MULTIVARIATE NONPARAMETRIC REGRESSION AND VISUALIZATION
To my parents
## CONTENTS IN BRIEF

### PART I METHODS OF REGRESSION AND CLASSIFICATION

1. Overview of Regression and Classification  
   3

2. Linear Methods and Extensions  
   77

3. Kernel Methods and Extensions  
   127

4. Semiparametric and Structural Models  
   229

5. Empirical Risk Minimization  
   241

### PART II VISUALIZATION

6. Visualization of Data  
   277

7. Visualization of Functions  
   295
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>xvii</td>
</tr>
<tr>
<td>Introduction</td>
<td>xix</td>
</tr>
<tr>
<td>I.1 Estimation of Functionals of Conditional Distributions</td>
<td>xx</td>
</tr>
<tr>
<td>I.2 Quantitative Finance</td>
<td>xxi</td>
</tr>
<tr>
<td>I.3 Visualization</td>
<td>xxi</td>
</tr>
<tr>
<td>I.4 Literature</td>
<td>xxi</td>
</tr>
</tbody>
</table>

## PART I METHODS OF REGRESSION AND CLASSIFICATION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Overview of Regression and Classification</td>
<td>3</td>
</tr>
<tr>
<td>1.1 Regression</td>
<td>3</td>
</tr>
<tr>
<td>1.1.1 Random Design and Fixed Design</td>
<td>4</td>
</tr>
<tr>
<td>1.1.2 Mean Regression</td>
<td>5</td>
</tr>
<tr>
<td>1.1.3 Partial Effects and Derivative Estimation</td>
<td>8</td>
</tr>
<tr>
<td>1.1.4 Variance Regression</td>
<td>9</td>
</tr>
<tr>
<td>1.1.5 Covariance and Correlation Regression</td>
<td>13</td>
</tr>
<tr>
<td>1.1.6 Quantile Regression</td>
<td>14</td>
</tr>
<tr>
<td>1.1.7 Approximation of the Response Variable</td>
<td>18</td>
</tr>
<tr>
<td>1.1.8 Conditional Distribution and Density</td>
<td>21</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>1.1.9</td>
<td>Time Series Data</td>
</tr>
<tr>
<td>1.1.10</td>
<td>Stochastic Control</td>
</tr>
<tr>
<td>1.1.11</td>
<td>Instrumental Variables</td>
</tr>
<tr>
<td>1.2</td>
<td><strong>Discrete Response Variable</strong></td>
</tr>
<tr>
<td>1.2.1</td>
<td>Binary Response Models</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Discrete Choice Models</td>
</tr>
<tr>
<td>1.2.3</td>
<td>Count Data</td>
</tr>
<tr>
<td>1.3</td>
<td><strong>Parametric Family Regression</strong></td>
</tr>
<tr>
<td>1.3.1</td>
<td>General Parametric Family</td>
</tr>
<tr>
<td>1.3.2</td>
<td>Exponential Family Regression</td>
</tr>
<tr>
<td>1.3.3</td>
<td>Copula Modeling</td>
</tr>
<tr>
<td>1.4</td>
<td><strong>Classification</strong></td>
</tr>
<tr>
<td>1.4.1</td>
<td>Bayes Risk</td>
</tr>
<tr>
<td>1.4.2</td>
<td>Methods of Classification</td>
</tr>
<tr>
<td>1.5</td>
<td><strong>Applications in Quantitative Finance</strong></td>
</tr>
<tr>
<td>1.5.1</td>
<td>Risk Management</td>
</tr>
<tr>
<td>1.5.2</td>
<td>Variance Trading</td>
</tr>
<tr>
<td>1.5.3</td>
<td>Portfolio Selection</td>
</tr>
<tr>
<td>1.5.4</td>
<td>Option Pricing and Hedging</td>
</tr>
<tr>
<td>1.6</td>
<td><strong>Data Examples</strong></td>
</tr>
<tr>
<td>1.6.1</td>
<td>Time Series of S&amp;P 500 Returns</td>
</tr>
<tr>
<td>1.6.2</td>
<td>Vector Time Series of S&amp;P 500 and Nasdaq-100 Returns</td>
</tr>
<tr>
<td>1.7</td>
<td><strong>Data Transformations</strong></td>
</tr>
<tr>
<td>1.7.1</td>
<td>Data Sphering</td>
</tr>
<tr>
<td>1.7.2</td>
<td>Copula Transformation</td>
</tr>
<tr>
<td>1.7.3</td>
<td>Transformations of the Response Variable</td>
</tr>
<tr>
<td>1.8</td>
<td><strong>Central Limit Theorems</strong></td>
</tr>
<tr>
<td>1.8.1</td>
<td>Independent Observations</td>
</tr>
<tr>
<td>1.8.2</td>
<td>Dependent Observations</td>
</tr>
<tr>
<td>1.8.3</td>
<td>Estimation of the Asymptotic Variance</td>
</tr>
<tr>
<td>1.9</td>
<td><strong>Measuring the Performance of Estimators</strong></td>
</tr>
<tr>
<td>1.9.1</td>
<td>Performance of Regression Function Estimators</td>
</tr>
<tr>
<td>1.9.2</td>
<td>Performance of Conditional Variance Estimators</td>
</tr>
<tr>
<td>1.9.3</td>
<td>Performance of Conditional Covariance Estimators</td>
</tr>
<tr>
<td>1.9.4</td>
<td>Performance of Quantile Function Estimators</td>
</tr>
<tr>
<td>1.9.5</td>
<td>Performance of Estimators of Expected Shortfall</td>
</tr>
<tr>
<td>1.9.6</td>
<td>Performance of Classifiers</td>
</tr>
<tr>
<td>1.10</td>
<td><strong>Confidence Sets</strong></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1.10.1 Pointwise Confidence Intervals 73
1.10.2 Confidence Bands 75

1.11 Testing 75

2 Linear Methods and Extensions 77

2.1 Linear Regression 78
  2.1.1 Least Squares Estimator 79
  2.1.2 Generalized Method of Moments Estimator 81
  2.1.3 Ridge Regression 84
  2.1.4 Asymptotic Distributions for Linear Regression 87
  2.1.5 Tests and Confidence Intervals for Linear Regression 90
  2.1.6 Variable Selection 92
  2.1.7 Applications of Linear Regression 94

2.2 Varying Coefficient Linear Regression 97
  2.2.1 The Weighted Least Squares Estimator 97
  2.2.2 Applications of Varying Coefficient Regression 98

2.3 Generalized Linear and Related Models 102
  2.3.1 Generalized Linear Models 102
  2.3.2 Binary Response Models 104
  2.3.3 Growth Models 107

2.4 Series Estimators 107
  2.4.1 Least Squares Series Estimator 107
  2.4.2 Orthonormal Basis Estimator 108
  2.4.3 Splines 110

2.5 Conditional Variance and ARCH Models 111
  2.5.1 Least Squares Estimator 112
  2.5.2 ARCH Model 113

2.6 Applications in Volatility and Quantile Estimation 116
  2.6.1 Benchmarks for Quantile Estimation 116
  2.6.2 Volatility and Quantiles with the LS Regression 118
  2.6.3 Volatility with the Ridge Regression 121
  2.6.4 Volatility and Quantiles with ARCH 122

2.7 Linear Classifiers 124

3 Kernel Methods and Extensions 127

3.1 Regressogram 129
3.2 Kernel Estimator 130
  3.2.1 Definition of the Kernel Regression Estimator 130
3.2.2 Comparison to the Regressogram 132
3.2.3 Gasser–Müller and Priestley–Chao Estimators 134
3.2.4 Moving Averages 134
3.2.5 Locally Stationary Data 136
3.2.6 Curse of Dimensionality 140
3.2.7 Smoothing Parameter Selection 140
3.2.8 Effective Sample Size 142
3.2.9 Kernel Estimator of Partial Derivatives 145
3.2.10 Confidence Intervals in Kernel Regression 146

3.3 Nearest-Neighbor Estimator 147

3.4 Classification with Local Averaging 148
3.4.1 Kernel Classification 148
3.4.2 Nearest-Neighbor Classification 149

3.5 Median Smoothing 151

3.6 Conditional Density Estimation 152
3.6.1 Kernel Estimator of Conditional Density 152
3.6.2 Histogram Estimator of Conditional Density 156
3.6.3 Nearest-Neighbor Estimator of Conditional Density 157

3.7 Conditional Distribution Function Estimation 158
3.7.1 Local Averaging Estimator 159
3.7.2 Time–Space Smoothing 159

3.8 Conditional Quantile Estimation 160

3.9 Conditional Variance Estimation 162
3.9.1 State–Space Smoothing and Variance Estimation 162
3.9.2 GARCH and Variance Estimation 163
3.9.3 Moving Averages and Variance Estimation 172

3.10 Conditional Covariance Estimation 176
3.10.1 State–Space Smoothing and Covariance Estimation 178
3.10.2 GARCH and Covariance Estimation 178
3.10.3 Moving Averages and Covariance Estimation 181

3.11 Applications in Risk Management 181
3.11.1 Volatility Estimation 182
3.11.2 Covariance and Correlation Estimation 193
3.11.3 Quantile Estimation 198

3.12 Applications in Portfolio Selection 205
3.12.1 Portfolio Selection Using Regression Functions 205
3.12.2 Portfolio Selection Using Classification 215
3.12.3 Portfolio Selection Using Markowitz Criterion 223
PART II VISUALIZATION

6 Visualization of Data 277

6.1 Scatter Plots 278
6.1.1 Two-Dimensional Scatter Plots 278
6.1.2 One-Dimensional Scatter Plots 278
6.1.3 Three- and Higher-Dimensional Scatter Plots 282

6.2 Histogram and Kernel Density Estimator 283
## 6.3 Dimension Reduction
- 6.3.1 Projection Pursuit 284
- 6.3.2 Multidimensional Scaling 286

## 6.4 Observations as Objects
- 6.4.1 Graphical Matrices 289
- 6.4.2 Parallel Coordinate Plots 290
- 6.4.3 Other Methods 293

## 7 Visualization of Functions 295
- 7.1 Slices 296
- 7.2 Partial Dependence Functions 298
- 7.3 Reconstruction of Sets
  - 7.3.1 Estimation of Level Sets of a Function 300
  - 7.3.2 Point Cloud Data 303
- 7.4 Level Set Trees 304
  - 7.4.1 Definition and Illustrations 304
  - 7.4.2 Calculation of Level Set Trees 308
  - 7.4.3 Volume Function 313
  - 7.4.4 Barycenter Plot 321
  - 7.4.5 Level Set Trees in Regression Function Estimation 322
- 7.5 Unimodal Densities
  - 7.5.1 Probability Content of Level Sets 327
  - 7.5.2 Set Visualization 327

## Appendix A: R Tutorial 329
- A.1 Data Visualization
  - A.1.1 QQ Plots 329
  - A.1.2 Tail Plots 330
  - A.1.3 Two-Dimensional Scatter Plots 330
  - A.1.4 Three-Dimensional Scatter Plots 331
- A.2 Linear Regression 331
- A.3 Kernel Regression
  - A.3.1 One-Dimensional Kernel Regression 332
  - A.3.2 Moving Averages 333
  - A.3.3 Two-Dimensional Kernel Regression 334
  - A.3.4 Three- and Higher-Dimensional Kernel Regression 336
  - A.3.5 Kernel Estimator of Derivatives 338
- A.4 Local Linear Regression 341
A.4.1 One-Dimensional Local Linear Regression 341
A.4.2 Two-Dimensional Local Linear Regression 342
A.4.3 Three- and Higher-Dimensional Local Linear Regression 343
A.4.4 Local Linear Derivative Estimation 343

A.5 Additive Models: Backfitting 344

A.6 Single-Index Regression 345
A.6.1 Estimating the Index 346
A.6.2 Estimating the Link Function 346
A.6.3 Plotting the Single-Index Regression Function 346

A.7 Forward Stagewise Modeling 347
A.7.1 Stagewise Fitting of Additive Models 347
A.7.2 Projection Pursuit Regression 348

A.8 Quantile Regression 349
A.8.1 Linear Quantile Regression 349
A.8.2 Kernel Quantile Regression 349

References 351

Author Index 361

Topic Index 365
The book is intended for students and researchers who want to learn to apply non-parametric and semiparametric methods and to use visualization tools related to these estimation methods. In particular, the book is intended for students and researchers in quantitative finance who want to apply statistical methods and for students and researchers of statistics who want to learn to apply statistical methods in quantitative finance. The book continues the themes of Klemelä (2009), which studied density estimation. The current book focuses on regression function estimation.

The book was written at the University of Oulu, Department of Mathematical Sciences. I wish to acknowledge the support provided by the University of Oulu and the Department of Mathematical Sciences.

The web page of the book is http://cc.oulu.fi/~jklmela/regstruct/.

JUSSI KLEMELÄ

Oulu, Finland
October 2013
We study regression analysis and classification, as well as estimation of conditional variances, quantiles, densities, and distribution functions. The focus of the book is on nonparametric methods. Nonparametric methods are flexible and able to adapt to various kinds of data, but they can suffer from the curse of dimensionality and from the lack of interpretability. Semiparametric methods are often able to cope with quite high-dimensional data and they are often easier to interpret, but they are less flexible and their use may lead to modeling errors. In addition to terms “nonparametric estimator” and “semiparametric estimator”, we can use the term “structured estimator” to denote such estimators that arise, for example, in additive models. These estimators obey a structural restriction, whereas the term “semiparametric estimator” is used for estimators that have a parametric and a nonparametric component.

Nonparametric, semiparametric, and structured methods are well established and widely applied. There are, nevertheless, areas where further work is useful. We have included three such areas in this book:

1. Estimation of several functionals of a conditional distribution; not only estimation of the conditional expectation but also estimation of the conditional variance and conditional quantiles.

2. Quantitative finance as an area of application for nonparametric and semiparametric methods.
3. Visualization tools in statistical learning.

I.1 ESTIMATION OF FUNCTIONALS OF CONDITIONAL DISTRIBUTIONS

One of the main topics of the book are the kernel methods. Kernel methods are easy to implement and computationally feasible, and their definition is intuitive. For example, a kernel regression estimator is a local average of the values of the response variable. Local averaging is a general regression method. In addition to the kernel estimator, examples of local averaging include the nearest-neighbor estimator, the regressogram, and the orthogonal series estimator.

We cover linear regression and generalized linear models. These models can be seen as starting points to many semiparametric and structured regression models. For example, the single index model, the additive model, and the varying coefficient linear regression model can be seen as generalizations of the linear regression model or the generalized linear model.

Empirical risk minimization is a general approach to statistical estimation. The methods of empirical risk minimization can be used in regression function estimation, in classification, in quantile regression, and in the estimation of other functionals of the conditional distribution. The method of local empirical risk minimization is a method which can be seen as a generalization of the kernel regression.

A regular regressogram is a special case of local averaging, but the empirical choice of the partition leads to a rich class of estimators. The choice of the partition is made using empirical risk minimization. In the one- and two-dimensional cases a regressogram is usually less efficient than the kernel estimator, but in high-dimensional cases a regressogram can be useful. For example, a method to select the partition of a regressogram can be seen as a method of variable selection, if the chosen partition is such that it can be defined using only a subset of the variables. The estimators that are defined as a solution of an optimization problem, like the minimizers of an empirical risk, need typically be calculated with numerical methods. Stagewise algorithms can also be taken as a definition of an estimator, even without giving an explicit minimization problem which they solve.

A regression function is defined as the conditional expectation of the distribution of a response variable. The conditional expectation is useful in making predictions as well as in finding causal relationships. We cover also the estimation of the conditional variance and conditional quantiles. These are needed to give a more complete view of the conditional distribution. Also, the estimation of the conditional variance and conditional quantiles is needed in risk management, which is an important area of quantitative finance. The conditional variance can be estimated by estimating the conditional expectation of the squared random variable, whereas a conditional quantile is a special case of the conditional median. In the time series setting the standard approaches for estimating the conditional variance are the ARCH and GARCH modeling, but we discuss nonparametric alternatives. The GARCH estimator is close
to a moving average, whereas the ARCH estimator is related to linear state space modeling.

In classification we are not interested in the estimation of functionals of a distribution, but the aim is to construct classification rules. However, most of the regression function estimation methods have a counterpart in classification.

I.2 QUANTITATIVE FINANCE

Risk management, portfolio selection, and option pricing can be identified as three important areas of quantitative finance. Parametric statistical methods have been dominating the statistical research in quantitative finance. In risk management, probability distributions have been modeled with the Pareto distribution or with distributions derived from the extreme value theory. In portfolio selection the multivariate normal model has been used together with the Markowitz theory of portfolio selection. In option pricing the Black-Scholes model of stock prices has been widely applied. The Black-Scholes model has also been extended to more general parametric models for the process of stock prices.

In risk management the $p$-quantile of a loss distribution has a direct interpretation as such threshold that the probability of the loss exceeding the threshold is less than $p$. Thus estimation of conditional quantiles is directly relevant for risk management. Unconditional quantile estimators do not take into account all available information, and thus in risk management it is useful to estimate conditional quantiles. The estimation of the conditional variance can be applied in the estimation of a conditional quantile, because in location-scale families the variance determines the quantiles. The estimation of conditional variance can be extended to the estimation of the conditional covariance or the conditional correlation.

We apply nonparametric regression function estimation in portfolio selection. The portfolio is selected either with the maximization of a conditional expected utility or with the maximization of a Markowitz criterion. When the collection of allowed portfolio weights is a finite set, then also classification can be used in portfolio selection. The squared returns are much easier to predict than the returns themselves, and thus in quantitative finance the focus has been in the prediction of volatility. However, it can be shown that despite the weak predictability of the returns, portfolio selection can profit from statistical prediction.

Option pricing can be formulated as a problem of stochastic control. We do not study the statistics of option pricing in detail, but give a basic framework for solving some option pricing problems nonparametrically.

I.3 VISUALIZATION

Statistical visualization is often considered as a visualization of the raw data. The visualization of the raw data can be a part of the exploratory data analysis, a first step to model building, and a tool to generate hypotheses about the data-generating mechanism. However, we put emphasis on a different approach to visualization.
In this approach, visualization tools are associated with statistical estimators or inference procedures. For example, we estimate first a regression function and then try to visualize and describe the properties of this regression function estimate. The distinction between the visualization of the raw data and the visualization of the estimator is not clear when nonparametric function estimation is used. In fact, nonparametric function estimation can be seen as a part of exploratory data analysis.

The SiZer is an example of a tool that combines visualization and inference, see Chaudhuri & Marron (1999). This methodology combines formal testing for the existence of modes with the SiZer maps to find out whether a mode of a density estimate of a regression function estimate is really there.

Semiparametric function estimates are often easier to visualize than nonparametric function estimates. For example, in a single index model the regression function estimate is a composition of a linear function and a univariate function. Thus in a single index model we need only to visualize the coefficients of the linear function and a one-dimensional function. The ease of visualization gives motivation to study semiparametric methods.

CART, as presented in Breiman, Friedman, Olshen & Stone (1984), is an example of an estimation method whose popularity is not only due to its statistical properties but also because it is defined in terms of a binary tree that gives directly a visualization of the estimator. Even when it is possible to find estimators with better statistical properties than CART, the possibility to visualization gives motivation to use CART.

Visualization of nonparametric function estimates, such as kernel estimates, is challenging. For the visualization of completely nonparametric estimates, we can use level set tree-based methods, as presented in Klemelä (2009). Level set tree-based methods have found interest also in topological data analysis and in scientific visualization, and these methods have their origin in the concept of a Reeb graph, defined originally in Reeb (1946).

In density estimation we are often interested in the mode structure of the density, defined as the number of local extremes, the largeness of the local extremes, and the location of the local extremes. The local extremes of a density function are related to the areas of concentration of the probability mass. In regression function estimation we are also interested in the mode structure. The local maxima of a regression function are related to the regions of the space of the explanatory variables where the response variable takes the largest values. The antimode structure is equally important to describe. The antimode structure means the number of local minima, the size of the local minima, and the location of the local minima. The local minima of a regression function are related to the areas of the space of the explanatory variables where the response variable takes the smallest values.

The mode structure of a regression function does not give complete information about the properties of the regression function. In regression analysis we are interested in the effects of the explanatory variables on the response variable and in the interaction between the explanatory variables. The effect of an explanatory variable can be formalized with the concept of a partial effect. The partial effect of an explanatory variable is the partial derivative of the regression function with respect to this variable. Nearly constant partial effects indicate that the regression function is
close to a linear function, since the partial derivatives of a linear function are constants. The local maxima of a partial effect correspond to the areas in the space of the explanatory variables where the increase of the expected value of the response variable, resulting from an increase of the value of the explanatory variable, is the largest. We can use level set trees of partial effects to visualize the mode structure and the antimode structure of the partial effects, and thus to visualize the effects and the interactions of the explanatory variables.

I.4 LITERATURE

