ICORS14
International Conference on Robust Statistics
Martin-Luther-University Halle-Wittenberg
August 10–15, 2014

Conference Guide &
Book of Abstracts

Claudia Becker
Kristin Dombrowski
Thomas Kirschstein
Steffen Liebscher
Giovanni Porzio
Giancarlo Ragozini
Lisiane Schnegelsberg
(local organizers)

Halle (Saale), Germany
## Contents

1 Foreword ................................................................. 2

2 Steering Committee ...................................................... 3

3 General Information ...................................................... 4

4 Scientific Program ....................................................... 6

5 Abstracts ................................................................. 10

6 Social Program .......................................................... 66

7 Halle@Night ............................................................. 67

8 Support & Sponsoring .................................................... 69

9 List of Participants ...................................................... 70
Dear participants,

it is our great pleasure to welcome you all to the ICORS 2014 conference in the beautiful city of Halle at the river Saale!

We all visit many rather different conferences, the huge ones and the small ones, the national and international ones. Some of us may be members of other scientific families as well, yet I think this robust family is exceptional.

We never defined a strict null model such as an international association of robust statistics. Nevertheless we meet at ICORS – it’s the family reunion working on a broader distributional neighborhood.

We have a steering committee for ICORS, but no one is professionally organizing the organizers, determining the dates, etc. ICORS is an outlier in the set of scientific conferences. Nevertheless members of our family volunteer in organizing ICORS – we are well aware that outliers may be of a certain value in themselves.

We experienced ICORS in various countries, in larger and smaller places, in urban and in rural environments. They were always coined by the personalities of the local organizers, meaning no ICORS was in any way average!

As with all academic families, the "seniors" take their academic offspring to the family reunion. So, particularly to all "offspring" being here this (first) time: meet the family!

We prepared a rich feast of scientific sessions to satisfy your academic hunger. Strive through the pages of this booklet to arrange your menu. But do not forget to satisfy your cultural appetite as well – you are in an environment with more than 1,200 years of history. And, finally, we take the "feast" rather literally: enjoy food and drinks and learn the lightness of being in the German-Italian branch of the family!

Claudia Becker
2 Steering Committee

Peter Rousseuw - chairman (Belgium)
Claudio Agostinelli (Italy)
Olcay Arslan (Turkey)
Claudia Becker (Germany)
Ana Bianco (Argentina)
Graciela Boente (Argentina)
Andrea Cerioli (Italy)
Frank Critchley (UK)
Christophe Croux (Belgium)
Juan Cuesta Albertos (Spain)
Laurie Davies (Germany)
Rudolf Dutter (Austria)
Luisa Fernholz (USA)
Chris Field (Canada)
Peter Filzmoser (Austria)
Ursula Gather (Germany)
Alfonso Gordaliza (Spain)
Marc Hallin (Belgium)
Xuming He (USA)
Mia Hubert (Belgium)
Jana Jureckova (Czech Republic)
Ricardo Maronna (Argentina)
Carlos Matran (Spain)
Stephan Morgenthaler (Switzerland)
Hannu Oja (Finland)
Daniel Pena (Spain)
Ana Pires (Portugal)
Marco Riani (Italy)
Isabel Rodrigues (Portugal)
Elvezio Ronchetti (Switzerland)
Georgy Shevlyakov (Russia)
David Tyler (USA)
Stefan Van Aelst (Belgium)
Roy Welsch (USA)
Victor Yohai (Argentina)
Ruben Zamar (Canada)
3 General Information

• Conference Venue

The conference will take place at the downtown campus (Melanchthonianum, Universitätsplatz 9) of the Martin-Luther-University in the center of Halle. The campus is located within walking distance (400m) from the Market place, which in turn is easily accessible by tram (lines 1, 2, 3, 5, 7, 8, 10, 16).

• Conference Office/Registration

The conference office is located in room "Sitzungszimmer" at the first floor of the Melanchthonianum (Universitätsplatz 9). The office is open between 8.30am and 5.30pm on Monday, Tuesday, Thursday and between 8.30am and 1pm on Wednesday and Friday. All participants should register at the conference desk once they arrive. Participants arriving on Sunday (August 10) can get their conference bag at the Welcome Reception (Große Steinstraße 73) between 5pm–9pm. In case of any concerns or questions don’t hesitate to contact the organizers at the conference office.
• Information for Presenters

All presentations (including the Keynotes) take place in room "XX" at the second floor of the Melanchthonianum (Universitätsplatz 9). The room is equipped with a Windows-PC, a beamer, an overhead projector and a board. It is also possible to connect an own laptop. Presenters are advised to appear in the lecture room 10min before their session starts in order to contact the session chair about their attendance and to copy the presentation. Presentations should either be in Powerpoint or PDF format.

• Poster Presentation

Poster displays are located at the first floor of the Melanchthonianum. Pins can be obtained at the conference office.

• Lunches/Coffee Breaks

Coffee breaks take place at the first floor of the Melanchthonianum. Lunch is served in the Bella SoSo (Universitätsring 6/6a).

• Internet Access

Free internet access is available using the wireless LAN of the Martin-Luther-University which covers large parts of the Universitätsplatz. Please use the following log-in data: TBA.

• Local Transportation

The best way to go around Halle is by tram. Tickets are available at the driver and at ticket machines (not available at all stations). Be advised that single tickets need to be stamped within the tram. You may also consider buying a weekly ticket (17.60€) either at http://www.havag.com/ or at the Havag Service Center at the market place (Marktplatz 11). Alternatively, cabs are waiting at numerous places in Halle. They are usually colored light yellow and wear the label "Taxi".

• Contact

In case of any concerns or questions please contact the conference office. In urgent cases directly contact any of the organizers (yellow name badge).

In case of an EMERGENCY call 110 (police) or 112 (ambulance/fire department).
4 Scientific Program

- Timetable

<table>
<thead>
<tr>
<th>Time</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>09.00-10.20</td>
<td>opening; then KN1</td>
<td>CS4</td>
<td>CS7</td>
<td>CS8</td>
<td>09.30-10.50: CS11</td>
</tr>
<tr>
<td>10.20-11.00</td>
<td>coffee break/poster</td>
<td>coffee break/poster</td>
<td>coffee break/poster</td>
<td>coffee break/poster</td>
<td>10.50-11.30: coffee break/poster</td>
</tr>
<tr>
<td>11.00-12.00</td>
<td>CS1</td>
<td>KN2</td>
<td>KN3</td>
<td>KN4</td>
<td>11.30-12.30: CS12</td>
</tr>
<tr>
<td>12.00-13.40</td>
<td>lunch</td>
<td>lunch</td>
<td>excursion</td>
<td>lunch</td>
<td>12.30-12.40: closing ceremony</td>
</tr>
<tr>
<td>13.40-15.00</td>
<td>CS2</td>
<td>CS5</td>
<td>excursion</td>
<td>CS9</td>
<td>12.40: lunch</td>
</tr>
<tr>
<td>15.00-15.40</td>
<td>coffee break/poster</td>
<td>coffee break/poster</td>
<td>excursion</td>
<td>coffee break/poster</td>
<td></td>
</tr>
<tr>
<td>15.40-17.20</td>
<td>CS3</td>
<td>CS6</td>
<td>excursion</td>
<td>CS10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17.30: guided city tour</td>
<td>17.20: steering committee</td>
<td></td>
<td></td>
<td>19.00: dinner</td>
</tr>
</tbody>
</table>

- Keynote Speaker

  KN1: Xuming He (University of Michigan), chair: Georgy Shevlyakov
  KN2: Victor Yohai (University of Buenos Aires), chair: Peter Rousseeuw
  KN3: Jana Jureckova (Charles University, Prague), chair: Christophe Croux
  KN4: Hannu Oja (University of Turku), chair: David Tyler

- CS1 (chair: Stefan Van Aelst)

  11.00-11.20 Jan Amos Visek (Treating the heteroscedasticity in robust estimation)
  11.20-11.40 Fatma Zehra Dogru (Robust mixture regression modelling based on the skew t distribution)
  11.40-12.00 Helena Ageeva (Statistical inferences for multiple regression under grouping distortions)
• CS2 (chair: Peter Filzmoser)
  13.40-14.00 Valentin Todorov (A Robust Tucker3 Model for Compositional Data)
  14.00-14.20 Amelia Kelly (Robust Integrative Analysis of Multi-block Contaminated Dataset)
  14.20-14.40 George Luta (Covariate-Adjusted Confidence Intervals for Relative Treatment Effects)
  14.40-15.00 Christophe Croux (Robust Sparse Canonical Correlation Analysis)

• CS3 (chair: Chris Field)
  15.40-16.00 Viktoria Öllerer (Sparse S- and MM-regression)
  16.00-16.20 Peter Filzmoser (Sparse partial robust M regression)
  16.20-16.40 Mara Velina (A Comparison of Robust Empirical Likelihood-based ANOVA Methods)
  16.40-17.00 Stefan Van Aelst (Least angle regression with categorical variables)

• CS4 (chair: Claudio Agostinelli)
  09.00-09.20 ArunKumar Kuchibhotla (A Minimum Distance Weighted Likelihood Method of Estimation)
  09.20-09.40 Anand Vidyashankar (Robust and Efficient Stochastic Optimization via Disparities)
  09.40-10.00 Nirian Martin (Robustness of phi-divergence test statistics based on exponentially tilted empirical likelihood estimators)
  10.00-10.20 Vartan Ohanes Choulakian (Taxicab Correspondence Analysis)

• CS5 (chair: Christine Müller)
  13.40-14.00 Stephanie Aerts (Robustness and efficiency of multivariate coefficients of variation)
  14.00-14.20 Georgy Shevlyakov (Robust M-estimates of the correlation coefficient in bivariate independent component distribution models)
  14.20-14.40 Alexey Kharin (Robustness of sequential decision making on parameters of the model under distortions)
  14.40-15.00 Natalia Stepanova (Goodness-of-fit tests based on sup-functionals of weighted empirical processes)
• CS6 (chair: Werner Stahel)
  15.40-16.00  Christine Müller (Data depth for autoregression)
  16.00-16.20  Pieter Segaert (Depth-based classification of multivariate functional data)
  16.20-16.40  Firat Ozdemir (Comparing k Independent Groups With a Method Based on Trimmed Means)
  16.40-17.00  Andrea Cerasa (Modelling international trade data with the Tweedie distribution)
  17.00-17.20  Neyko Neykov (Robust variable selection in joint modeling of location, scale and shape for high dimensional data through trimming)

• CS7 (chair: Gentiane Haesbroeck)
  09.00-09.20  Pavel Cizek (Robust GMM estimation and moment selection in dynamic panel data models)
  09.20-09.40  Markus Asser Naftali Matilainen (Blind source separation for multivariate conditionally heteroscedastic time series)
  09.40-10.00  Mehmet Niyazi Cankaya (Robust Estimation for the Parameters of the Distributed Lag Models Based on Heavy Tailed Skew Distributions)
  10.00-10.20  Masoud Yarmohammadi (Comparing Singular Spectrum Analysis and Artificial Neural Networks in forecasting exchange rates)

• CS8 (chair: Graciela Boente)
  09.00-09.20  Samuel Orso (Bounded-influence robust estimation of copulas)
  09.20-09.40  Peter Ruckdeschel (Optimally Robust Covariances (Also Covering Weighted Observations))
  09.40-10.00  Andrea Cerioli (On the properties of the multivariate Forward Search estimator)
  10.00-10.20  Alfonso Garcia-Perez (A von Mises approximation to the small sample distribution of the trimmed mean)

• CS9 (chair: Peter Ruckdeschel)
  14.00-14.20  Ximing Xu (Robust state space models for estimating fish stock maturities)
  14.20-14.40  Debbie Dupuis (Robust Value-at-Risk Estimation)
  14.40-15.00  Yogesh Kumar Bichpuriya (Robust probability density forecast of yearly peak load)
• CS10 (chair: Alfonso Garcia-Perez)
  15.40-16.00 Shojaeddin Chenouri (Robust Dimension Reduction)
  16.00-16.20 Graciela Boente (Robust backfitting estimators for additive models)
  16.20-16.40 David Tyler (Regularized M-estimators of scatter matrices)
  16.40-17.00 Rauf Ahmad (A simple and practical solution to the classical multivariate
    Behrens-Fisher problem under non-normality)
  17.00-17.20 Davide Buttarazzi (A boxplot for circular data)

• CS11 (chair: Valentin Todorov)
  09.30-09.50 Silvelyn Zwanzig (On RM-Estimators in Regression)
  09.50-10.10 Kang-Mo Jung (Weighted regression estimator with SCAD)
  10.10-10.30 Chris Chatzinakos (A new Algorithm for Least Trimmed Absolute Deviation
    (LTAD) Location and Regression)
  10.30-10.50 Mark Hannay (Redescending psi in GLM (without Mallows weights) and a
    new robust test for forward search)

• CS12 (chair: Andrea Cerioli)
  11.30-11.50 Claudio Agostinelli (Robust estimators of the generalized loggamma regression
    model with censored observations)
  11.50-12.10 Yogesh Kumar Bichpuriya (On optimal combination of probability density
    forecasts using Kullback-Leibler divergence criterion)
  12.10-12.30 Giuseppe Pandolfo (Nonparametric classification of circular data)

• Poster
  – Yetkin Tuac (Analysis of Regression Models with AR(1) Error Terms Based on Skew
    Distributions: Parameter Estimation)
  – Yesim Güney (Robust Estimation for the Parameter of the Zipfian Distribution: Properties
    and Applications)
  – Said Alkarni (A compound class of geometric and lifetimes distributions)
5 Abstracts

Keynote Speaker ......................................................... 11
CS1 ................................................................. 15
CS2 ................................................................. 19
CS3 ................................................................. 23
CS4 ................................................................. 27
CS5 ................................................................. 31
CS6 ................................................................. 35
CS7 ................................................................. 40
CS8 ................................................................. 44
CS9 ................................................................. 48
CS10 .............................................................. 51
CS11 .............................................................. 56
CS12 .............................................................. 60
Poster ............................................................... 63
Robust low-rank data matrix approximation

Xuming He$^{1,\ast}$

$^{1}$ University of Michigan, USA; xmhe@umich.edu
$\ast$ Corresponding author

Abstract. The singular value decomposition is widely used to approximate data matrices with lower rank matrices. Low rank approximation is closely related to traditional multivariate techniques such as principal component analysis and factor analysis, and is also used in a test of uni-dimensionality of the mean of a data matrix as proposed by Feng and He [Ann. Appl. Stat. 3 (2009) 1634-1654]. However, the first singular values and vectors can be driven by a small number of outlying measurements, and low rank approximations could mislead in the presence of outliers. In this talk, we discuss robust alternatives that moderate the effect of outliers in low-rank approximations. Factors of consideration include algorithmic convergence, statistical efficiency, and power for detecting the need for additional singular structures in the approximating matrix. We use oligonucleotide gene microarray data to demonstrate how robust singular value decomposition compares with its traditional counterparts. Examples show that the robust methods often lead to a more meaningful assessment of the dimensionality of gene intensity data matrices. The talk is based on joint work with Dr. Xingdong Feng at Shanghai University of Finance and Economics.

Keywords.
High finite-sample efficiency and robustness based on distance-constrained maximum likelihood

R.A. Maronna\textsuperscript{1} and V.J. Yohai\textsuperscript{2,*}

\textsuperscript{1} Universidad Nacional de La Plata; rmaronna@retina.ar
\textsuperscript{2} Universidad de Buenos Aires; victoryohai@gmail.com
* Corresponding author

Abstract. Good robust estimators can be tuned to combine a high breakdown point and a specified asymptotic efficiency at a central model. This happens in regression with MM and \( \tau \) estimators among others. However, the finite-sample efficiency of these estimators can be much lower than the asymptotic one. To overcome this drawback, an approach is proposed for parametric models, which is based on a distance between parameters. Given a robust estimator, the proposed one is obtained by maximizing the likelihood under the constraint that the distance is less than a given threshold. For the linear model with normal errors and using the MM estimator and the distance induced by the Kullback-Leibler divergence, simulations show that the proposed estimator attains a finite-sample efficiency close to one, while its maximum mean squared error is smaller than that of the MM estimator. In this case we found the asymptotic distribution which is non normal. Using this result we are able to obtain asymptotic confidence intervals for the regression coefficients. We also prove that the proposed estimator keeps the breakdown point of the initial estimator. The same approach is used to estimate multivariate location and scatter using as starting point an S estimator. Monte Carlo results show that in this case the finite sample efficiency of the proposed estimator is higher than that of the S estimator while keeping a similar degree of robustness.

Keywords. Finite-sample efficiency; Contamination bias; MM estimators; S estimators; Linear model; Scatter matrix
Averaged regression and extreme regression quantile

Jana Jurečková

1 Charles University in Prague; jurecko@karlin.mff.cuni.cz
* Corresponding author

Abstract. We show that the averaged standardized alpha-regression quantile equals, for every n>p, to the weighted mean of a subvector of p model errors, corresponding to the optimal base of the respective linear program; p is the dimension of regression parameter. It supplements the asymptotic result, showing that the averaged standardized alpha-regression quantile is asymptotically equivalent to the empirical alpha-quantile of model errors. The special case is the averaged extreme regression quantile, which is equal to the weighted mean of p largest order statistics of model errors.

Keywords. Alpha-regression quantile, Extreme regression quantile, Two-step regression quantile, Averaged version of regression quantile.
Scatter matrices and linear dimension reduction

H. Oja

1 University of Turku, Finland; Hanna.Oja@utu.fi

Abstract. In this talk, unsupervised and supervised linear dimension reduction methods based on various uses of scatter matrices are addressed. In our approach, dimension reduction is done by preprocessing the data, that is, by transforming the data to an invariant coordinate system (ICS). The unsupervised ICS functionals in Tyler et al. (2009) are based on the use of two unsupervised scatter matrix functionals while the supervised functionals such as the sliced inverse regression (SIR) utilize both unsupervised and supervised scatter matrices (Liski et al., 2013). See also Ilmonen et al. (2012). The use of several scatter functionals is discussed as well.

In the so called Blind Source Separation (BSS) model, it is assumed that the observed $p$-variate observations are all linear combinations of $p$ latent variables. This may be seen also as a dimension reduction problem if only few of the latent variables are relevant in data analysis. The aim is then to find an estimate for an unmixing matrix, which transforms the observed variables back to the latent variables. The most popular BSS approach is independent component analysis (ICA), and the so called AMUSE and SOBI solutions are used to find latent stationary time series in the analysis of multivariate time series. Unmixing matrix estimates are often found as ICS or SICS solutions. See e.g. Miettinen et al. (2014).

Typically, dimension reduction procedures proposed in the literature are highly non-robust as the first step in the algorithms is to standardize the observations using the regular covariance matrix. In this talk, we also briefly discuss the possibility to robustify these procedures. The talk is based on joint work with several coauthors.

Keywords. Blind Source Separation; Independent component analysis; Invariant coordinate system; Sliced inverse regression; Time series.

References


T reating the heteroscedasticity in robust estimation

J. Á. Víšek

Abstract. Having ignored the heteroscedasticity (when employing the LS or ML) can lead to the misspecification of model with the decreased quality of identification of the underlying model. Therefore the classical regression paid a lot of attention to it, including tools for detecting the heteroscedasticity (Breusch, Pagan (1979), Cook, Weisberg (1983)), for evaluating properly the significance of explanatory variables (White, 1980) or for removing the heteroscedasticity (Wooldridge (2001, 2006); Fahrmeir et al. (2013), among many others).

In the robust regression we did exercise an effort to cope with the problem, too. There are several (asymptotic) theoretical findings (Koenker, Bassett Jr (1982); Carroll, Ruppert (1982); Erceg-Hurn, Mirosevich (2008); Víšek (2002, 2008), giving some among many others). Unfortunately, the most of them suffer by the loss of regression equivariance (Bickel, 1975). Moreover, the results for finite samples established by simulations are lacking completely.

The paper addresses both theoretical as well as application part of problem when the underlying model is estimated by the least weighted squares (LWS) - the robustified version of LS. It shows that the diagnostic tools tailored for the LS can be modified for the LWS (although involved technicalities occasionally appeared). It also demonstrates that due to the smooth weighting down of influential observations, the estimator can cope with mild heteroscedasticity without any additional steps. Finally, it shows that the estimation of the model of heteroscedasticity by LWS gives a good results even under heavy contamination and/or high heteroscedasticity.

Keywords. Regression; Heteroscedasticity; Detecting it; Depressing it; Removing it.

References


Robust mixture regression modelling based on the skew t distribution

F. Z. Doğru1,∗ and O. Arslan1

1 Ankara University, Faculty of Science, Department of Statistics, 06100, Ankara/Turkey; fzdogru@ankara.edu.tr, oarslan@ankara.edu.tr

Abstract. The parameter estimation of mixture regression models using the normal distribution is sensitive to outliers and heavy-tailed error distributions. To model the heavy tailed errors in mixture regression models, Yao et al. (2014) propose to use mixtures of t distribution. In this work, we study the robust mixture regression based on the skew t distribution to model the skewness and heavy tailedness in mixture regression (Doğru, 2014). We will use the skew t distribution introduced by Azzalini & Capitaino (2003). Since the skew t distribution can be written as a scale mixture of normal, truncated normal and gamma distributions, we present an EM algorithm (Dempster et al., 1977) to compute the estimates for the unknown regression parameters. Finally, the performance of proposed method is demonstrated by a simulation study and a real data example.

Keywords. EM algorithm; Mixture regression models; Skew t distribution.

References


Abstract. Applying classical models to real data can be inefficient because most observations are distorted, coarsed or incomplete (Kharin, 2013). One of the possible distortions is grouping, i.e. when instead of knowing the exact value we observe only a category it belongs to. We focus on a particular case of grouping in regression. Let

\[ Y_t = F^0(X_t, \theta^0) + u_t, \quad t = 1, \ldots, n, \]

be a regression model, where \( X_t \in X \subseteq \mathbb{R}^N, t = 1, \ldots, n, \) are independent regressors, \( F^0(\cdot, \cdot) : X \times \Theta \to \mathbb{R}^1 \) is some regression function specified by a parameter \( \theta^0 \in \Theta \subseteq \mathbb{R}^m, \) \( \{u_t\}_{t=1}^n \) are i.i.d. Gaussian random variables, \( \mathcal{L}\{u_t\} = \mathcal{N}(0, (\sigma^0)^2) \). Consider a set of \( K \) nonintersecting intervals:

\[ A_1 = (a_0, a_1], \quad A_2 = (a_1, a_2], \ldots, A_{K-1} = (a_{K-2}, a_{K-1}], \quad A_K = (a_{K-1}, a_K], \quad -\infty = a_0 < a_1 < \cdots < a_K = +\infty. \]

We do not observe true values of \( \{Y_t\}_{t=1}^n \), instead we observe discrete values \( \{\nu_t\}_{t=1}^n \), where \( \nu_t = i, \) if \( Y_t \in A_i, \) \( i \in \{1, 2, \ldots, K\} \).

Under this grouping distortion of the dependent variable we present the following results:

- conditions of model identifiability;
- asymptotic properties of MLE;
- statistical tests on regression function;
- results of computer experiments.

Keywords. Incomplete data; Grouping distortion; Multiple regression.

References

A Robust Tucker3 Model for Compositional Data

V. Todorov1,∗, M. A. Di Palma2 and M. Gallo2

1 United Nations Industrial Development Organization, Vienna, Austria; v.todorov@unido.org
2 University of Naples-L’Orientale, Naples, Italy; madipalma@unior.it, mgallo@unior.it
∗Corresponding author

Abstract. Double counting is inherent to the output concept, therefore it is preferable to use manufacturing value added (MVA) instead to measure the manufacturing production. While the issue of double counting in production statistics is successfully addressed by using MVA, commodity exchange in trade data is still measured as output. The relevance of value added has increased in the recent years due to the unbundling of the production process, where different stages of value chain take place in different countries. We want to represent the export statistics through value added to output ratio using data from international statistical databases. The data sets considered are organized by country, commodity or activity and year (activities are classified according to the International Standard Industrial Classification of all economic activities (ISIC)) and thus they are three-way compositional data.

Different methods exist for analysis array and we choose Tucker3 because it provides a better subspace approximation for the data. The Tucker3 method as most of the N-way methods is based on alternating least squares (ALS) which makes it vulnerable to the presence of outliers in the data. Even a single outliying data point can strongly influence the resulting model and the conclusions based on it. A robust version of Tucker3 was presented by Pravdova et al. (2001) but it suffers from two main deficiencies. First of all the robust initialization of the algorithm is based on MCD which will not work in high dimensions. And secondly, the method is not suitable for applying on compositional data. We propose to select the initial subset using robust PCA and to transform the compositional data applying ilr transformation (Egozcue et al., 2003; Gallo, 2013). Furthermore, since to our knowledge there is no readily available software for computing robust Tucker3 models, we provide implementation of the proposed algorithm in R. The method is compared to its competitors both in terms of its efficiency and the computational effort needed.

Keywords. Multi-way; Robust; Outliers; Compositional data; Manufacturing value added

References


Robust Integrative Analysis of Multi-block Contaminated Dataset

Valeriy Korostyshevskiy¹,*, Junrui Di¹ and Amelia Kelly¹

¹ Department of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University; vrk@georgetown.edu, jd1470@georgetown.edu, ark77@georgetown.edu
*Corresponding author

Abstract. Research in several fields now requires the analysis of “multi-block” data, in which multiple high-dimensional and fundamentally disparate datatypes are available for a common set of objects. Various techniques have been developed to analyze multi-block data. In 2011, Lock et al. developed the “Joint and Individual variation Explained” (JIVE) to investigate the association between disparate datatypes. This method decomposes a multi-block dataset into a sum of three terms: a low rank approximation corresponding to joint structure, low rank approximations corresponding to individual structures, and residual noise under certain rank and orthogonality constraints. Given that real-life multi-block data (especially those with high dimension) are likely to have outliers, in our paper, we propose to develop a robust version of JIVE considering the use of several different robust regression techniques. We also compare the performance of these robust techniques in terms of approximation accuracy and computational feasibility. Data sets resulting from independent biological samples are often large and highly variable, inciting a need to reduce dimension. We consider several effective rank estimation techniques to identify directions containing the greatest amount of variability in order to appropriately reduce data dimension. These methods rely upon Principal Component Analysis (PCA) and Singular Value Decomposition (SVD). We test and compare seven methods using a Monte Carlo procedure on simulated and real data. We further couple the methods with a robust algorithm to investigate outliers and contamination.

Keywords. JIVE; SVD; PCA; Matrix Factorization; Robust Regression.

References

Covariate-Adjusted Confidence Intervals for Relative Treatment Effects

G. Luta¹,*, A. Barbo² and A. Dragomir¹

¹ Georgetown University; gl77@georgetown.edu, ad576@georgetown.edu
² University of Texas MD Anderson Cancer Center; agbarbo@mdanderson.org
* Corresponding author

Abstract. A nonparametric approach is proposed for inference regarding the relative treatment effect (e.g. the probability that a participant from one treatment group has a higher response value than a participant from the other treatment group) with adjustment for baseline covariates. The new covariate-adjusted confidence intervals are constructed using jackknife empirical likelihood methods. The empirical coverage of the new confidence intervals is compared with that of confidence intervals constructed using alternative methods through a simulation study. The use of the new method is illustrated using data from a randomized clinical trial.

Keywords. Empirical likelihood; Jackknife; Relative effect.
Robust Sparse Canonical Correlation Analysis

C. Croux$^1$ and I. Wilms$^{1,*}$

$^1$ KU Leuven; christophe.croux@kuleuven.be; ines.wilms@kuleuven.be
$^*$ Corresponding author

Abstract. Canonical correlation analysis (CCA) describes the associations between two sets of variables by maximizing the correlation between linear combinations of the variables in each data set. This talk discusses a method for Robust Sparse CCA. Sparse estimation produces linear combinations of only a subset of variables from each data set (Witten et al., 2009). More precisely, some of the elements of the canonical vectors will be estimated as exactly zero. As such, the interpretability of the canonical variates is increased. We also robustify the method such that it can cope with outliers in the data (Branco et al., 2005). To estimate the canonical vectors, we convert the CCA problem into an alternating regression framework. Sparse canonical vectors are obtained by adding a lasso penalty on the coefficient estimates to the Least Squares estimator. The lasso, however, is not robust to outliers. The method can be easily robustified by using the sparse Least Trimmed Squares estimator (Alfons et al., 2013). We illustrate the good performance of the Robust Sparse CCA method in several simulation studies.

Keywords. Canonical Correlation Analysis; Robust regression; Sparsity

References


Sparse S- and MM-regression

V. Öllerer¹,∗ and C. Croux¹

¹ KU Leuven, Belgium; viktoria.ollerer@kuleuven.be, christophe.croux@kuleuven.be
∗Corresponding author

Abstract. The S- and MM-estimators are two of the most commonly used robust regression estimators. Their appealing properties include a high breakdown point and a high efficiency. However, they cannot be applied to high-dimensional data, data with more variables than observations. Adding an $L_1$-penalty to their objective function yields the sparse S- and sparse MM-estimator. These estimators combine the robust properties of the original estimators with model selection induced by the sparsity constraint. In this talk, we present the breakdown point and the influence function of the sparse S and sparse MM-estimators. In a simulation study, we compare their performance to the ordinary S- and MM-estimators, as well as to other sparse estimators like lasso and sparse LTS. Additionally, we apply the estimators to a chemometrics data set with dimension ranging in the thousands that is known to contain several outliers.

Keywords. High-dimensions; Regression; Robustness; Sparsity
Sparse partial robust M regression

S. Serneels¹, P. Filzmoser²*, I. Hoffmann² and C. Croux³

¹ BASF Corp.; sven.serneels@basf.com
² Vienna University of Technology; P.Filzmoser@tuwien.ac.at, ica_hoffmann@yahoo.de
³ KU Leuven; christophe.croux@econ.kuleuven.be

Abstract. Sparse partial robust M regression is introduced as a new regression method. It is the first dimension reduction and regression algorithm that yields estimates with a partial least squares alike interpretability that are both sparse and robust with respect to vertical outliers and leverage points. Comparisons with the classical counterpart Sparse Partial Least Squares (SPLS) regression (Chun & Keles, 2010), as well as to the non-sparse counterparts are made (Serneels et al., 2005).

Keywords. partial least squares regression; sparse regression; high-dimensional data.

References


A Comparison of Robust Empirical Likelihood-based ANOVA Methods

M. Velina¹*, J. Valeinis¹, R. Nedovis¹ and G. Luta²

¹ University of Latvia, Riga, Latvia; mara.velina@lu.lv, janis.valeinis@lu.lv; roberts.nedoviss@inbox.lv
² Georgetown University, Washington D.C., USA; g777@georgetown.edu
* Corresponding author

Abstract. Bondell & Stefanski (2013) proposed linear regression estimators based on a two-stage generalized empirical likelihood method that have good efficiency properties while being robust to outliers. We apply the Bondell and Stefanski method to the situation involving a categorical predictor and obtain a robust ANOVA version of the method. We use the new method on simulated and real datasets (involving heterogenous variances and/or non-normal data) and compare its results with other robust ANOVA methods including the empirical likelihood-based ANOVA procedure for trimmed means proposed by Velina, Valeinis & Luta (2014).

Keywords. Empirical likelihood; Generalized empirical likelihood; ANOVA

References


Least angle regression with categorical variables

S. Van Aelst¹,*, R. Zamar² and H. Zhang²

¹ Department of Mathematics, KU Leuven, Belgium; Stefan.VanAelst@kuleuven.be
² Department of Statistics; ruben@stat.ubc.ca, fred@metaforsoftware.com
* Corresponding author

Abstract. Least angle regression (LARS) has been introduced as a very efficient manner for variable selection in linear regression based on least squares. LARS is closely related to LASSO and provides a list of predictors ordered in terms of importance. The non-robustness of LARS has been addressed by Khan et al. (2007). They showed that LARS only depends on correlations and introduced a fast and robust correlation measure to obtain robust LARS. Another problem of LARS is that it cannot handle well categorical variables. To overcome this problem we introduce new methods to measure correlation between a categorical and numerical variable as well as between two categorical variables. We show that these new correlation measures can be used successfully in LARS and robust LARS which largely extends the applicability of this method. The results show good behavior of the extended LARS method compared to forward selection and group LASSO. Moreover, with these measures of correlation (robust) LARS can in principle even be used to handle problems with categorical response variables.

Keywords. LARS; categorical variables; correlation measures.

References

A Minimum Distance Weighted Likelihood Method of Estimation

K. Arun Kumar

* Indian Statistical Institute, Kolkata; karun3kumar@gmail.com.

Abstract. Minimum disparity estimation is a popular parametric estimation technique which combines strong robustness properties with full first order efficiency, and hence it emerges as useful alternative to the standard maximum likelihood estimation method. See, for example, Lindsay (1994) and Basu et al. (2011). In continuous models, however, the evaluation of the minimum disparity estimator involves a complicated numerical exercise. The computation of the estimating function requires numerical evaluation of integrals at every stage of the iteration which can lead to many coding and convergence problems particularly when the support is infinite and/or the data are multivariate. In this paper, we propose a novel method of minimum disparity estimation which bypasses the integral calculations and prove the asymptotic efficiency of the estimator thus obtained. In spirit the method of estimation has similarity with the weighted likelihood estimation method of Markatou et al. (1998) in that our method also reduces the estimating function to a sum over the observed data points rather than an integral over the entire support. At the end our method also stands as a particular weighted likelihood estimation method. However, our method is fundamentally different from that of Markatou et al. (1998) in that it corresponds to the minimization of a well defined objective function rather than originating from the estimating equation itself.

Keywords. Disparity; Residual Adjustment Function; Kernel Density Estimator; Efficiency.

References


Robust and Efficient Stochastic Optimization via Disparities

Anand N. Vidyashankar\textsuperscript{1,*}, and Jie Xu\textsuperscript{2}

\textsuperscript{1} Department of Statistics, Volgenau School of Engineering, George Mason University; avidyash@gmu.edu
\textsuperscript{2} Department of Systems Engineering and Operations Research, Volgenau School of Engineering, George Mason University; jxu13@gmu.edu
\textsuperscript{*} Corresponding author

Abstract. Stochastic optimization problems arise in a variety of scientific applications. Typically, they involve optimizing the expectation of a random function (with a stochastic input that is typically not i.i.d.) which does not have a closed form expression but can be simulated given an input sequence. These are sometimes referred to as black-box simulations and several examples abound in areas such as queueing networks, airline scheduling, revenue management, and semiconductor manufacturing. Questions of significant interest concern evaluating the input uncertainty and simulation variability in downstream decision making. In this presentation, we describe a disparity-based statistical framework for robust and efficient stochastic optimization and establish several fundamental results concerning the estimated optimizer. Specifically, we introduce an asymptotic regime involving the size of the input and the number of simulations and provide weak regularity conditions for the consistency of the estimated optimizer. In such situations, we also establish the limit distribution of the estimated optimizer under appropriate centering and scaling and identify the limit to be non-Gaussian. Additionally, we describe a novel nested simulation algorithm to estimate the variability of the estimated optimizer and provide methods to construct confidence interval. Finally, we provide theoretical and computational methods for assessing the robustness of the proposed methods. Comparisons with the work of Delage (2010) and data analyses examples will also be presented.

Keywords. Stochastic Optimization; Hellinger distance; COMPASS; Learning; Weak dependence

References

Robustness of phi-divergence test statistics based on exponentially tilted empirical likelihood estimators

A. Felipe¹, N. Martín²,*, P. Miranda¹ and L. Pardo¹

¹ Department of Statistics and Operations Research I, Complutense University of Madrid, 28040 Madrid, Spain; felipe@mat.ucm.es, pmiranda@mat.ucm.es, lpardo@mat.ucm.es
² Department of Statistics, Carlos III University of Madrid, 28903 Getafe (Madrid), Spain; nirian.martin@uc3m.es
*Corresponding author

Abstract. It is well-known that empirical likelihood (EL) estimator exhibit desirable higher-order asymptotic properties when the model is correctly specified. In the presence of model misspecification the EL does not have the same desirable properties. Schennach (2007) introduced the exponentially tilted empirical likelihood (ETEL) that combines the EL and the exponential tilting (ET) estimator. This estimator has a good behavior under model misspecification, preserving the desirable higher-order asymptotic properties of the EL under the correct model. On the other hand, in Balakrishnan et al. (2014) it was established that the empirical phi-divergence test statistics have a better behavior in the sense of efficiency and robustness in comparison with the empirical likelihood ratio test statistic. The main purpose of this communication is to study the robustness properties of the phi-divergence test statistics based on ETEL estimators.

Keywords. Empirical likelihood estimator; Exponential tilting estimator; Exponentially tilted empirical likelihood; Empirical likelihood ratio test; Phi-divergence test statistics; Robustness.

References


Abstract. This talk has two aims: First, we introduce taxicab correspondence analysis (TCA) as a robust L1 variant of correspondence analysis (CA), and we describe the many aspects in which the robustness of TCA appear. Second, it is well known that, the chi-square distance used in CA has an often cited and discussed problem: Rare observations may have an unduly large influence on the analysis, see in particular, Rao (1996) and Greenacre (2013). We shall highlight another facet of CA: For some datasets, CA emphasizes some aspects of the data not necessarily related to rare observations. This will be done by comparing the maps obtained by CA with the maps obtained by TCA on many real data sets. If the projections of view of both maps are quite different, we refer to this phenomenon as parallax. In astronomy, parallax means the apparent change in the position of an object as seen from two different points. The existence of the parallax highlights the nonrobust role of the Euclidean geometry coupled with the chi-square distance in CA.

Keywords. Taxicab correspondence analysis; Correspondence analysis; Influential points; Geometry; Parallax.

References


Robustness and efficiency of multivariate coefficients of variation

S. Aerts\textsuperscript{1,*}, G. Haesbroeck\textsuperscript{2} and C. Ruwet\textsuperscript{2,3}

\textsuperscript{1} HEC-ULg, University of Liège, Belgium; stephanie.aerts@ulg.ac.be,
\textsuperscript{2} Department of Mathematics, University of Liège, Belgium; g.haesbroeck@ulg.ac.be, c.ruwet@ulg.ac.be
\textsuperscript{3} Service de mathématiques, Haute Ecole de la Province de Liège, Belgium; christel.ruwet@hepl.be
\textsuperscript{*}Corresponding author

Abstract. The coefficient of variation is a well-known measure used in many fields to compare the variability of univariate variables having really different means or expressed in different scales. However, when the dimension of the problem is greater than one, comparing the variability only marginally may lead to controversial results. Several multivariate extensions of the univariate coefficient of variation have been introduced in the literature in order to summarize global relative variability in one single index (see \textsuperscript{?}, for a review). These multivariate coefficients are defined in terms of the covariance matrix (via its trace, determinant,...) and of the mean vector of the underlying distribution.

In practice, these coefficients can be estimated by plugging any pair of location and covariance estimators in their definitions. However, as soon as the classical mean and covariance matrix are under consideration, the influence functions are unbounded, while the use of any robust estimators yields bounded influence functions.

While useful in their own right, the influence functions of the multivariate coefficients of variation will be further exploited in this talk to derive a general expression for the corresponding asymptotic variances under elliptical symmetry. Simulations compare the finite-sample efficiency of the classical estimator and the robust approach based on the Minimum Covariance Determinant estimator.

Then, focusing on two of the considered multivariate coefficients, a diagnostic tool based on their influence functions, as suggested in \textsuperscript{?} is derived. It allows to detect those observations having the tendency to increase or decrease the relative dispersion and it will be compared, on a real-life dataset, with the usual DD-plot.

Keywords. Coefficient of Variation ; Influence Function ; Minimum Covariance Determinant Estimator

References


Robust $M$-estimates of the correlation coefficient in bivariate independent component distribution models

G. Shevlyakov* and P. Smirnov

Department of Applied Mathematics, St. Petersburg State Polytechnic University, Polytechnicheskaya, 29, St. Petersburg, 195251, RUSSIA; Georgy.Shevlyakov@phmf.spbstu.ru, s.paul@mail.ru
*Corresponding author

Abstract. The family of bivariate independent component distributions (BICD) firstly introduced in (Shevlyakov & Vilchevski., 2002) proved to be rather advantageous for construction of highly robust and efficient estimates of a correlation coefficient (Shevlyakov et al., 2012). Basing on the maximum likelihood estimates of the correlation coefficient for BICD models, we introduce a family of $M$-estimates of the sought parameter. Their asymptotic normality, variance and bias are obtained. Monte Carlo experiment confirms high robustness and efficiency of the proposed $M$-estimates on small and large samples.

Keywords. Robustness; $M$-estimate; Correlation Coefficient.

References


Robustness of sequential decision making on parameters of the model under distortions

A. Kharin1,*

1 Belarusian State University, Independence av. 4, Minsk 220030, Belarus; KharinAY@bsu.by
*Corresponding author

Abstract. Sequential analysis is an effective approach in the decision making problems (Mukhopadhyay & de Silva, 2009) where the number of observations used is an important characteristic provided the accuracy of decisions is satisfied. The optimal properties of sequential tests are often broken in practice as the data deviate from the hypothetical model, the model is distorted (Huber & Ronchetti, 2009). As a result, the characteristics of sequential decision rules (error probabilities and expected sample sizes) increase significantly w.r.t. the hypothetical values (Kharin, 2014).

The theory is developed and is presented in the talk to analyze robustness of sequential tests for cases of simple and composite hypotheses. The following models of data are considered: independent observations, Markov chains, time series with a trend, autoregressive time series. Distortions of the following types are analyzed: “contamination” of the prior probability distribution of parameters, “outliers” in observations, functional distortions of the probability density functions in the $L_1$- and $L_2$-metrics.

Using the method of asymptotic expansions construction (Kharin & Shlyk, 2009) for the characteristics of sequential tests w.r.t. the distortion level (Kharin, 2013) the instability coefficients are obtained to analyze robustness quantitatively. Families of generalized sequential tests are proposed to increase the robustness, and the robust sequential decision rules are constructed to minimize the maximal possible risk under distortions. The results are illustrated by computer experiments.

Keywords. Sequential decision rule; Distortion; Asymptotic expansion; Robustness; Instability coefficient.

References


Goodness-of-fit tests based on sup-functionals of weighted empirical processes

N. Stepanova¹∗ and T. Pavlenko²

¹ School of Mathematics and Statistics, Carleton University; nstep@math.carleton.ca
² Department of Mathematics, KTH Royal Institute of Technology; pavlenko@math.kth.se
∗ Corresponding author

Abstract. A large class of goodness-of-fit test statistics based on sup-functionals of weighted empirical processes is studied. The weight functions employed are Erdős-Feller-Kolmogorov-Petrovski upper-class functions of a Brownian bridge. The whole class of test statistics is shown to be consistent against a fixed alternative. Based on the result of Csörgő et al. (1986) obtained for this type of test statistics, we provide the asymptotic null distribution theory for the class of tests at hand, and present an algorithm for tabulating the limit distribution functions under the null hypothesis. A new family of nonparametric confidence bands is constructed for the true distribution function and it is found to perform very well. The results obtained, together with a new result on the convergence in distribution of the higher criticism statistic introduced in Donoho & Jin (2004) demonstrate the advantage of our approach over a common approach that utilizes a family of regularly varying weight functions. Furthermore, we show that, in various subtle problems of detecting sparse heterogeneous mixtures, the proposed test statistics achieve the detection boundary found by Ingster (1997) and, when distinguishing between the null and alternative hypotheses, perform optimally adaptively to unknown sparsity and size of the non-null effects.

Keywords. Goodness-of-fit; Weighted empirical processes; Multiple comparisons; Sparse normal means; Confidence bands.

References


Data depth for autoregression

Christine H. Müller$^{1,*}$

$^1$ TU University Dortmund; cmueller@statistik.tu-dortmund.de
$^*$Corresponding author

Abstract. We show that the simplicial depth of an AR(1) process reduces to a simple sign statistic if the AR process is almost surely strictly increasing, i.e. is an explosive process. Then it is easy to see that the asymptotic distribution of the simplicial depth of the AR(1) process is the same as for linear regression through the origin if only one parameter is unknown. However, this is not the case, if more than one parameter are unknown. If two parameters are unknown, then the asymptotic distribution is given by an integrated two dimensional Gauss process for the AR process while the asymptotic distribution is an infinite sum of chi-squared random variables for linear regression with intercept. To avoid complicated asymptotic distributions, we propose modifications of the simplicial depth of a p-dimensional parameter with p greater than one so that the asymptotic distribution is the standard normal distribution.

Keywords. Simplicial depth, Asymptotic distribution, AR(1) process.
Depth-based classification of multivariate functional data

P. Segaert\textsuperscript{1,\ast}, M. Hubert\textsuperscript{1} and P. Rousseeuw\textsuperscript{1}

\textsuperscript{1} Affiliation of authors; Department of Mathematics, KU Leuven, Belgium
pieter.segaert@wis.kuleuven.be, mia.hubert@wis.kuleuven.be, peter@rousseeuw.net
\ast Corresponding author

Abstract. Recently, multivariate functional halfspace depth (MFHD) has been proposed as a new tool to study multivariate functional data. This depth function allows to estimate the central tendency and the variability of multiple sets of curves. The application of this depth function to supervised classification is presented. We use the MFHD to propose a new distance measure, from which we derive a new classifier. Comparisons are made with other classification methods, such as the functional k-nearest neighbour, the maximum depth classifier, and a method based on the depth-depth plot. Their behaviour is also studied in the presence of outlying curves. We illustrate their performance by means of a thorough simulation study and the analysis of real data.

Keywords. Functional data; Supervised classification; Multivariate functional depth; Outliers

References

Comparing $k$ Independent Groups With a Method Based on Trimmed Means

A.Fırat Özdemir¹*, Rand Wilcox²

¹Department of Statistics, Dokuz Eylül University, Turkey; firat.ozdemir@deu.edu.tr
²Department of Psychology, University of Southern California, USA; rwilcox@usc.edu
*Corresponding Author

Abstract. The ANOVA F test is the most popular and commonly used procedure for comparing $k$ independent groups. But it is well known that this method is very sensitive to non-normality, which has led to the derivation of alternative techniques based on robust estimators. In this work, ANOVA F test, Welch test with trimmed mean, Welch test with trimmed mean and a bootstrap-t, Schrader and Hettmansperger method with trimmed means, a percentile bootstrap method with trimmed means and a newly proposed method were compared in terms of both the Type I error probability and power. The proposed method coped well with ANOVA F and other alternatives under various experimental settings.

Keywords. $k$ independent groups; Trimmed mean; Bootstrap-t; Percentile bootstrap

References


Modelling international trade data with the Tweedie distribution

Lucio Barabesi 1, Andrea Cerasa 2∗, Domenico Perrotta 2, Andrea Cerioli 3

1 University of Siena, Siena, Italy; lucio.barabesi@unisi.it
2 European Commission, Joint Research Centre, Ispra, Italy; andrea.cerasa@jrc.ec.europa.eu∗, domenico.perrotta@ec.europa.eu
3 University of Parma; andrea.cerioli@unipr.it

Abstract.
Robust methods are indispensable in the analysis of international trade data, as outliers and deviations from the classical model assumptions are recurring issues. The objective evaluation of such methods requires a large number of data sets artificially generated with known statistical properties, which must comply with real world scenarios.
This contribution shows the potential of the Tweedie distribution (Tweedie, 1984) in mimicking the distribution of the quantities of given products imported into the European Union. The availability of this flexible three-parameter model for describing traded quantities can also provide direct support to policy makers, in the form of tools for monitoring the effect of policy measures and for deciding how to react against international trade distortions originating in and outside the European Union.
We see the advantages of adopting the Tweedie model in several data sets which are particularly relevant in the anti-fraud context and which show non-trivial features. Indeed, we typically have to face with markedly skew empirical distributions with heavy tails, a large number of rounding errors in small-scale transactions due to data registration problems, and structural zeros arising because of confidentiality issues related to national regulations. We also provide a systematic outline of the genesis of the Tweedie distribution and we address a number of relevant computational issues, such as the development of efficient algorithms both for parameter estimation and for random variate generation.

Keywords. Keyword1; Keyword2. Include up to five keywords separated by semicolons starting with capital letters.

References
Robust variable selection in joint modeling of location, scale and shape for high dimensional data through trimming

N. Neykov\textsuperscript{1,*}, P. Filzmoser\textsuperscript{2} and P. Neytchev\textsuperscript{1}

\textsuperscript{1} National Institute of Meteorology and Hydrology, Bulgarian Academy of Sciences, 66 Tsarigradsko chaussee, Sofia 1784, Bulgaria; neyko.neykov@meteo.bg, plamen.neytchev@meteo.bg
\textsuperscript{2} Dept. of Statistics & Probability Theory, Vienna University of Technology, Wiedner Hauptstr. 8-10 A-1040 Vienna, Austria; P.Filzmoser@tuwien.ac.at
\textsuperscript{*} Corresponding author

Abstract. Generalized linear and additive models (GAMLSSs) for location, scale and shape were discussed by Rigby & Stasinopoulos (2005) as a class of statistical models for regression problems with univariate response. A key feature of GAMLSSs is that every parameter of the conditional response distribution can be modeled by its own predictor and an associated link function. The GAMLSSs estimation is based on the penalized likelihood whereas the generalized version of AIC and BIC is used for variable selection. In order to avoid the GAMLSSs deficiencies in presence of high dimensional data settings Mayr et al. (2012) developed the gradient boosting algorithm gamboostLSS to handle the computations within the GAMLSSs framework. However, the proposed boosting estimators can be very sensitive to outliers in the data, especially to outliers in the covariates (leverage points). In order to overcome these disadvantages, the usage of the boosting estimators based on trimming are recommended to estimate the unknown parameters in a robust way. The superiority of this approach is illustrated by examples in a simulation study. As a prominent measure of robustness, the finite sample breakdown point of the considered estimators are characterized in this setting using the notion of d-fullness following Neykov et al. (2014).

Keywords. Penalized Maximum Trimmed Likelihood Estimator; Breakdown point; GLMs.

References


Robust GMM estimation and moment selection in dynamic panel data models

P. Čížek 1,∗ and M. Aquaro 2

1 Department of Econometrics & OR, Tilburg University; P.Cizek@tilburguniversity.edu
2 Department of Economics, University of Warwick; M.Aquaro@warwick.ac.uk
∗Corresponding author

Abstract. Shorter and longer panel data are increasingly used hand in hand with wider availability of the panel data. Although random-effect models allow the use of standard linear regression techniques, the fixed-effect models are more complex to deal with, especially when the lagged dependent variable is included. Mostly locally robust methods based on the generalized method of moments (GMM) adjusted to have a bounded influence function were studied up to now such as Ronchetti and Trojani (2001) and Lucas et al. (2007). Globally robust alternatives have been recently proposed and based on the medians of the differenced data, see Dhaene and Zhu (2009) and Aquaro and Cizek (2014). While the median-based estimators exhibits positive breakdown point and finite gross-error sensitivity, their finite-sample performance both in terms of precision and robustness significantly depends on the unknown underlying data generating process and kind of data contamination.

To design a robust estimator performing well irrespective of the sample structure, data generating process, and the type of contamination, we study robust properties of the medians of the first- and higher-order differenced data and combine the corresponding moment conditions by means of a two-step GMM-type procedure. Contrary to the standard GMM, the proposed estimation procedure is based on weights reflecting both the variance and bias of the moment conditions and its asymptotic and robust properties are derived. Additionally, the procedure is enhanced by a robust moment-selection procedure that selects only the moment conditions relevant for a given data generating process.

Keywords. Dynamic panel-data models; Generalized method of moments; Influence function; Robust estimation.

References


Blind source separation for multivariate conditionally heteroscedastic time series

M. Matilainen*, K. Nordhausen and H. Oja

Department of Mathematics and Statistics, University of Turku; markus.matilainen@utu.fi, klaus.nordhausen@utu.fi, hannu.oja@utu.fi
*Corresponding author

Abstract.
The Blind Source Separation (BSS) model for time series assumes that the components of the observed \( p \)-variate time series are linear combinations of \( p \) independent and stationary time series, and the idea is then to try to transform them back to latent time series (sources) (Miettinen et al., 2012). In the so called AMUSE and SOBI solutions (Belouchrani et al., 1997) the aim is to diagonalize cross-autocovariance matrices with one or several lags, correspondingly, while the unmixing matrix estimate in the classical FOBI solution (Cardoso, 1989), diagonalizes the covariance matrix and the (marginal) matrix of fourth moments. AMUSE and SOBI find the independent sources if the sources are ARMA processes, for example, but fail for conditionally heteroscedastic GARCH processes. On the other hand, FOBI seems to find solution for GARCH models but, unfortunately, ignores the serial dependence.

In this talk, we introduce a new generalized FOBI method, gFOBI, which jointly diagonalizes several lagged fourth moment matrices and seems to work nicely both for ARMA processes and for GARCH processes. The properties of the new estimate are discussed and illustrated with a real data example. The performances of the SOBI, FOBI and gFOBI estimates are compared in a simulation study. The possibility to robustify these moment based estimates is also briefly discussed.

Keywords. Independent Component Analysis; FOBI; GARCH; SOBI.

References


Robust Estimation for the Parameters of the Distributed Lag Models Based on Heavy Tailed Skew Distributions

M.N. Çankaya, Y. Tuaç and O. Arslan

1 Ankara University, Faculty of Science, Department of Statistics, 06100 Ankara/Turkey; mn-cankaya@ankara.edu.tr, ytuac@ankara.edu.tr, oarslan@ankara.edu.tr
* Corresponding author

Abstract. This study proposes to estimate the parameters of the distributed lag models based on heavy tailed skew distributions. The performance of the estimators is illustrated with Manto Carlo simulation study. Also, real data examples are provided to show the modeling strength of the distributions used in the study.

Keywords. Distributed Lag Models; Outliers; Heavy tailed skew distributions.

References

Comparing Singular Spectrum Analysis and Artificial Neural Networks in forecasting exchange rates

M. Yarmohammadi

1 Department of Statistics, Payame Noor University, 19395-4697 Tehran, Islamic Republic of Iran; masyar@pnu.ac.ir

Abstract. Exchange rate between two currencies is regarded as the value of one country’s currency in terms of another currency. The exchange rate is used when simply converting one currency to another or trading in the foreign exchange market. Level of imports and exports, the domestic and international markets, inflation and many other economical indexes are dramatically affected by exchange rate. Many attempts have done to model and forecast exchange rates for the different currencies, but their results are not reliable for all periods and locations. The performance of nonlinear models such as Artificial Neural Network with ARIMA and GARCH linear models have compared. (for example see Marzban et al. (2005) and Fahimifard et al. (2009)). In this research we compare two nonparametric time series analysis and forecasting techniques known as Artificial Neural Networks (NN) and Singular Spectrum Analysis (SSA) for forecasting exchange rates. The SSA technique itself is experiencing a rapid growth in popularity with diverse applications in a variety of time series data. On the other hand, Neural Network models have been evaluated for exchange rate forecasting both historically and more recently. The SSA technique has been used as a filtering method recently, Hassani et al. (2010). However, improper selection of the window length and the number of singular values, that the SSA technique is very sensitive to, provide a non desirable filtered series and consequently non accurate forecasting results. The other object of this paper is to find the optimal values of these choices.

Keywords. Exchange rate, Artificial Neural Networks, Singular Spectrum Analysis, Forecasting.

References


Bounded-Influence Robust Estimation of Copulas

Samuel Orso\(^1\),\(^\ast\), Stéphane Guerrier\(^2\) and Maria-Pia Victoria-Feser\(^1\)

\(^1\) Research Center for Statistics & GSEM, University of Geneva, Switzerland; Samuel.Orso@unige.ch, Maria-Pia.VictoriaFeser@unige.ch
\(^2\) Department of Statistics & Applied Probability, University of California, Santa Barbara, USA; guerrier@pstat.ucsb.edu
\(^\ast\) Corresponding author

Abstract. Copula functions are very convenient for modelling multivariate observations. Popular estimation methods are the (two-stage) maximum likelihood and an alternative semi-parametric with empirical cumulative distribution functions of the margins. Unfortunately, they can often be biased whenever relatively small model deviations occur at the marginal and/or copula levels. In this paper we propose two robust estimators that do not share this undesirable feature. Since skewed and heavy tailed parametric marginals are considered in many applications, we also propose a bounded-bias robust estimator for such distributions that is corrected for consistency by means of indirect inference. In a simulation study we show that our robust estimators outperform the conventional approaches.

Keywords. M-estimators; Indirect Inference; Income distribution; Semi-parametric estimation; Two-stage estimation; Influence function; Robust; Gumbel copula; Clayton copula.
Optimally Robust Covariances (Also Covering Weighted Observations)

P. Ruckdeschel

* Fraunhofer ITWM, Kaiserslautern, Germany; peter.ruckdeschel@itwm.fraunhofer.de

Abstract. For weighted data as it arises, e.g., in stratified sampling, in model based clustering, Gaussian mixture models, or Hidden Markov Models, the Minimum covariance determinant (MCD) estimator (Rousseeuw, 1984, 1986) as one of the most important robust alternatives so far is not available.

We close this gap, (re-)implementing the (fast) MCD estimator from Rousseeuw and Van Driessen (1999) in pure R, with additional coverage of weighted data. Moreover the new pure R extends more easily to refined data structures as in Bates and Maechler (2013) than the current FORTRAN code.

The original proposal for a reweighting step to achieve higher efficiency in the ideal model by Croux and Haesbroeck (1999) is modified in two respects (and covers weighted data most easily): (A) We replace the Mahalanobis norm by an alternative equivariant norm induced by the joint likelihood of multivariate location and scale, and (B) we replace hard thresholding by optimally-robust weights, compare Hampel et al. (1986); Rieder et al. (2008). Equivariance ensures computability for arbitrary observation dimensions. Looking at the respective efficiencies achieved in this approach, this gives considerable gains compared to the original reweighting.

The finite sample behaviour of the new procedure is checked in a simulation study.

Keywords. MCD estimator; weighted data; robust optimality; robust covariances; equivariance

References


On the properties of the multivariate Forward Search estimator

A. Cerioli\textsuperscript{1, *}, A. Farcomeni\textsuperscript{2} and M. Riani\textsuperscript{1}

\textsuperscript{1} University of Parma, Via Kennedy 6; 43125 Parma, Italy; andrea.cerioli@unipr.it, mriani@unipr.it
\textsuperscript{2} Sapienza - University of Rome, piazzale Aldo Moro 5; 00186 Roma, Italy; alessio.farcomeni@uniroma1.it
\textsuperscript{*}Corresponding author

Abstract. The Forward Search is a powerful general method for detecting anomalies in structured data, whose diagnostic power has been shown in many statistical contexts (Atkinson et al., 2004; Riani et al, 2009; Mavridis and Moustaki, 2009; Riani et al, 2014\textsuperscript{a}). However, despite the wealth of empirical evidence in favour of the method, only few theoretical properties have been established regarding the resulting estimators. These properties mainly concern the univariate regression framework (Johansen and Nielsen, 2013). We thus focus on the multivariate case and show that the Forward Search estimators of multivariate location and scatter are strongly consistent at the normal model (Cerioli et al, 2014). We also obtain their finite sample breakdown point. Our results put the Forward Search approach for multivariate data on a solid statistical ground, which formally motivates its use in robust applied statistics. Furthermore, they allow us to compare the Forward Search estimators with other well known multivariate high-breakdown techniques.

Keywords. Forward Search; Multivariate estimation; Strong Consistency; Breakdown point.

References


A von Mises approximation to the small sample distribution of
the trimmed mean

A. García-Pérez

Departamento de Estadística, I. O. y C. N., Universidad Nacional de Educación a Distancia (UNED),
Paseo Senda del Rey 9, 28040-Madrid, Spain; agar-per@ccia.uned.es

Abstract. The small sample distribution of the trimmed mean is usually approximated with
a Student’s $t$. But this approximation is valid only in the case that the sample comes from a
standard normal model and the sample size is not very small.
Although there are some accurate saddlepoint approximations when the sample size is small
and the distribution not normal, these are very difficult to apply and the elements involved in
it difficult to interpret.
In this paper we propose a new approximation based on the von Mises expansion of the tail
probability functional of the trimmed mean that improves the usual Student’s $t$ approximation
in the normal case and that is able to be applied for other more general models.
This new approximation allows, for instance, a better choice of the trimming fraction in a
context of hypothesis testing problem using the new tool, P value line.

Keywords. Robustness; Trimmed mean; Small sample distribution; Von Mises expansion; P
value line
Robust state space models for estimating fish stock maturities

Ximing Xu\textsuperscript{1,*}, Eva Cantoni\textsuperscript{2}, Chris Field\textsuperscript{1} and Joanna Mills Flemming\textsuperscript{1}

\textsuperscript{1} Dalhousie University; ximing.xu@dal.ca, chris.field@dal.ca, joanna.flemming@dal.ca
\textsuperscript{2} University of Geneva; Eva.Cantoni@unige.ch
\*Corresponding author

Abstract. In this paper we develop a weighted likelihood method for state space models used in fisheries science that allow errors in the observation process. Standard methods for fitting state space models have been shown to be sensitive to noisy data. By adjusting the threshold parameter, the weighted likelihood allows us to achieve a good balance between efficiency and robustness. Simulation studies show that the proposed method outperforms the maximum likelihood in the presence of measurement errors, and has similar performance when there is no contamination. Finally the weighted likelihood approach is applied to a dataset of cod maturity collected by Fisheries and Oceans Canada.

\textbf{Keywords.} State space model; Weighted likelihood; Robustness; Laplace approximation.
Robust Value-at-Risk Estimation

D.J. Dupuis\textsuperscript{1,*}, N. Papageorgiou\textsuperscript{2} and B. Rémillard\textsuperscript{1}

\textsuperscript{1} Department of Management Sciences, HEC Montréal, Montréal (Québec), Canada; debbie.dupuis@hec.ca, bruno.remillard@hec.ca
\textsuperscript{2} Department of Finance, HEC Montréal, Montréal (Québec), Canada; nicolas.papageorgiou@hec.ca
\* Corresponding author

\textbf{Abstract.} Losses due to idiosyncratic events can have a disproportionate impact on traditional value-at-risk estimates, upwardly biasing these estimates, increasing capital requirements and unnecessarily reducing the available capital and profitability of financial institutions. We propose new bias-robust conditional variance estimators based on weighted likelihood at heavy-tailed models, as well as value-at-risk estimators based on the latter and on volatility updated historical simulation. The new value-at-risk estimators use optimally chosen rolling window length and smoothing parameter value. A simulation study and extensive backtesting results illustrate the strong performance of the proposed methodology during all periods and highlight the model’s ability to mitigate the potentially costly upward bias generated by idiosyncratic shocks.

\textbf{Keywords.} M-estimator; Bias-robust; Exponentially weighted moving average.
Robust probability density forecast of yearly peak load

Yogesh K. Bichpuriya\textsuperscript{1,*}, S. A. Soman\textsuperscript{1} and A. Subramanyam\textsuperscript{2}

\textsuperscript{1}Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India; bichpuriya@ee.iitb.ac.in, soman@ee.iitb.ac.in
\textsuperscript{2}Department of Mathematics, Indian Institute of Technology Bombay, Mumbai, India; as@math.iitb.ac.in

*Corresponding author

Abstract. We consider the problem of long term probability density forecast of yearly peak load (YPL) for an electricity distribution company. The problem is characterized by a sparse sample as only one observation of YPL is available in a year. The problem has primarily two dimensions viz., forecasting trend and modelling explanatory variables like GDP and population. We assume that probability density forecasts of explanatory variables are available (Survey of Professional Forecasters, RBI). This is at variance with earlier approaches e.g., Hyndman and Fan (2010), which treated explanatory variables as categorical variables. The primary applications of such forecasts are in portfolio optimization and network expansion planning. Hence, the forecast should be robust against influential observations, error in econometric projections and discrepancies in estimating trend function.

We achieve robustness against influential observations and error in econometric projections as follows. By using a method akin to jackknifing, we obtain multiple instances of YPL per scenario of explanatory variables. The median of these instances is used as a point estimate for the corresponding scenario. The process is repeated for different instances of explanatory variables resulting in multiple point estimates. The density forecast of YPL is obtained using kernel density estimation.

We use alternating condition expectation (Breiman and Friedman, 1985) to discover trend without making any assumption on its functional form. We also obtain density forecasts using parametric trend models like linear, parabolic and exponential with explanatory variables factored in them. Finally, we combine all the forecasts to achieve robustness against discrepancies in trend models. We present a real life case study to illustrate the approach.

Keywords. Non-parametric methods; Probability density forecast; Yearly peak load.

References


Robust Dimension Reduction

Shojaeddin Chenouri\textsuperscript{1,*}, Jiaxi Liang\textsuperscript{1} and Christopher G. Small\textsuperscript{1}

\textsuperscript{1} University of Waterloo; schenouri@uwaterloo.ca, j34liang@uwaterloo.ca, cgsmall@uwaterloo.ca
\textsuperscript{*} Corresponding author

Abstract. Information in the data often has far fewer degrees of freedom than the number of variables that encode it. Dimensionality reduction attempts to reduce the number of variables used to describe the data. How to assess or compare the performances of different dimension reduction methods is still an open problem. It is known that some commonly used methods are sensitive to the presence of outliers, or the violation of model assumptions. In this presentation, we shall first survey some dimension reduction techniques which are robust against departure from certain assumptions. This survey includes robust approaches to nonlinear dimension reduction under a unifying framework of kernel PCA. By using a kernel trick, the robust methods available for PCA can be extended to nonlinear cases. In this talk we also introduce a goodness measure called local Spearman correlation for assessing the performance of dimensionality reduction methods. Based on the goodness measure, a type of influence function and breakdown point are defined to study the robustness of dimensionality reduction methods.

Keywords. principal component analysis; outlier; robust statistics; kernel; manifold; dimension reduction
Robust backfitting estimators for additive models

G. Boente\textsuperscript{1,}\textsuperscript{*}, A. Martínez\textsuperscript{1} and M. Salibian–Barrera\textsuperscript{2}

\textsuperscript{1} Universidad de Buenos Aires and CONICET; gboente@dm.uba.ar, ale_m_martinez@hotmail.com
\textsuperscript{2} University of British Columbia; matias@stat.ubc.ca
\textsuperscript{*}Corresponding author

Abstract. As is well known, kernel estimators of the regression function in nonparametric multivariate regression models suffer from the so-called curse of dimensionality, which occurs because the number of observations lying in neighbourhoods of fixed radii decreases exponentially with the dimension. Additive models are widely used to avoid the difficulty of estimating regression functions of several covariates without using a parametric model. They generalize linear models, are easily interpretable, and are not affected by the curse of the dimensionality. Different estimation procedures for these methods have been proposed in the literature. It is easy to see that most of these estimators can be unduly affected by a small proportion of atypical observations, since they are based on local averages or local polynomials. In particular, the effect of a response outlier will be large if the related covariates are close to the point in which the regression function needs to be estimated. For that reason, robust procedures to estimate the components of an additive model are needed. To solve this problem, we consider a robust back–fitting algorithm based on local kernel polynomials. These estimators simultaneously avoid the curse of dimensionality and the sensitivity to atypical observations. Fisher–consistency is derived under mild conditions and a simulation simulation is performed to compare our proposal with the usual back–fitting algorithm.

Keywords. Additive Models; Backfitting; Robustness
Regularized M-estimators of scatter matrices

Esa Ollila\textsuperscript{1}, David E. Tyler\textsuperscript{2,*}

\textsuperscript{1} esa.ollila@aalto.fi
\textsuperscript{2} The State University of New Jersey; dtyler@rci.rutgers.edu
\textsuperscript{*} Corresponding author

\textbf{Abstract.} A general class of regularized M-estimators for the scatter matrix of a multivariate complex data set is proposed which is also applicable to low or insufficient sample support (small $n$, large $p$) problems. This class constitutes a natural generalization of M-estimators of the scatter matrix and are defined as a solution to a penalized M-estimation cost function that depend on a pair of regularization parameters.

Using the concept of geodesic convexity, we prove the uniqueness of the regularized M-estimators of scatter and the uniqueness of the solution to the corresponding M-estimating equations. An iterative algorithm with proven convergence to the solution of the regularized M-estimating equation is also given. Furthermore, we derive a simple, closed form and data dependent solution for choosing the regularization parameters based on shape matrix matching in the mean squared sense.

Finally, some simulations studies illustrate the improved accuracy of the proposed regularized M-estimators of scatter compared to their non-regularized counterparts in low sample support problems. An example of radar detection using normalized matched filter (NMF) illustrate that an adaptive NMF detector based on regularized M-estimators are able to maintain accurately the preset CFAR level and at the same time provide similar probability of detection as the (theoretical) NMF detector.

\textbf{Keywords.} Geodesic convexity, Complex elliptical distributions, M-estimator of scatter, Regularization, Robustness, Normalized matched filter.
A simple and practical solution to the classical multivariate Behrens-Fisher problem under non-normality

M. Rauf Ahmad1,∗, Silvelyn Zwanzig2

1 Department of Statistics, Uppsala University, Uppsala, Sweden; rauf.ahmad@statistik.uu.se
2 Department of Mathematics, Uppsala University, Uppsala, Sweden; silvelyn.zwanzig@math.uu.se
∗Corresponding author

Abstract. A solution to the classical multivariate Behrens-Fisher problem is proposed under fairly general conditions. In particular, the solution is robust to the normality assumption. The two-sample case is first discussed to explain the idea in detail, which is then straightforwardly extended to the general case of several populations. Several salient features of the proposed solution distinguish it from the competing solutions in the literature. For example, it focuses on a simple modification to the usual Hotelling’s $T^2$ statistic under heteroscedasticity and shows that it, asymptotically, follows the same Chi-squared distribution as obtained under normality assumption with known covariance matrices. Further, the modified test and its asymptotic distribution are based on the theory of $U$-statistics.

Several useful special cases, particularly concerning the power of test statistic, are considered to show the outperforming properties of the test statistic in comparison to the ones already available in the literature. The rank-based completely non-parametric Behrens-Fisher problem is also considered. The underlying philosophy and approach of the proposed solution closely follows that of Ahmad (2014a,b) where two- and multi-sample test statistics for similar problem are proposed when the data are high-dimensional. Simulation results are used to show the accuracy of the proposed solution and its comparison with its competitors. A few real-life applications are also used to illustrate the practical use of the proposed solution.

Keywords. Behrens-Fisher problem; $U$-statistics; Non-normality;

References


A boxplot for circular data

D. Buttarazzi*, G.C. Porzio

Department of Economics and Law, University of Cassino and Southern Lazio;
davidebuttarazzi@outlook.com, porzio@eco.unicas.it
* Corresponding author

Abstract. It is more than 40 years that statisticians use the boxplot as a quick and effective tool for exploratory data analysis. However, the corresponding tools available for circular data are not fully satisfactory. Both Fisher (1995) and Abuzaid et al. (2011) adopted a ‘linear’ approach that turns out to be not always appropriate. While on the line the structure of the box is uniquely defined once the observations are sorted in non-decreasing order, in the circular space this is not allowed given the lack of a natural ordering. For this reason, a proper boxplot is introduced here. In analogy with the bagplot for bivariate data (Rousseeuw et al., 1999), data depth functions are used to provide a depth-based circular boxplot. By means of a simulation study, a rule of thumb useful to indentify the whiskers is then offered, along with a discussion on the choice of the depth function. A visual display of ‘parallel’ circular boxplots is finally proposed.

Keywords. Data depth; Parallel circular boxplots; Whiskers’ overlap.

References


On RM-Estimators in Regression

S. Zwanzig

Uppsala University, Department of Mathematics, Box 480, 75106 Uppsala Sweden, zwanzig@math.uu.se

Abstract. Rank estimators and M-estimators are defined as a minimizer of a rank dispersion or
of a respective M criteria. Here we assume a combination of both methods. The RM-estimator
is a minimizer of a criteria which includes ranks of residuals as well as weighted generalized
residuals. In case of an errors-in-variables model orthogonal residuals are proposed, compare
Zwanzig (2013). These criteria are L statistics, thus using the theory of L statistics the
consistency of RM-estimators is shown. In a small simulation study the behavior of different
R, M and RM-estimators is studied.

Keywords. Regression; Rank estimators; M-estimators; Errors-in-variables

References

Weighted regression estimator with SCAD

K.-M. Jung

1 Department of Statistics and Computer Science, Kunsan National University, South Korea; kmjung@kunsan.ac.kr
*Corresponding author

Abstract. The least absolute deviation estimator is an alternative to the least squares estimator in the robust point of view. However, the former is sensitive to leverage points, even though it is robust to the regression outliers. We propose a weighted regression estimator which is robust to both leverage points and regression outliers. The estimator is a penalized regression estimator with the weighted absolute deviation loss function and the smoothly clipped absolute deviation penalty (Fan & Li, 2001). Since the considered penalty function is not convex, we consider the local linear approximation algorithm (Zou & Li, 2008). Numerical simulation shows that the proposed estimator is robust for contaminated data.

Keywords. Least absolute deviation; Local linear approximation; Robust; Smooth clipped absolute deviation; Weight.

References


A new Algorithm for Least Trimmed Absolute Deviation (LTAD) Location and Regression

C. Chatzinakos\textsuperscript{1,*}, G. Zioutas\textsuperscript{2}

\textsuperscript{1} Faculty of Electric Engineering Aristotle University of Thessaloniki, Greece; Chatzin@auth.gr
\textsuperscript{2} Faculty of Electric Engineering Aristotle University of Thessaloniki, Greece; Zioutas@eng.auth.gr
\textsuperscript{*} Corresponding author

Abstract. It is well known that classical methods using sample average suffer from the present of outliers. Using the median instead of the mean can particularly resolve this issue but not completely. A better approach is to use the Least Trimmed Absolute Deviation (LTAD) which is known to have desirable properties as has been studied by Tableman (1994). This method users at least half of the sample observations. However, this weighting function can make it difficult to identify intermediate outliers that are relatively close to the bulk of the data.

We propose a new trimming procedure that reformulates the LTAD problem as a mixed integer linear program which is then shown to be equivalent to a linear program (LP) under some transformations. The new robust procedure LP-LTAD can reduce the coverage of the sample observations below 50\%, without losing efficiency which enables us to identifying the intermediate outliers. Further the proposed LP-LTAD algorithm can be consider as an alternative approach to LTAD robust regression studied by Hawkins & Olive (1999).

Simulations results show that LP-LTAD is attractive alternative to other algorithms for LTAD regression, particular for large data samples as its computational load is significant low. Moreover, the linear formula permits downweight of \( x \) outliers and yielded regression estimates are robust against outliers on high leverages cases.

Keywords. Robust location estimation; Least trimmed absolute deviation; Outlier detection; Linear programming; Robust regression estimates

References


58
Redescending psi in GLM (without Mallows weights) and a new robust test for forward search

Mark Hannay

1 University of Geneva; mark.hannay@unige.ch
∗Corresponding author

Abstract. Cantoni and Ronchetti (2001) proposed a now widely used robust score equation in GLM. In order for these score equations to be robust against outliers in x, they proposed to use Mallows weights on the covariates. Although robustness is achieved, depending on the covariates, the relative loss of efficiency isn’t bounded.

We propose a slightly different scores equation, which while using redescending psi achieves the $\frac{1}{2}$ breakdown point. We also propose a new robust test, similar to fitting residuals in the linear case. It has the advantage over the already existing robust test, that it is computationally much faster in a forward search situation, where one has a fixed reduced model and many possible extensions, that all need testing and thus fitting.

Keywords. Redescending psi, Robust GLM, High breakdown point.
Robust estimators of the generalized loggamma regression model with censored observations

C. Agostinelli1∗, I. Locatelli, A. Marazzi and V. Yohai

1 Ca’ Foscari University, Department of Environmental Sciences, Informatics and Statistics; *Corresponding author

Abstract. Robust estimators for accelerated failure time models with censored observations are proposed. It is assumed that the error model belongs to a location-scale-shape family of distributions on the logscale. In particular, the generalized loggamma distribution is considered. A three steps procedures is proposed. In the first step, the procedure computes a regression high breakdown point (hbdp) S-estimate of the slope parameters. Then, it calculates a hbdp Qtau-estimate of the remaining parameters, minimizing a tau-scale of the differences between empirical (Kaplan-Meier) and theoretical quantiles of the residuals. Finally, it computes a weighted likelihood estimator with weights based on a disparity measure between the error model density and a kernel density estimate of the residuals. The final estimate attains full efficiency at the model, with respect to the maximum likelihood estimate, while maintaining the breakdown point of the S and Qtau estimators.

Keywords. Loggamma regression; Censored data
On optimal combination of probability density forecasts using Kullback-Leibler divergence criterion

Yogesh K. Bichpuriya\textsuperscript{1,*}, S. A. Soman\textsuperscript{1} and A. Subramaniam\textsuperscript{2}

\textsuperscript{1}Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India; bichpuriya@ee.iitb.ac.in, soman@ee.iitb.ac.in
\textsuperscript{2}Department of Mathematics, Indian Institute of Technology Bombay, Mumbai, India; as@math.iitb.ac.in
\textsuperscript{*}Corresponding author

Abstract. We consider the problem of forecasting probability density of yearly peak load in the long term. Typically, multiple forecasts are available parametrized by explanatory variables like GDP and population (Hyndman and Fan, 2010). We address the question of combining them in a meaningful way. We consider convex combination of the available probability density forecasts. Kullback-Leibler information criterion (KLIC) is used to measure divergence between two densities (Cover and Thomas, 2006). It has been used to measure divergence between estimated density and true density (Hall and Mitchell, 2007). Since, the true density is unknown, the problem of obtaining optimal weights, i.e., minimization of KLIC, can be translated to that of maximization of Expected log score (ELS). This problem is solved by discretization.

Given \((y_i, f^k_i), i = 1, \cdots, N_p\) and \(k = 1, \cdots, N_f\), i.e., \(N_p\) instances of density forecasts where \(f^k_i\) is probability of instance \(y_i\) by \(k\)th forecast and not all zero in sets of \(\{f^k_i, \forall k\}, \forall i\), we show that maximization of ELS is a convex programming problem. We show that an optimal solution, \(w^\ast\), exists and leads to a global maximum. We, then, consider the problem of determining minimal set of forecasters from a given set. We show that a set of forecasters is minimal iff \(\begin{bmatrix} G \\ 1^T \end{bmatrix}\) is full column rank where \(G\) is \(N_p \times N_f\) matrix of forecast probabilities and \(1^T = [1, 1, \cdots, 1]\). We show that addition of a new forecast can not degrade quality of combination. We illustrate the approach with a real life data set of an electricity distribution company.

Keywords. Combination; Kullback-Leibler information criterion; Probability density forecast.

References


Nonparametric classification of circular data

Giuseppe Pandolfo$^1, *$, Giovanni C. Porzio$^2$ and Mia Hubert$^3$

$^1$ Department of Economics and Law, University of Cassino and Southern Lazio; g.pandolfo@unicas.it
$^2$ Department of Economics and Law, University of Cassino and Southern Lazio; porzio@eco.unicas.it
$^3$ Department of Mathematics & LStat, KU Leuven; mia.hubert@wis.kuleuven.be
* Corresponding author

Abstract. A procedure is developed in order to deal with the classification problem of objects in circular statistics. It is fully nonparametric and based on depth functions for directional data. Using the so-called DD-plot, we apply the $k$-nearest neighbors algorithm in order to discriminate between competing groups. We adopt three different notions of data depth for directional data: the angular simplicial, the angular Tukey and the arc distance. We investigate and compare their performances through the average misclassification rate with respect to different distributional settings by using simulated and real data sets. Results show that the use of the arc distance depth is to be generally preferred, and in some cases it outperforms the classifier based both on the angular simplicial and Tukey depths.

Keywords. Classification; Circular data; Data Depth; DD-plot; k-Nearest Neighbors.

References


Analysis of Regression Models with AR(1) Error Terms Based on Skew Distributions: Parameter Estimation

Y. Tuac1,* and O. Arslan1

1 Ankara University, Faculty of Science, Department of Statistics, 06100 Ankara/Turkey
ytuac@ankara.edu.tr, oarslan@ankara.edu.tr
*Corresponding author

Abstract. In this study we consider a linear regression model with AR(1) errors. We use heavy tailed and heavy tailed skew distributions to estimate parameters of interests. We give a small simulation study and real data example to illustrate performance of the purposed estimators.

Keywords. Regression; Skew Distributions; Autoregressive; Parameter Estimation; EM Algorithm.

References


Abstract. The Zipf distribution appears very often in practice when modelling nature as well as man-made phenomena. In this fact the Zipf distribution, which is suggested by G.K. Zipf (1949), have attracted particular attention among scientists. The Zipf distribution is a particular case of the discrete Power Law distribution with positive integers support. It has only one parameter. In this study, we propose robust estimators for the parameter of Zipfian distribution. We give some real data examples to illustrate the performance of the proposed estimators over the existing ML and the frequency estimators.

Keywords. Power Law; Robust; Zipf.

References


A compound class of geometric and lifetimes distributions

S. Alkarni¹,*

¹ King Saud University, Saudi Arabia; salkarni@ksu.edu.sa
* Corresponding author

Abstract. A new lifetime class with decreasing failure rate is introduced by compounding truncated Geometric distribution and any proper continuous lifetime distribution. The properties of the proposed class are discussed, including a formal proof of its probability density function, distribution function and explicit algebraic formulae for its reliability and failure rate functions. A simple EM-type algorithm for iteratively computing maximum likelihood estimates is presented. A formal equation for Fisher information matrix is derived in order to obtaining the asymptotic covariance matrix. This new class of distributions generalizes several distributions which have been introduced and studied in the literature.

Keywords. Lifetime distributions; Decreasing failure rate; Geometric distribution.
6 Social Program

- Sunday (August 10), 5pm–9pm, Große Steinstraße 73 (School of Business and Economics)
  Welcome reception/German barbecue

- Monday (August 11), 5.30pm–7pm, Start: Universitätsplatz 9 (Melanchthonianum)
  Guided tour through Halle (by foot)

- Wednesday (August 13), 12am–6pm, Start: Universitätsplatz 9 (Melanchthonianum)
  - Trip to Wittenberg: Wittenberg is a small town (with a population of approx. 50,000) situated some 80km to the north-east of Halle. Wittenberg is about 800 years old and a historically important political and cultural center especially due to close connections with Martin Luther, the protestant reformer. The Castle Church in Wittenberg is the place where in 1517 Martin Luther nailed his 95 theses and where today his tomb can be found. A city tour will cover numerous historical sites in Wittenberg. The trip to Wittenberg and back to Halle will be by bus.
  - Trip to the BMW factory near Leipzig: The BMW factory Leipzig is located some 40km east of Halle. It was opened in 2005 and is one of the most modern car factories in the world. The factory employs about 6,000 workers and has a production capacity of up to 750 cars per day. Beside BMW’s classical vehicle series (1 & 2 series) also electric cars (i series) are manufactured. There will be an extensive tour through the factory. The trip to the factory and back to Halle will be by bus.

- Thursday (August 14), 7pm–12pm, Franckestraße 1 (DORMERO Kongress- & Kulturzentrum)
  The Conference dinner will take place at the DORMERO Congress Center. Participants have to use their own means of transport (walking or public transport) to get to the location as well as for the return trip.
Explore the city under your own steam. Listed below, you will find a selection of recommended restaurants, bars, clubs and beer gardens.

RESTAURANTS
1 Zum Schad Kleine Klausstraße 3 German, reasonable prices
2 Trattoria Da Luca Kleine Ulrichstraße 26 Italian, price: mid-range
3 Restaurant Hermes Marktplatz 15 Greek, price: mid-range
4 Prager Bierstuben Große Nikolaistraße 9 Czech, price: mid-range
5 Große Sternstraße Große Sternstraße Several restaurants offer different kinds of food, price: mid-range
6 Ökooase Kleine Ulrichstraße 2 Vegetarian restaurant, open until 7.00 p.m.

BARS
7 Enchilada Universitätsring 6 trendy bar, Mexican, wide-range of cocktails
8 Kaffeeschuppen Kleine Ulrichstraße 11 cosy Irish pub, Jazz
9 Lujah Kleine Ulrichstraße 36 cosy, trendy bar
10 NT Cafe Große Ulrichstraße 51 cosy café & bar
11 Potemkin Kleine Ulrichstraße 27 hip cocktail bar
12 Sonar Dachritzstraße 6 cosy cocktail bar
13 Czech Mittelstraße 7 cosy café & cocktail bar

CLUBS
14 Flower Power Moritzburgering 1 every night, mainly 60s and 70s music
15 Tanzbar Palette Große Nikolaistraße 9-11 modern music, predominantly electronic
16 Turm Friedemann-Bach-Platz 5 Tuesday and Wednesday, 3 dance floors in a medieval tower, modern music

BEER GARDENS
17 Objekt 5 Seebener Str. 5 beer garden, German food
18 Volkspark Burgstraße 27 big beer garden
19 Fritzengarten Jägerplatz 14 cosy beer garden
The city of Halle has a wide range of restaurants, bars, clubs and beer gardens which offer different kinds of foods and drinks for each taste. Starting from Marktplatz, most of those locations are within walking distance.
8  Support & Sponsoring

Sponsors

- Deutsche Forschungsgemeinschaft (DFG)
- Halloren Schokoladenfabrik AG
- Joint Research Centre (JRC) of the European Commission
- Ministerium für Wissenschaft und Wirtschaft des Landes Sachsen-Anhalt

Exhibitors

- Springer-Verlag GmbH
9 List of Participants

Aerts, Stéphanie, University of Liege
Ageeva, Helena, Belarusian State University
Agostinelli, Claudio, Ca’ Foscari University
Ahmad, M. Rauf, Uppsala University
Alkarni, Said, King Saud University
Arslan, Olcay, Ankara University
Becker, Claudia, Martin-Luther-University
Bichpuriya, Yogesh Kumar, Indian Institute of Technology Bombay
Boente Boente, Graciela L., Universidad de Buenos Aires
Buttarazzi, Davide, University of Cassino and Southern Lazio
Cankaya, Mehmet Niyazi, Ankara University
Celik, Nuri, Bartin University
Cerasa, Andrea, Joint Research Centre
Cerioli, Andrea, University of Parma, Italy
Cetin, Meral, Hacettepe University
Chatzinakos, Chris, Aristotle University of Thessaloniki
Chenouri, Shojaaeddin, University of Waterloo
Choulakian, Jonathan, Université de Moncton
Choulakian, Vartan Ohanes, Université de Moncton
Cizek, Pavel, Tilburg University
Cömert, Songül, Turkish Statistical Institute
Croux, Christophe, KU Leuven
Dogru, Fatma Zehra, Ankara University
Dombrowski, Kristin, Martin-Luther-University
Dragomir, Anca, Georgetown University
Dupuis, Debbie J., HEC Montreal
Field, Christopher Arnell, Dalhousie University
Filzmoser, Peter, Vienna University of Technology
García-Pérez, Alfonso, UNED
Ghosh, Abhik, Indian Statistical Institute
Güney, Yesim, Ankara University
Haesbroeck, Gentiane, University of Liège
Hannay, Mark, University of Geneva
He, Xuming, University of Michigan
Hubert, Mía, KU Leuven
Jung, Kang-Mo, Kunsan National University
Jung, Keonwoo, Dongwon middle school
Jureckova, Jana, Charles University in Prague
Kapusta, Alexander, Belarusian State University
Kelly, Amelia, LMI
Kharin, Alexey, Belarusian State University
Kirschstein, Thomas, Martin-Luther-University
Kloss, Mathias, Leibniz-Institut für Agrarentwicklung in Transformationsökonomien
Kuchibhotla, Arun Kumar, Indian Statistical Institute
Leßmann, Sophia, Martin-Luther-University
Liebscher, Steffen, Martin-Luther-University
Löscher, Anne, Martin-Luther-University
Luta, George, Georgetown University
Martin, Nirian, Carlos III University of Madrid
Matilainen, Markus Asser Naftali, University of Turku
Meric, Asiyan, Energy Market Regulatory Authority
Müller, Christine, TU Dortmund
Neykov, Neyko Mateev, Bulgarian Academy of Sciences
Oja, Hannu, University of Turku
Öllerer, Viktoria, KU Leuven
Orso, Samuel, University of Geneva
Ozdemir, A. Firtat, Dokuz Eylul University
Pandolfo, Giuseppe, University of Cassino and Southern Lazio
Pardo, Leandro, Complutense University of Madrid
Pohle, Anna, Martin-Luther-University
Porzio, Giovanni C., University of Cassino and Southern Lazio
Ragozini, Giancarlo, University of Naples Federico II
Rosteck, Veronika, Springer-Verlag GmbH
Roussseeuw, Peter, KU Leuven
Ruckdeschel, Peter, Fraunhofer ITWM Kaiserslautern
Schneegolsberg, Lisiane, Martin-Luther-University
Schyns, Michael, University of Liege
Segaert, Pieter, KU Leuven
Shevlyakov, Georgy, St. Petersburg State Polytechnic University
Stahel, Werner A., ETH Zürich
Stepanova, Natalia, Carleton University
Todorov, Valentin, United Nations Industrial Development Organization (UNIDO)
Todorova, Bistra
Tuac, Yetkin, Ankara University
Tyler, David E., Rutgers, The State University of New Jersey
Valeinis, Janis, University of Latvia
Van Aelst, Stefan, KU Leuven
Velina, Mara, University of Latvia
Vidyashankar, Anand N., George Mason University
Visek, Jan Amos, Charles University in Prague
Xu, Ximing, Dalhousie University
Yarmohammadi, Masoud, Payame Noor University
Yohai, Victor Jaime, University of Buenos Aires
Zioga, Krystallia, University of Macedonia
Zwanzig, Silvelyn, Uppsala Universitet