

*Härdle • Klinke • Müller*  
XploRe – Learning Guide

**Springer**

*Berlin*

*Heidelberg*

*New York*

*Barcelona*

*Hong Kong*

*London*

*Milan*

*Paris*

*Singapore*

*Tokyo*

W. Härdle • S. Klinke • M. Müller

# XploRe – Learning Guide



Springer

*Wolfgang Härdle*

*Sigbert Klinke*

*Marlene Müller*

Humboldt-Universität zu Berlin

Institut für Statistik und Ökonometrie

Spandauer Straße 1

10178 Berlin, Germany

*e-mail:* haerdle/sigbert/marlene@wiwi.hu-berlin.de

---

Cover art: Detail of “Portrait of the Franciscan monk Luca Pacioli, humanist and mathematician (1445-1517)” by Jacopo de Barbari, Museo di Capodimonte, Naples, Italy

---

CIP data applied for

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

**Härdle, Wolfgang:**

XploRe - learning guide / W. Härdle; S. Klinke; M. Müller.- Berlin; Heidelberg; New York; Barcelona; Hong Kong; London; Milan; Paris; Singapore; Tokyo: Springer, 2000

ISBN-13: 978-3-540-66207-5

e-ISBN-13: 978-3-642-60232-0

DOI: 10.1007/978-3-642-60232-0

Additional material to this book can be downloaded from <http://extras.springer.com>.

ISBN 978-3-540-66207-5 Springer-Verlag Berlin Heidelberg New York

---

Mathematics Subject Classification (1991): 62-04, 62-00

---

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 2000

The use of general descriptive names, registered names, trademarks etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready copy by authors

SPIN 10733011 40/3143CK-5 4 3 2 1 0 – Printed on acid-free paper

# Contents

<b>Preface</b>	<b>15</b>
<b>Part I: First Steps</b>	<b>19</b>
<b>1 Getting Started</b>	<b>21</b>
1.1 Using XploRe . . . . .	21
1.1.1 Input and Output Windows . . . . .	21
1.1.2 Simple Computations . . . . .	22
1.1.3 First Data Analysis . . . . .	23
1.1.4 Exploring Data . . . . .	24
1.1.5 Printing Graphics . . . . .	27
1.2 Quantlet Examples . . . . .	29
1.2.1 Summary Statistics . . . . .	30
1.2.2 Histograms . . . . .	30
1.2.3 2D Density Estimation . . . . .	31
1.2.4 Interactive Kernel Regression . . . . .	33
1.3 Getting Help . . . . .	34
1.4 Basic XploRe Syntax . . . . .	36
1.4.1 Operators . . . . .	36
1.4.2 Variables . . . . .	38

---

1.4.3	Variable Names . . . . .	38
1.4.4	Functions . . . . .	39
1.4.5	Quantlet files . . . . .	40
<b>2</b>	<b>Descriptive Statistics</b>	<b>43</b>
	<i>Marlene Müller</i>	
2.1	Data Matrices . . . . .	43
2.1.1	Creating Data Matrices . . . . .	44
2.1.2	Loading Data Files . . . . .	46
2.1.3	Matrix Operations . . . . .	47
2.2	Computing Statistical Characteristics . . . . .	49
2.2.1	Minimum and Maximum . . . . .	50
2.2.2	Mean, Variance and Other Moments . . . . .	51
2.2.3	Median and Quantiles . . . . .	52
2.2.4	Covariance and Correlation . . . . .	55
2.2.5	Categorical Data . . . . .	57
2.2.6	Missing Values and Infinite Values . . . . .	59
2.3	Summarizing Statistical Information . . . . .	63
2.3.1	Summarizing Metric Data . . . . .	63
2.3.2	Summarizing Categorical Data . . . . .	66
<b>3</b>	<b>Graphics</b>	<b>69</b>
	<i>Sigbert Klinke</i>	
3.1	Basic Plotting . . . . .	70
3.1.1	Plotting a Data Set . . . . .	70
3.1.2	Plotting a Function . . . . .	71
3.1.3	Plotting Several Functions . . . . .	72

---

3.1.4	Coloring Data Sets . . . . .	73
3.1.5	Plotting Lines from Data Sets . . . . .	74
3.1.6	Several Plots . . . . .	76
3.2	Univariate Graphics . . . . .	79
3.2.1	Boxplots . . . . .	80
3.2.2	Dotplots . . . . .	82
3.2.3	Bar Charts . . . . .	83
3.2.4	Quantile-Quantile Plots . . . . .	85
3.2.5	Histograms . . . . .	86
3.3	Multivariate Graphics . . . . .	90
3.3.1	Three-Dimensional Plots . . . . .	91
3.3.2	Surface Plots . . . . .	92
3.3.3	Contour Plots . . . . .	92
3.3.4	Sunflower Plots . . . . .	94
3.3.5	Linear Regression . . . . .	95
3.3.6	Bivariate Plots . . . . .	97
3.3.7	Star Diagrams . . . . .	99
3.3.8	Scatter-Plot Matrices . . . . .	100
3.3.9	Andrews Curves . . . . .	102
3.3.10	Parallel Coordinate Plots . . . . .	103
3.4	Advanced Graphics . . . . .	105
3.4.1	Moving and Rotating . . . . .	105
3.4.2	Simple Predefined Graphic Primitives . . . . .	106
3.4.3	Color Models . . . . .	108
3.5	Graphic Commands . . . . .	109
3.5.1	Controlling Data Points . . . . .	110
3.5.2	Color of Data Points . . . . .	111

---

3.5.3	Symbol of Data Points . . . . .	113
3.5.4	Size of Data Points . . . . .	115
3.5.5	Connection of Data Points . . . . .	116
3.5.6	Label of Data Points . . . . .	121
3.5.7	Title and Axes Labels . . . . .	124
3.5.8	Axes Layout . . . . .	125
<b>4</b>	<b>Regression Methods</b>	<b>129</b>
	<i>Jörg Aßmus</i>	
4.1	Simple Linear Regression . . . . .	131
4.2	Multiple Linear Regression . . . . .	137
4.3	Nonlinear Regression . . . . .	143
<b>5</b>	<b>Teachware Quantlets</b>	<b>147</b>
	<i>Nathaniel Derby</i>	
5.1	Visualizing Data . . . . .	149
5.2	Random Sampling . . . . .	150
5.3	The $p$ -Value in Hypothesis Testing . . . . .	153
5.4	Approximating the Binomial by the Normal Distribution . . . . .	155
5.5	The Central Limit Theorem . . . . .	157
5.6	The Pearson Correlation Coefficient . . . . .	159
5.7	Linear Regression . . . . .	162
	Bibliography . . . . .	165



**Part II: Statistical Libraries 167**

**6 Smoothing Methods 169**

*Marlene Müller*

- 6.1 Kernel Density Estimation . . . . . 169
  - 6.1.1 Computational Aspects . . . . . 171
  - 6.1.2 Computing Kernel Density Estimates . . . . . 173
  - 6.1.3 Kernel Choice . . . . . 176
  - 6.1.4 Bandwidth Selection . . . . . 177
  - 6.1.5 Confidence Intervals and Bands . . . . . 181
- 6.2 Kernel Regression . . . . . 185
  - 6.2.1 Computational Aspects . . . . . 185
  - 6.2.2 Computing Kernel Regression Estimates . . . . . 186
  - 6.2.3 Bandwidth Selection . . . . . 188
  - 6.2.4 Confidence Intervals and Bands . . . . . 192
  - 6.2.5 Local Polynomial Regression and Derivative Estimation 194
- 6.3 Multivariate Density and Regression Functions . . . . . 197
  - 6.3.1 Computational Aspects . . . . . 197
  - 6.3.2 Multivariate Density Estimation . . . . . 197
  - 6.3.3 Multivariate Regression . . . . . 201
- Bibliography . . . . . 203

**7 Generalized Linear Models 205**

*Marlene Müller*

- 7.1 Estimating GLMs . . . . . 206
  - 7.1.1 Models . . . . . 206
  - 7.1.2 Maximum-Likelihood Estimation . . . . . 208

---

7.2	Computing GLM Estimates . . . . .	208
7.2.1	Data Preparation . . . . .	208
7.2.2	Interactive Estimation . . . . .	209
7.2.3	Noninteractive Estimation . . . . .	213
7.3	Weights & Constraints . . . . .	215
7.3.1	Prior Weights . . . . .	216
7.3.2	Replications in Data . . . . .	217
7.3.3	Constrained Estimation . . . . .	217
7.4	Options . . . . .	218
7.4.1	Setting Options . . . . .	219
7.4.2	Weights and Offsets . . . . .	219
7.4.3	Control Parameters . . . . .	219
7.4.4	Output Modification . . . . .	221
7.5	Statistical Evaluation and Presentation . . . . .	221
7.5.1	Statistical Characteristics . . . . .	221
7.5.2	Output Display . . . . .	223
7.5.3	Significance of Parameters . . . . .	223
7.5.4	Likelihood Ratio Tests for Comparing Nested Models . . . . .	225
7.5.5	Subset Selection . . . . .	226
	Bibliography . . . . .	228
<b>8</b>	<b>Neural Networks</b>	<b>229</b>
	<i>Wolfgang Härdle and Heiko Lehmann</i>	
8.1	Feed-Forward Networks . . . . .	231
8.2	Computing a Neural Network . . . . .	232
8.2.1	Controlling the Parameters of the Neural Network . . . . .	234
8.2.2	The Resulting Neural Network . . . . .	235

8.3 Running a Neural Network . . . . . 236

    8.3.1 Implementing a Simple Discriminant Analysis . . . . . 237

    8.3.2 Implementing a More Complex Discriminant Analysis . . . . . 241

Bibliography . . . . . 246

**9 Time Series 247**

*Petr Franěk and Wolfgang Härdle*

9.1 Time Domain and Frequency Domain Analysis . . . . . 247

    9.1.1 Autocovariance and Autocorrelation Function . . . . . 248

    9.1.2 The Periodogram and the Spectrum of a Series . . . . . 251

9.2 Linear Models . . . . . 253

    9.2.1 Autoregressive Models . . . . . 253

    9.2.2 Autoregressive Moving Average Models . . . . . 254

    9.2.3 Estimating ARMA Processes . . . . . 256

9.3 Nonlinear Models . . . . . 259

    9.3.1 Several Examples of Nonlinear Models . . . . . 259

    9.3.2 Nonlinearity in the Conditional Second Moments . . . . . 264

    9.3.3 Estimating ARCH Models . . . . . 266

    9.3.4 Testing for ARCH . . . . . 267

Bibliography . . . . . 270

**10 Kalman Filtering 273**

*Petr Franěk*

10.1 State–Space Models . . . . . 273

    10.1.1 Examples of State–Space Models . . . . . 275

    10.1.2 Modeling State–Space Models in XploRe . . . . . 276

10.2 Kalman Filtering and Smoothing . . . . . 277

10.3 Parameter Estimation in State–Space Models . . . . .	280
Bibliography . . . . .	283
<b>11 Finance</b>	<b>285</b>
<i>Stefan Sperlich and Wolfgang Härdle</i>	
11.1 Outline of the Theory . . . . .	286
11.1.1 Some History . . . . .	286
11.1.2 The Black–Scholes Formula . . . . .	287
11.2 Assets . . . . .	289
11.2.1 Stock Simulation . . . . .	290
11.2.2 Stock Estimation . . . . .	292
11.2.3 Stock Estimation and Simulation . . . . .	292
11.3 Options . . . . .	294
11.3.1 Calculation of Option Prices and Implied Volatilities . .	294
11.3.2 Option Price Determining Factors . . . . .	298
11.3.3 Greeks . . . . .	301
11.4 Portfolios and Hedging . . . . .	304
11.4.1 Calculation of Arbitrage . . . . .	304
11.4.2 Bull-Call Spreads . . . . .	305
<b>12 Microeconometrics and Panel Data</b>	<b>307</b>
<i>Jörg Breitung and Axel Werwatz</i>	
12.1 Limited-Dependent and Qualitative Dependent Variables . . .	308
12.1.1 Probit, Logit and Tobit . . . . .	308
12.1.2 Single Index Models . . . . .	310
12.1.3 Average Derivatives . . . . .	311
12.1.4 Average Derivative Estimation . . . . .	312

12.1.5	Weighted Average Derivative Estimation . . . . .	314
12.1.6	Average Derivatives and Discrete Variables . . . . .	315
12.1.7	Parametric versus Semiparametric Single Index Models	318
12.2	Multiple Index Models . . . . .	320
12.2.1	Sliced Inverse Regression . . . . .	321
12.2.2	Testing Parametric Multiple Index Models . . . . .	322
12.3	Self-Selection Models . . . . .	324
12.3.1	Parametric Model . . . . .	325
12.3.2	Semiparametric Model . . . . .	327
12.4	Panel Data Analysis . . . . .	330
12.4.1	The Data Set . . . . .	333
12.4.2	Time Effects . . . . .	335
12.4.3	Model Specification . . . . .	336
12.4.4	Estimation . . . . .	338
12.4.5	An Example . . . . .	339
12.5	Dynamic Panel Data Models . . . . .	343
12.6	Unit Root Tests for Panel Data . . . . .	347
	Bibliography . . . . .	349

**13 Extreme Value Analysis** **353**

*Rolf-Dieter Reiss and Michael Thomas*

13.1	Extreme Value Models . . . . .	354
13.2	Generalized Pareto Distributions . . . . .	356
13.3	Assessing the Adequacy: Mean Excess Functions . . . . .	358
13.4	Estimation in EV Models . . . . .	359
13.4.1	Linear Combination of Ratios of Spacings (LRS) . . . . .	359
13.4.2	ML Estimator in the EV Model . . . . .	360

13.4.3	ML Estimator in the Gumbel Model . . . . .	360
13.5	Fitting GP Distributions to the Upper Tail . . . . .	361
13.6	Parametric Estimators for GP Models . . . . .	362
13.6.1	Moment Estimator . . . . .	363
13.6.2	ML Estimator in the GP Model . . . . .	364
13.6.3	Pickands Estimator . . . . .	364
13.6.4	Drees–Pickands Estimator . . . . .	365
13.6.5	Hill Estimator . . . . .	366
13.6.6	ML Estimator for Exponential Distributions . . . . .	366
13.6.7	Selecting a Threshold by Means of a Diagram . . . . .	367
13.7	Graphical User Interface . . . . .	368
13.8	Example . . . . .	369
	Bibliography . . . . .	373
<b>14</b>	<b>Wavelets</b>	<b>375</b>
	<i>Yuri Golubev, Wolfgang Härdle, Zdeněk Hlávka, Sigbert Klinke, Michael H. Neumann and Stefan Sperlich</i>	
14.1	Quantlib twave . . . . .	377
14.1.1	Change Basis . . . . .	378
14.1.2	Change Function . . . . .	379
14.1.3	Change View . . . . .	380
14.2	Discrete Wavelet Transform . . . . .	381
14.3	Function Approximation . . . . .	383
14.4	Data Compression . . . . .	385
14.5	Two Sines . . . . .	388
14.6	Frequency Shift . . . . .	389
14.7	Thresholding . . . . .	392

14.7.1 Hard Thresholding . . . . .	393
14.7.2 Soft Thresholding . . . . .	395
14.7.3 Adaptive Thresholding . . . . .	397
14.8 Translation Invariance . . . . .	402
14.9 Image Denoising . . . . .	404
Bibliography . . . . .	407

**Part III: Programming 409**

**15 Reading and Writing Data 411**

*Sigbert Klinke, Jürgen Symanzik and Marlene Müller*

15.1 Reading and Writing Data Files . . . . .	411
15.2 Input Format Strings . . . . .	414
15.3 Output Format Strings . . . . .	417
15.4 Customizing the Output Window . . . . .	419
15.4.1 Headline Style . . . . .	421
15.4.2 Layer Style . . . . .	422
15.4.3 Line Number Style . . . . .	424
15.4.4 Value Formats and Lengths . . . . .	425
15.4.5 Saving Output to a File . . . . .	426

**16 Matrix Handling 429**

*Yasemin Boztug and Marlene Müller*

16.1 Basic Operations . . . . .	429
16.1.1 Creating Matrices and Arrays . . . . .	430
16.1.2 Operators for Numeric Matrices . . . . .	435
16.2 Comparison Operators . . . . .	440
16.3 Matrix Manipulation . . . . .	442

---

16.3.1	Extraction of Elements . . . . .	442
16.3.2	Matrix Transformation . . . . .	445
16.4	Sums and Products . . . . .	448
16.5	Distance Function . . . . .	450
16.6	Decompositions . . . . .	451
16.6.1	Spectral Decomposition . . . . .	451
16.6.2	Singular Value Decomposition . . . . .	454
16.6.3	LU Decomposition . . . . .	455
16.6.4	Cholesky Decomposition . . . . .	456
16.7	Lists . . . . .	457
16.7.1	Creating Lists . . . . .	457
16.7.2	Handling Lists . . . . .	459
16.7.3	Getting Information on Lists . . . . .	462
<b>17</b>	<b>Quantlets and Quantlibs</b>	<b>465</b>
<i>Wolfgang Härdle, Zdeněk Hlávka and Sigbert Klinke</i>		
17.1	Quantlets . . . . .	465
17.2	Flow Control . . . . .	476
17.2.1	Local and Global Variables . . . . .	476
17.2.2	Conditioning . . . . .	478
17.2.3	Branching . . . . .	480
17.2.4	While-Loop . . . . .	481
17.2.5	Do-Loop . . . . .	482
17.2.6	Optional Input and Output in Procedures . . . . .	483
17.2.7	Errors and Warnings . . . . .	486
17.3	User Interaction . . . . .	488
17.4	APSS . . . . .	495



17.5 Quantlibs . . . . . 499

**Appendix 503**

**A Customizing XploRe 505**

A.1 XploRe.ini . . . . . 505

A.1.1 The ini File . . . . . 505

A.1.2 Composing Paths . . . . . 507

A.2 startup.xpl . . . . . 508

**B Data Sets 509**

B.1 Netincome–Food Expenditures . . . . . 509

B.2 U.S. Companies . . . . . 509

B.3 CPS 1985 . . . . . 510

B.4 Boston Housing . . . . . 510

B.5 Lizard Data . . . . . 511

B.6 Kyphosis Data . . . . . 512

B.7 Swiss Bank Notes . . . . . 513

B.8 Earnings Data . . . . . 513

B.9 Westwood Data . . . . . 514

B.10 Pullover Data . . . . . 514

B.11 Geysers Data . . . . . 514

Bibliography . . . . . 515

**Index 516**

# Preface

It is generally accepted that training in statistics must include some exposure to the mechanics of computational statistics. This learning guide is intended for beginners in computer-aided statistical data analysis. The prerequisites for XploRe — the statistical computing environment — are an introductory course in statistics or mathematics. The reader of this book should be familiar with basic elements of matrix algebra and the use of HTML browsers. This guide is designed to help students to XploRe their data, to learn (via data interaction) about statistical methods and to disseminate their findings via the HTML outlet. The XploRe APSS (Auto Pilot Support System) is a powerful tool for finding the appropriate statistical technique (quantlet) for the data under analysis. Homogeneous quantlets are combined in XploRe into quantlibs.

The XploRe language is intuitive and users with prior experience of other statistical programs will find it easy to reproduce the examples explained in this guide. The quantlets in this guide are available on the CD-ROM as well as on the Internet. The statistical operations that the student is guided into range from basic one-dimensional data analysis to more complicated tasks such as time series analysis, multivariate graphics construction, microeconometrics, panel data analysis, etc.

The guide starts with a simple data analysis of pullover sales data, then introduces graphics. The graphics are interactive and cover a wide range of displays of statistical data. The regression chapter guides the user first to linear scatter-plot regression and then to multiple regression and nonlinear models. The teachware quantlets comprise a basic set of interactive, illustrative examples in introductory statistics. For the student, they provide an opportunity to understand some important basic concepts in statistics through trial and error. For the teacher, they can aid the instruction by allowing the students to independently work on exploratory examples at their own pace. Additionally, with a modicum of understanding of the XploRe programming language, the teacher can modify these examples to fit his/her own preferences.

In the second part, nonparametric smoothing methods are introduced. They provide a flexible analysis tool, often based on interactive graphical data representation. The focus there is on function estimation by kernel smoothing. More refined tools are the neural networks. They consist of many simple processing units (neurons) that are connected by communication channels. This guide shows how to adapt such networks to real data. The time series quantlets cover the standard analysis techniques and offer nonlinear methods for modern time series data analysis. The Kalman filtering module is a set of recursive quantlets based on the construction of an estimate from the previous periods and observations available at the time. The `finance` quantlib considers the statistical analysis of financial markets. This quantlib offers functions to predict, to simulate and to estimate finance data series such as, for example, stock returns, to determine option prices and to evaluate different scenarios (e.g. for portfolio strategies).

The microeconometrics section introduces the tools available in XploRe for analyzing micro data, i.e. data sets consisting of observations on individual units, such as persons, households or firms. The quantlib `metrics` provides the statistical tools such as panel analysis for analyzing observed individual behavior. The analysis of risk is one of the primary objectives of insurance firms and the profit and loss analysis on financial markets. It is also applied to flood discharges and high concentration of air pollutants. The ‘value at risk’ and the extreme (upper or lower) parts of a sample are estimated and analyzed in this module. Wavelets comprise a powerful tool which can be used for a wide range of applications, namely describing a signal for parsimonious representation, for data compression, for smoothing and image denoising and for jump detection. This guide introduces the wavelet quantlib `twave`.

XploRe offers a large variety of commands and tools for creating and manipulating multidimensional objects called matrices and lists. The last part of this guide presents the basic instructions for matrix handling and illustrates further topics in matrix algebra and list handling with XploRe. The quantlet construction technique and the generation of a user defined APSS are the topics of the final chapter on quantlets and quantlibs.

XploRe and this learning guide have benefited at several stages from cooperation with many colleagues and students. We want to mention in particular: Gökhan Aydinli, Silke Baars, Alexander Benkwitz, Marco Bianchi, Rong Chen, Vila Co, Stefan Daske, Sven Denkert, Ulrich Dorazelski, Michaela Dranganska, Dorit Feldmann, Jörg Feuerhake, Frank Geppert, Birgit Grund, Janet Grassmann, Christian Hafner, Susanne Hannappel, Nicholas Hengartner, Holger

---

Jakob, Christopher Kath, Jussi Klemelä, Petra Korndörfer, Gunnar Krause, Katharina Kröl, Stephan Lauer, Pascal Lavergne, Michel Lejeune, Torsten Kleinow, Thomas Kötter, Thomas Kühn, Hua Liang, Steve Marron, Danilo Mercurio, Hans-Joachim Mucha, Fabian Nötzel, Isabel Proenca, Andreas Rupp, Jerome Saracco, Svetlana Schmelzer, Levon Shahbaghyan, Hizir Sofyan, Hans Gerhard Strohe, Wolfgang Stockmeyer, Elke Tscharf, Berwin Turlach, Rodrigo Witzel, Lijian Yang, Kerstin Zanter.

We owe special thanks to M. Feith, Springer Verlag, for professional editorship and continuous encouragement at several critical junctions. We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 373 “Quantifikation und Simulation Ökonomischer Prozesse”.

Berlin, May 1999

## Contents of the CD-ROM

<b>Acrobat_Reader</b>	
ar32e301.exe	Setup file for Acrobat Reader 3.01
<b>e-books</b>	
mva.pdf	Multivariate Analysis textbook in PDF format
xplbook.pdf	This Learning Guide in PDF format
xplbook.html	This Learning Guide in HTML format
<b>Netscape</b>	
cc32e46.exe	Setup file for Netscape Communicator 4.6
<b>Readme_Installation</b>	
	Installation instructions and Registration Form
<b>XploRe</b>	
Setup.exe	Setup file for XploRe
<b>XploRe-Client</b>	
xplinst.exe	XploRe Java client for worldwide Quantlet Service