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.