



COMPSTAT

Proceedings
in Computational Statistics

15th Symposium Held in Berlin,
Germany, 2002

Edited by
Wolfgang Härdle
Bernd Rönz



With 205 Figures
and 89 Tables

Professor Dr. Wolfgang Härdle
Humboldt-Universität zu Berlin
CASE – Centre of Applied Statistics and Economics
Spandauer Straße 1
10178 Berlin, Germany

Professor Dr. Bernd Rönz
Humboldt-Universität zu Berlin
Wirtschaftswissenschaftliche Fakultät
Institut für Statistik und Ökonometrie
Spandauer Straße 1
10178 Berlin, Germany

ISBN 978-3-7908-1517-7 ISBN 978-3-642-57489-4 (eBook)
DOI 10.1007/978-3-642-57489-4

Library of Congress Cataloging-in-Publication Data applied for
Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Proceedings in Computational Statistics:
15th Symposium Held in Berlin, Germany, 2002; with 89 Tables /
COMPSTAT. Ed. by Wolfgang Härdle; Bernd Rönz. –
Heidelberg; New York: Physica-Verl., 2002

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 Physica-Verlag. Violations are liable for prosecution under the German Copyright Law.

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2002
Originally published by Physica-Verlag Heidelberg New York in 2002

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.

Cover design: Erich Kirchner, Heidelberg

SPIN 10886369 88/2202-5 4 3 2 1 0 – Printed on acid-free paper

Preface

This COMPSTAT 2002 book contains the Keynote, Invited, and Full Contributed papers presented in Berlin, August 2002. A companion volume including Short Communications and Posters is published on CD.

The COMPSTAT 2002 is the 15th conference in a serie of biannual conferences with the objective to present the latest developments in Computational Statistics and is taking place from August 24th to August 28th, 2002.

Previous COMPSTATs were in Vienna (1974), Berlin (1976), Leiden (1978), Edinburgh (1980), Toulouse (1982), Prague (1984), Rome (1986), Copenhagen (1988), Dubrovnik (1990), Neuchâtel (1992), Vienna (1994), Barcelona (1996), Bristol (1998) and Utrecht (2000).

COMPSTAT 2002 is organised by CASE, Center of Applied Statistics and Economics at Humboldt-Universität zu Berlin in cooperation with Freie Universität Berlin and University of Potsdam.

The topics of COMPSTAT include methodological applications, innovative software and mathematical developments, especially in the following fields: statistical risk management, multivariate and robust analysis, Markov Chain Monte Carlo Methods, statistics of E-commerce, new strategies in teaching (Multimedia, Internet), computerbased sampling/questionnaires, analysis of large databases (with emphasis on computing in memory), graphical tools for data analysis, classification and clustering, new statistical software and historical development of software.

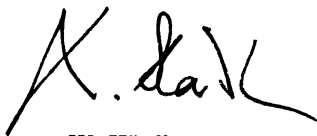
A conference like COMPSTAT is not possible without an active scientific input from many high class researchers. The contributions were classified by the Scientific Program Committee (SPC): Prof. Wolfgang Härdle (Berlin), Prof. Antony Unwin (Augsburg), Prof. Jaromir Antoch (Prague), Prof. Adrian Bowman (Glasgow), Prof. Michel Delecroix (Paris), Prof. Wenceslao Gonzalez Manteiga (Santiago de Compostela), Prof. Junji Nakano (Tokyo), Prof. Michael Schimek (Graz) and Prof. Peter van der Heijden (Utrecht).

The papers were refereed by the SPC and we would like to thank our colleagues for all their help. The resulting proceedings provide a broad overview of the current research areas in statistical computing.

We are also very grateful to the members of the Local Organizing Committee: Dr. Yasemin Boztug, Prof. Dr. Herbert Büning, Prof. Dr. Wolfgang Härdle, Prof. Dr. Lutz Hildebrandt, Dr. Sigbert Klinke, Prof. Dr. Uwe Küchler, Prof. Dr. Bernd Rönz, Dr. Peter Schirmbacher, Dipl.-Ing. Benjamin Schüler, Prof. Dr. Vladimir Spokoiny, Prof. Dr. Hans Gerhard Strohe, Prof. Dr. Jürgen Wolters and Uwe Ziegenhagen.

We would also like to thank Patricia Ahrend and Luise Richter who together with Dr. Klinke did a great job in managing the papers.

Berlin, August 2002:



W. Härdle



B. Rönz

Contents

Preface	v
Invited Papers	1
An Implementation for Regression Quantile Estimation	3
<i>T. W. Yee</i>	
Computational Methods for Time Series Analysis	15
<i>G. Kitagawa, T. Higuchi, S. Sato</i>	
Forecasting PC-ARIMA Models for Functional Data	25
<i>M.J. Valderrama, F.A. Ocaña, A.M. Aguilera</i>	
KyPlot as a Tool for Graphical Data Analysis	37
<i>K. Yoshioka</i>	
Mice and Elephants Visualization of Internet Traffic	47
<i>J.S. Marron, F. Hernandez-Campos, F.D. Smith</i>	
Relativity and Resolution for High Dimensional Information Visualization with Generalized Association Plots (GAP)	55
<i>S.-C. Chang, C.-H. Chen, Y.-Y. Chi, C.-W. Ouyoung</i>	
Supervised Learning from Microarray Data	67
<i>T. Hastie, R. Tibshirani, B. Narasimhan, G. Chu</i>	
Teaching Statistics with Electronic Textbooks	79
<i>J. Symanzik, N. Vukasinovic</i>	
Trans-Dimensional Markov Chains and their Applications in Statistics . .	91
<i>S.P. Brooks</i>	
Full Papers	103
A Bayesian Model for Compositional Data Analysis	105
<i>M.J. Brewer, C. Soulsby, S.M. Dunn</i>	
A Comparison of Marginal Likelihood Computation Methods	111
<i>C.S. Bos</i>	
A Hotelling Test Based on MCD	117
<i>G. Willems, G. Pison, P.J. Rousseeuw, S. van Aelst</i>	
A Resampling Approach to Cluster Validation	123
<i>V. Roth, T. Lange, M. Braun, J. Buhmann</i>	
A Self Documenting Programming Environment for Weighting	129
<i>W. Grossmann, P. Ofner</i>	
A State Space Model for Non-Stationary Functional Data	135
<i>M. Ortega-Moreno, M.J. Valderrama, J.C. Ruiz-Molina</i>	
A Wildlife Simulation Package (WiSP)	141
<i>W. Zucchini, M. Erdelmeier, D. Borchers</i>	
Algorithmical and Computational Procedures for a Markov Model in Survival Analysis	147
<i>J.E. Ruiz-Castro, R. Pérez-Ocón, D. Montoro-Cazorla</i>	

An Algorithm for the Construction of Experimental Designs with Fixed and Random Blocks	153
<i>P. Goos, A. N. Donev, M. Vandebroek</i>	
An Algorithm to Estimate Time Varying Parameter SURE Models under Different Type of Restrictions	159
<i>S. Orbe, E. Ferreira, J.M. Rodríguez-Póo</i>	
Analyzing Data with Robust Multivariate Methods and Diagnostic Plots	165
<i>G. Pison, S. van Aelst</i>	
Application of "Aggregated Classifiers" in Survival Time Studies	171
<i>A. Benner</i>	
Application of Hopfield-like Neural Networks to Nonlinear Factorization	177
<i>D. Husek, A.A. Frolov, H. Rezankova, V. Snasel</i>	
Bagging Tree Classifiers for Glaucoma Diagnosis	183
<i>T. Hothorn, B. Lausen</i>	
Bayesian Automatic Parameter Estimation of Threshold Autoregressive (TAR) Models using Markov Chain Monte Carlo (MCMC)	189
<i>E. Amiri</i>	
Bayesian Semiparametric Seemingly Unrelated Regression	195
<i>S. Lang, S.B. Adebayo, L. Fahrmeir</i>	
Blockmodeling Techniques for Web Mining	201
<i>G. Schoier</i>	
Bootstrapping Threshold Autoregressive Models	207
<i>J. Öhrvik, G. Schoier</i>	
Canonical Variates for Recursive Partitioning in Data Mining	213
<i>G. Cappelli, C. Conversano</i>	
CA _{no} Va _© : a Software for Causal Modeling	219
<i>O. Wüthrich-Martone, C. Nachtigall, M. Müller, R. Steyer</i>	
Classification Based on the Support Vector Machine, Regression Depth, and Discriminant Analysis	225
<i>A. Christmann, P. Fischer, T. Joachims</i>	
Clockwise Bivariate Boxplots	231
<i>A. Corbellini</i>	
Combining Graphical Models and PCA for Statistical Process Control	237
<i>R. Fried, U. Gather, M. Imhoff, M. Keller, V. Lanius</i>	
Comparing Two Partitions: Some Proposals and Experiments	243
<i>G. Saporta, G. Youness</i>	
Comparison of Nested Simulated Annealing and Reactive Tabu Search for Efficient Experimental Designs with Correlated Data	249
<i>N. Coombes, R. Payne, P. Lisboa</i>	
Computational Connections between Robust Multivariate Analysis and Clustering	255
<i>D.M. Rocke, D.L. Woodruff</i>	
Computer Intensive Methods for Mixed-effects Models	261
<i>J.A. Sanchez, J. Ocaña</i>	
Construction of T-Optimum Designs for Multiresponse Dynamic Models	267
<i>D. Uciński, B. Bogacka</i>	
Data Compression and Selection of Variables, with Respect to Exact Inference	273
<i>J. Läuter, S. Kropf</i>	
Data Extraction from Dense 3-D Surface Models	279
<i>M. Bock, A. Bowman, J. Bowman, P. Siebert</i>	
Detection of Locally Stationary Segments in Time Series	285
<i>U. Ligges, C. Weihs, P. Hasse-Becker</i>	
Detection of Outliers in Multivariate Data: A Method Based on Clustering and Robust Estimators	291
<i>C.M. Santos-Pereira, A.M. Pires</i>	
Development of a Framework for Analyzing Process Monitoring Data with Applications to Semiconductor Manufacturing Process	297
<i>Y.-H. Yoon, Y.-S. Kim, S.-J. Kim, B.-J. Yum</i>	

Different Ways to See a Tree – KLIMT	303
<i>S. Urbanek</i>	
e-stat: A Web-based Learning Environment in Applied Statistics	309
<i>E. Cramer, K. Cramer, U. Kamps</i>	
e-stat: Automatic Evaluation of Online Exercises	315
<i>K. Bartels</i>	
e-stat: Basic Stochastic Finance at School Level	321
<i>C. Mohn, D. Pfeifer</i>	
e-stat: Development of a Scenario for Statistics in Chemical Engineering	327
<i>C. Weihs, M. Kappler</i>	
e-stat: Web-based Learning and Teaching of Statistics in Secondary Schools	333
<i>C. Pahl, P. Lipinski, K. Reiss</i>	
EMILeA-stat: Structural and Didactic Aspects of Teaching Statistics through an Internet-based, Multi-medial Environment	339
<i>U. Genschel, U. Gather, A. Busch</i>	
Evaluating the GPH Estimator via Bootstrap Technique	343
<i>S. Golia</i>	
Evolutionary Algorithms with Competing Heuristics in Computational Statistics	349
<i>J. Tvrđík, I. Krivý, L. Mišík</i>	
Exact Nonparametric Inference in R	355
<i>T. Hothorn, K. Hornik</i>	
Exploring the Structure of Regression Surfaces by using SiZer Map for Additive Models	361
<i>R. Raya Miranda, M.D. Martínez Miranda, A. González Carmona</i>	
Fast and Robust Filtering of Time Series with Trends	367
<i>R. Fried, U. Gather</i>	
Functional Principal Component Modelling of the Intensity of a Doubly Stochastic Poisson Process	373
<i>A.M. Aguilera, P.R. Bouzas, N. Ruiz-Fuentes</i>	
Growing and Visualizing Prediction Paths Trees in Market Basket Analysis	377
<i>M. Ária, F. Mola, R. Siciliano</i>	
Improved Fitting of Constrained Multivariate Regression Models using Automatic Differentiation	383
<i>T. Ringrose, S. Forth</i>	
Imputation of Continuous Variables Missing at Random using the Method of Simulated Scores	389
<i>G. Calzolari, L. Neri</i>	
Induction of Association Rules: Apriori Implementation	395
<i>C. Borgelt, R. Kruse</i>	
Intelligent WBT: Specification and Architecture of the Distributed Multimedia e-Learning System e-stat	401
<i>C. Möbus, B. Albers, S. Hartmann, J. Zurborg</i>	
Interactive Exploratory Analysis of Spatio-Temporal Data	407
<i>J.M. Dreesman</i>	
Interactive Graphics for Data Mining	413
<i>D. Di Benedetto</i>	
Least Squares Reconstruction of Binary Images using Eigenvalue Optimization	419
<i>S. Chrétien, F. Corset</i>	
Locally Adaptive Function Estimation for Categorical Regression Models	425
<i>A. Jerak, S. Lang</i>	
Maneuvering Target Tracking by using Particle Filter Method with Model Switching Structure	431
<i>N. Ikoma, T. Higuchi, H. Maeda</i>	
mathStatica: Mathematical Statistics with <i>Mathematica</i>	437
<i>C. Rose, M.D. Smith</i>	

MCMC Model for Estimation Poverty Risk Factors using Household Budget Data	443
<i>E. Käärrik, E.-M. Tiit, M. Vähi</i>	
MD*Book online & e-stat: Generating e-stat Modules from LaTeX	449
<i>R. Witzel, S. Klinke</i>	
Missing Data Incremental Imputation through Tree Based Methods	455
<i>C. Conversano, C. Cappelli</i>	
Missing Values Resampling for Time Series	461
<i>A.M. Alonso, D. Peña, J.J. Romo</i>	
ModelBuilder – an Automated General-to-specific Modelling Tool	467
<i>M. Kurcewicz</i>	
On the Use of Particle Filters for Bayesian Image Restoration	473
<i>K. Nittono, T. Kamakura</i>	
Optimally Trained Regression Trees and Occam’s Razor	479
<i>P. Šavický, J. Klaschka</i>	
Parallel Algorithms for Inference in Spatial Gaussian Models	485
<i>M. Whitley, S.P. Wilson</i>	
Parameters Estimation of Block Mixture Models	491
<i>M. Nadif, G. Govaert</i>	
Pattern Recognition of Time Series using Wavelets	497
<i>E.A. Maharaj</i>	
Representing Knowledge in the Statistical System Jasp	503
<i>I. Kobayashi, Y. Yamamoto, T. Fujwara</i>	
Robust Estimation with Discrete Explanatory Variables	509
<i>P. Čížek</i>	
Robust Principal Components Regression	515
<i>S. Verboven, M. Hubert</i>	
Robust Time Series Analysis through the Forward Search	521
<i>L. Grossi, M. Riani</i>	
Rough Sets and Association Rules – Which is Efficient?	527
<i>D. Delic, H.-J. Lenz, M. Neiling</i>	
Skewness and Fat Tails in Discrete Choice Models	533
<i>R. Capobianco</i>	
Standardized Partition Spaces	539
<i>U. Sondhauß, C. Weihs</i>	
StatDataML: An XML Format for Statistical Data	545
<i>D. Meyer, F. Leisch, T. Hothorn, K. Hornik</i>	
Statistical Computing on Web Browsers with the Dynamic Link Library	551
<i>A. Takeuchi, H. Yadohisa, K. Yamaguchi, C. Asano, M. Watanabe</i>	
Statistical Inference for a Robust Measure of Multiple Correlation	557
<i>C. Dehon, C. Croux</i>	
Statistical Software VASMM for Variable Selection in Multivariate Methods	563
<i>M. Iizuka, Y. Mori, T. Tarumi, Y. Tanaka</i>	
Structural Equation Models for Finite Mixtures – Simulation Results and Empirical Applications	569
<i>D. Temme, J. Williams, L. Hildebrandt</i>	
Sweave: Dynamic Generation of Statistical Reports using Literate Data Analysis	575
<i>F. Leisch</i>	
Testing for Simplification in Spatial Models	581
<i>L. Scaccia, R.J. Martin</i>	
The Forward Search	587
<i>A. Atkinson</i>	
The MISSION Client: Navigating Ontology Information for Query Formulation and Publication in Distributed Statistical Information Systems	593
<i>Y. Bi</i>	

Time Series Modelling using Mobile Devices and Broadband Internet . . .	599
<i>A. Prat</i>	
Unbiased Partial Spline Fitting under Autoregressive Errors	605
<i>M.G. Schimek</i>	
Unobserved Heterogeneity in Store Choice Models	611
<i>I.R. del Bosque, A. Suárez-Vázquez, I. Moral-Arce, J.M. Rodríguez-Póo</i>	
Using the Forward Library in S-plus	617
<i>K. Konis</i>	
Variance Stabilization and Robust Normalization for Microarray Gene Expression Data	623
<i>A. von Heydebreck, W. Huber, A. Poustka, M. Vingron</i>	
Weights and Fragments	629
<i>S. Morgenthaler</i>	
XQS/MD*Crypt as a Means of Education and Computation	635
<i>J. Feuerhake</i>	
Author's Index	641
Keyword Index	645