat{\sigma}_1 = \cdots = \hat{\sigma}_d = 1$.

Figure 3.10(a) shows kernel estimates of the distribution of S&P 500 monthly net returns (blue) and of the distribution of US 10-year bond monthly net returns (red). The data set of monthly returns of S&P 500 and US 10-year bond is described in Section 2.4.3. Panel (b) shows kernel density estimates of S&P

10 The definition of the kernel density estimator can be motivated in the following way. The density a point $x \in \mathbf{R}^d$ can be approximated by

$$f(x) \approx \frac{P(B_h(x))}{\lambda(B_h(x))},$$

where $B_h(x) = \{y \in \mathbf{R}^d : \|x - y\| \le h\}$, $h > 0$ is small, and $\lambda(B_h(x))$ is the Lebesgue measure of $B_h(x)$. We have that

$$\frac{P(B_h(x))}{\lambda(B_h(x))} \approx \frac{1}{\lambda(B_h(x))} \frac{1}{T} \sum_{i=1}^{T} I_{B_h(x)}(X_i) = \frac{1}{T} \sum_{i=1}^{T} K_h(x - X_i),$$

when $K(x) = I_{B_1(0)}(x)$. We arrive into (3.43) by allowing other kernel functions than only the indicator function.