



International Conference on Trends and  
Perspectives in Linear Statistical Inference

# Book of Abstracts

**20** LINSTAT  
KÖPING  
NIT **14**



**Linköping University**  
**INSTITUTE OF TECHNOLOGY**

Linköping, Sweden  
August 24-28, 2014



International Conference on Trends and  
Perspectives in Linear Statistical Inference

# Book of Abstracts

**20**LINSTAT  
KÖPING  
LIN**14**



**Linköping University**  
**INSTITUTE OF TECHNOLOGY**

Linköping, Sweden  
August 24-28, 2014

**Edited by**

Martin Singull

*Linköping University*

*Linköping, Sweden*

Dietrich von Rosen

*Swedish University of Agricultural Sciences*

*Uppsala, Sweden*

and

*Linköping University*

*Linköping, Sweden*

ISBN: 978-91-7519-242-0

Copyright © 2014 Martin Singull

# Contents

<b>1</b>	<b>Program</b>	<b>15</b>
<b>2</b>	<b>Abstract - Invited Speakers</b>	<b>23</b>
	Alan Agresti – <i>Some Perspectives about Generalized Linear Modeling</i> . . . . .	23
	S. Ejaz Ahmed – <i>Big Data, Big Bias, Small Surprise!</i> . . . . .	23
	R. Dennis Cook – <i>Envelopes: Methods for Efficient Estimation in Multivariate Statistics</i> . . . . .	24
	Sat Gupta – <i>Ratio and Regression Estimation of Finite Population Mean Using Optional Randomized Response Models</i> . . . . .	25
	Lynn Roy LaMotte – <i>Factor-effect formulations in linear models such that ANOVA effects can be tested with deleted-terms sums of squares</i> . . . . .	25
	Emmanuel Lesaffre – <i>Joint modeling of a multilevel factor analytic model and a multilevel covariance regression model</i> . . . . .	25
	Yonghui Liu – <i>Influence diagnostics in a vector autoregressive model</i> . . . . .	26
	Thomas Mathew – <i>Small Sample Asymptotics: Two Applications</i> . . . . .	26
	Jamal Najim – <i>Spiked Models in Large Random Matrices and two statistical applications</i> . . . . .	27
	Muni S. Srivastava and Martin Singull – <i>Test for the mean in a Growth Curve model for high dimensions</i> . . . . .	28
	Yongge Tian – <i>Formulas, equalities and inequalities for covariance matrices of estimations under a general linear model</i> . . . . .	29
	Julia Volaufova, Lynn R. LaMotte, and Ondrej Blaha – <i>Generalized <math>R^2</math> in Linear Mixed Models</i> . . . . .	29
	Alena Bachratá and Radoslav Harman – <i>Experimental Designs under Resource Constraints: Algorithmic Construction</i> . . . . .	30
	Fatma Sevinc Kurnaz and Peter Filzmoser – <i>A Robust Liu Regression Estimator</i> . . . . .	30
	Jolanta Pielaszkiewicz, Dietrich von Rosen and Martin Singull – <i>Free Probability approach to Random Matrices – An alternative Free Moment–Cumulant relation formula</i> . . . . .	31
	Maryna Prus and Rainer Schwabe – <i>Linear Aspects in Random Coefficient Regression Models: Optimal Designs for Different Design Criteria</i> . . . . .	32
<b>3</b>	<b>Abstract - Contributed Speakers</b>	<b>33</b>
	Esra Akdeniz Duran – <i>Efficiency of the Generalized Difference-based Liu Estimator in Partially Linear Models with Correlated Errors</i> . . . . .	33
	Shengqiang Guo and Md. Moudud Alam – <i>Fitting spatial hurdle models using a hierarchical likelihood method</i> . . . . .	34
	Öykum Esra Askin and Deniz Inan – <i>Gamma frailty for the exponential-zero truncated Poisson distribution</i> . . . . .	34

<b>Anthony C. Atkinson</b> – <i>Optimum Designs for Two Treatments with Unequal Variances</i> . . . . .	35
<b>Selahattin Aydogdu</b> and Gulhayat Golbasi Simsek – <i>Bootstrap approach in parallel analysis to determine non-zero canonical correlations</i> . . . . .	36
<b>Ufuk Beyaztas</b> and Aylin Alin – <i>Some Notes on Sufficient Bootstrap</i> . . . . .	37
Kabir Soeny, <b>Barbara Bogacka</b> , Byron Jones and Thomas Bouillon – <i>Optimum Dose Regimen in Clinical Trials</i> . . . . .	37
<b>Camila Caiado</b> – <i>Bayesian uncertainty analysis for tipping points modelled by computer simulators</i> . . . . .	38
<b>Francisco Carvalho</b> , Roman Zmšlony and João T. Mexia – <i>Mean Driven Balance and Uniformly Best Linear Unbiased Estimators</i> . . . . .	38
<b>Deliang Dai</b> – <i>Mahalanobis distances of factor structure data</i> . . . . .	39
<b>Dennis Dobler</b> and Markus Pauly – <i>Resampling Aalen-Johansen Processes: Why Efron’s Bootstrap and Permutation Techniques fail and Multiplier Approaches succeed.</i> . . . . .	39
<b>Sandra Donevska</b> – <i>Variable Selection Stepwise Procedure for Compositional Data</i> . . . . .	40
<b>Hilmar Drygas</b> – <i>Gauss-algorithm and Gram-Schmidt orthogonalization</i> . . . . .	41
<b>Esra Emiroglu</b> and Kadri Ulas Akay – <i>An alternative approach to solve the LAD-LASSO problem</i> . . . . .	41
<b>Birsen Eygi Erdogan</b> – <i>Prediction of Turkish Bank Failures Using Panel Factor Logistic Regression</i> . . . . .	42
<b>Kamila Fačevicoá</b> and Karel Hron – <i>Coordinate representation of compositional tables</i> . . . . .	43
<b>Majbritt Felleki</b> – <i>A Hierarchical Model with a Correlation between the Levels: An Example from Animal Breeding</i> . . . . .	44
<b>Katarzyna Filipiak</b> , Daniel Klein and Anuradha Roy – <i>Hypothesis testing for a separable covariance structure with AR(1) under the two-level multivariate model</i> . . . . .	45
<b>Eva Fišerová</b> – <i>Regression with Compositional Response</i> . . . . .	45
<b>Miguel Fonseca</b> and Anuradha Roy – <i>Inference in k-Exchangeable Multivariate Models</i> . . . . .	46
<b>Johannes Forkman</b> – <i>A simple parametric bootstrap method for testing principal components in normally distributed data</i> . . . . .	47
<b>Micheal Goldstein</b> – <i>Bayes linear uncertainty analysis for complex computer models</i> . . . . .	47
<b>Huseyin Guler</b> and Cigdem Kosar – <i>Nonlinear Inequality Constrained Ridge Regression Estimator</i> . . . . .	48
<b>Chengcheng Hao</b> , Dietrich von Rosen and Tatjana von Rosen – <i>Explicit influence analysis for count data under AB-BA crossover trials</i> . . . . .	49
<b>Stephen Haslett</b> – <i>Small area estimation methods and their links to linear and linear mixed models</i> . . . . .	49
<b>Jan Hauke</b> , Charles R. Johnson and Tadeusz Ostrowski – <i>On some applications of saddle-point matrices</i> . . . . .	50
<b>Thomas Holgersson</b> and Deliang Dai – <i>High-dimensional CLTs for individual Mahalanobis distances</i> . . . . .	50
<b>Klára Hružová</b> , Karel Hron, Peter Filzmoser and Valentin Todorov – <i>Orthogonal regression among parts of compositional data</i> . . . . .	51

<b>Su-Yun Huang</b> , Ting-Li Chen, Hung Hung, I-Ping Tu and Pei-Shien Wu – <i>Statistical Properties for Multilinear Principal Component Analysis</i> . . . . .	52
<b>Shinpei Imori</b> and Dietrich von Rosen – <i>Covariance Components Selection in High-Dimensional Growth Curve Model with Random Coefficients</i> . . . . .	53
<b>Deniz Inan</b> , Öyküm Esra Askin and Ali Hakan Buyuklu – <i>A Comparison of compound Poisson class distributions</i> . . . . .	54
<b>Geoff Jones</b> and Penny Bilton – <i>'Hard' versus 'Soft' Predictions from Unit-level Models for Small Area Estimation of Proportions</i> . . . . .	55
Alexandra Jauhainen, José Sánchez and <b>Rebecka Jörnsten</b> – <i>Network Component Resolution and Cancer Landscapes</i> . . . . .	55
<b>Alžběta Kalivodová</b> , Karel Hron and Peter Filzmoser – <i>Partial least squares discriminant analysis for compositional data: logratio methodology and its application to metabolomics</i> . . . . .	56
<b>Emil Karlsson</b> and Martin Singull – <i>Explicit Estimators for a Banded Covariance Matrix in a Multivariate Normal Distribution</i> . . . . .	57
<b>Daniel Klein</b> , Anuradha Roy, Ricardo Leiva and Ivan Žežula – <i>Multivariate data with block compound symmetry covariance structure</i> . . . . .	58
<b>Petr Koldanov</b> – <i>Statistical inferences for market network analysis</i> . . . . .	58
<b>Sergii Kulishov</b> and Olexandr Iakovenko – <i>Fractal and Antifractal Oxymerons, Moebius Strip Like Transformations of Biomedical Data as Basis for Exploratory Subgroup Analysis</i> . . . . .	59
<b>Fikriye Kurtoglu</b> and M. Revan Özkale – <i>Restricted ridge estimation of generalized linear models</i> . . . . .	60
<b>Ying Li</b> and Dietrich von Rosen – <i>A two-step model for linear prediction with connections to PLS</i> . . . . .	61
<b>Yuli Liang</b> , Dietrich von Rosen and Tatjana von Rosen – <i>Hypothesis testing in multilevel models with block circular covariance structures</i> . . . . .	61
<b>Tsung-I Lin</b> and Wan-Lun Wang – <i>Multivariate skew-normal linear mixed models for multi-outcome longitudinal data</i> . . . . .	62
<b>Måns Magnusson</b> , Leif Jonsson and Mattias Villani – <i>A Supervised Topic Models for Textual Data using a Multivariate probit model</i> . . . . .	63
<b>Augustyn Markiewicz</b> and Barbora Arendacká – <i>On sufficiency of quadratically sufficient statistics in possibly mixed model</i> . . . . .	63
<b>Hiba Nassar</b> , Boualem Djehiche and Astrid Hilbert – <i>A Functional Hodrick Prescott Filter</i> . . . . .	64
<b>Innocent Ngaruye</b> – <i>Small Area Estimation for multivariate repeated measures data</i> . . . . .	64
<b>Joseph Nzabanita</b> – <i>Estimation in the multivariate linear normal models with linearly structured covariance matrices</i> . . . . .	65
<b>Kazuhiro Ozawa</b> , Iwona Mejza, Stanisław Mejza and Shinji Kuriki – <i>Repeated Latin and Youden squares with whole plot and subplot control treatments</i> . . . . .	65
Khurshid Alam and <b>Sudhir Paul</b> – <i>Testing Equality of Scale Parameters of Two Weibull Distributions in the Presence of Unequal Shape Parameters</i> . . . . .	66
<b>Marcin Przystalski</b> and Miguel Fonseca – <i>Hypothesis testing in variance components with constraints</i> . . . . .	67
<b>Simo Puntanen</b> – <i>It's 30 years</i> . . . . .	68
<b>Matias Quiroz</b> , Mattias Villani and Robert Kohn – <i>Speeding Up MCMC by Efficient Data Subsampling</i> . . . . .	68

<b>Martin Ricker</b> , Víctor Peña Ramírez and Dietrich von Rosen – <i>A New Method to Compare Statistical Growth Curves: The PL-GMANOVA Model and its Application</i> . . . . .	69
<b>Paulo Canas Rodrigues</b> , Andreia Forte and Vanda Lourenço – <i>Robust principal component analysis with application to multi-location plant breeding trials</i> . . . . .	70
<b>Anuradha Roy</b> , Ricardo Leiva, Rubén Bageta and Juan Carlos Pina – <i>Model of Fatigue Failure Due to Multiple Cracks using Extended Birnbaum-Saunders Distribution</i> . . . . .	71
<b>Patrik Rydén</b> , Rafael Björk, and Tatjana Pavlenko – <i>Structure learning for improved classification accuracy for high-dimensional omics data</i> . . . . .	72
<b>Lars Rönnegård</b> – <i>Rotating a Linear Mixed Model for Multiple Testing in Genome-wide Association Studies having Repeated Measurements</i> . . . . .	72
<b>José Sánchez</b> , Tatjana Pavlenko and Rebecka Jörnsten – <i>Joint estimation of transcription networks and expression with applications to classification</i> . . . . .	73
<b>Timo Schmid</b> , Ray Chambers, Ralf Münnich and Nikos Tzavidis – <i>Bias correction of robust small area estimators under spatial correlation</i> . . . . .	74
<b>Beste H. Sertdemir</b> and Esin Firuzan – <i>New Block Bootstrap Methods: Sufficient and/or Ordered</i> . . . . .	75
<b>Mahendran Shitan</b> , A. Thavaneswaran and Turaj Vazifedan – <i>First-Order Random Coefficient Autoregressive (RCA(1)) Model: Joint Whittle Estimation and Information</i> . . . . .	75
<b>Natalia Stepanova</b> and Tatjana Pavlenko – <i>A new class of goodness-of-fit tests with applications to the problem of detecting sparse heterogeneous mixtures</i> . . . . .	76
<b>Annika Tillander</b> and Tatjana Pavlenko – <i>Higher criticism for estimating proportion of non-null effect in high-dimensional multiple comparison</i> . . . . .	77
<b>Secil Yalaz Toprak</b> , Müjgan Tez and H. İlhan Tutalar – <i>Semiparametric Regression in the Presence of Measurement Error</i> . . . . .	78
<b>Vicenç Torra</b> – <i>New results on the Choquet integral based distributions</i> . . . . .	78
<b>Ozlem Turkсен</b> and Müjgan Tez – <i>Optimization of Compartment Models by Using Metaheuristic Approaches</i> . . . . .	79
<b>Marju Valge</b> , Tonu Kollo and Dietrich von Rosen – <i>Testing Covariance Structure in High-Dimensional Setup: Comparison of Score Test and Likelihood Ratio Test</i> . . . . .	80
<b>Patrik Waldmann</b> – <i>SIS-BEN: an efficient two-stage approach for high dimensional analysis of correlated data</i> . . . . .	81
<b>Wan-Lun Wang</b> and Tsung-I Lin – <i>Estimation in Multivariate <math>t</math> Nonlinear Mixed-effects Models with Missing Outcomes</i> . . . . .	81
<b>Bertil Wegmann</b> , Anders Eklund, Stephen LaConte and Mattias Villani – <i>Bayesian inference for heteroscedastic Rician time series with applications to fMRI</i> . . . . .	82
<b>Silvelyn Zwanzig</b> – <i>SimSel - a Method for Variablen Selection</i> . . . . .	83
<b>M. Revan Özkale</b> – <i>Diagnostic statistics for binary logistic regression in the presence of multicollinearity</i> . . . . .	83
<b>4 Abstract - Posters</b>	<b>85</b>
<b>Silvie Bělašková</b> and Eva Fišerová – <i>Simulation Study of the Proportional Hazards Model with Delayed Entry</i> . . . . .	85



<b>Bandar Bin-Mohsin</b> and Daniel Lesnic – <i>Numerical reconstruction of the support of a source in an elliptic equation</i> . . . . .	86
<b>Justyna Brzezińska</b> – <i>The comparative log-linear analysis of unemployment in Poland in R</i> . . . . .	87
<b>Martina Chvosteková</b> – <i>Confidence Bands for Percentile line in the Linear Regression Model</i> . . . . .	87
<b>Nesrin Güler</b> and Simo Puntanen – <i>IPM method for the BLUEs of subparameters under multiple-partitioned linear model</i> . . . . .	88
<b>Ting Hsiang Lin</b> and Min-Hsiao Tsai – <i>Model Selection Information Criteria for Inflated Data</i> . . . . .	89
<b>Ramidha Srihera</b> , Alongkorn Chamsranoi, Manutchanok Noiyanon and Pakamas Chaiyawan – <i>A comparison of R and Stata in sample surveys</i> . . . . .	90
<b>Nan-Cheng Su</b> and Arjun K. Gupta – <i>On some sampling distributions for skew normal population</i> . . . . .	91
Pakize Taylan and Ersin Uysal and <b>Müjgan Tez</b> – <i>On the Curvature Measurements of the Nonlinear Errors In Variables Models (NEIVM)</i> . . . . .	91
<b>5 List of Participants</b>	<b>93</b>
<b>Index</b>	<b>101</b>



# Preface

The International Conference on Trends and Perspectives in Linear Statistical Inference (LinStat2014) is held 24-28 August, 2014, in Linköping, Sweden, at Linköping University. This is the follow-up of the 2012 edition held in Bedlewo near Poznań, Poland.

The purpose of the conference is to bring together researchers sharing an interest in a variety of aspects related to linear statistical inference and offer them a possibility to discuss current developments in these subjects. The topics of interest include, but are not limited to:

- Statistical Inference: Theory and Applications
- Multilevel models: Theory and Applications
- Model Selection and Dimension Reduction: Theory and Applications
- Design of Experiments: Theory and Applications
- Optimal Design: Theory and Applications
- Mixed Linear Models: Theory and Applications
- High-Dimensional Statistical Analysis: Theory and Applications
- Categorical Data Analysis: Theory and Applications
- Survey Methodology including Small Area Estimation
- Numerical Methods and Linear Models: Theory and Applications
- Biostatistics: Theory and Applications
- Multivariate Analysis: Theory and Applications
- Bayesian Statistics: Theory and Applications
- Matrix Theory and Linear Models: Theory and Applications

The conference consists of invited speakers sessions, contributed speakers sessions, young scientist speaker session and a poster session. Sessions on special topics are also be organized. Before the conference a mini course on Survey Sampling and Linear Models will take place and is given by professor Stephen Haslett, Massey University, New Zealand and professor Simo Puntanen, University of Tampere, Finland.

The work of young scientists is highly appreciated. The Scientific Committee will award the best presentation and best poster. The awarded will be invited to speak at the next edition of LinStat.

## Proceeding

The conference proceeding will be published in a special volume of the journal *Acta et Commentationes Universitatis Tartuensis de Mathematica* (Reviewed by Mathematical Reviews and Zentralblatt für Mathematik). For details see the website <http://math.ut.ee/acta/>.

All papers submitted must meet the publication standards of mentioned journal and

will be subject to normal refereeing procedure. All submitted papers should have become accepted for presentation at the LinStat2104 meeting. The volume will be co edited by Martin Singull and Dietrich von Rosen.

- Submission interval: May 1 - November 15, 2014.

Papers should be submitted to the Editorial Coordinator Enno Kolk ([enno.kolk@ut.ee](mailto:enno.kolk@ut.ee)) with an indication that it is a LinStat2014 paper.

## Committees of LinStat2014

### Scientific Committee

- Dietrich von Rosen (Sweden) - Chair
- Augustyn Markiewicz (Poland)
- Anthony C. Atkinson (United Kingdom)
- João T. Mexia (Portugal)
- Simo Puntanen (Finland)
- Götz Trenkler (Germany)
- Roman Zmyślony (Poland)
- Müjgan Tez (Turkey)

### Organizing Committee

- Martin Singull (Sweden) - Chair
- Aylin Alin (Turkey)
- Francisco Carvalho (Portugal)
- Elaine Billberg (Sweden)
- Katarzyna Filipiak (Poland)
- Innocent Ngaruye (Sweden)
- Joseph Nzabanita (Sweden)
- Jolanta Pielaszkiewicz (Sweden)

## Previous Conferences in the LinStat series

- Wisła, Poland (1977, 1978, 1980),
- Poznań, Poland (1984),
- Olsztyn, Poland (1988),
- Poznań, Poland - LINSTAT'93,
- Jachranka, Poland (1996),
- Łagów, Poland - STAT'98,
- Bedlewo, Poland - STATLIN'03,
- Bedlewo, Poland - LINSTAT'08 - in celebration of Tadeusz Caliński's 80th birthday
- Tomar, Portugal - LINSTAT'2010
- Bedlewo, Poland - LINSTAT'2012

## Acknowledgments

We would like to thank our sponsors

- Vetenskapsrådet - Swedish Research Council
- Ericsson AB - world-leading provider of telecommunications equipment and services to mobile and fixed network operators
- GRAPES - Swedish Network for Graduate and Postgraduate Education in Statistics
- Svenska Statistikfrämjandet
- Springer

Finally we would like to welcome you all to Linköping University and LinStat2014!

With the hope of good work, many ideas and great progress!

Best wishes Martin Singull and Dietrich von Rosen



# 1 | Program

---

---

Sunday August 24 (13:00-20:00)	
<b>13:00-13:30</b>	<b>Opening (Vallfarten)</b> Helen Dannetun – Rector of Linköping University
<b>13:30-14:25</b>	<b>Invited speaker session (1 – Vallfarten)</b> Chair: Simo Puntanen Dennis Cook <i>Envelopes: Methods for Efficient Estimation in Multivariate Statistics</i>
<b>14:25-14:35</b>	<b>Pause</b>
<b>14:35-15:30</b>	<b>Invited speaker session (2 – Vallfarten)</b> Chair: Thomas Mathew Muni S. Srivastava <i>Test for the mean in a Growth Curve model for high dimensions</i>
<b>15:30-16:10</b>	<b>Coffee break (University Club)</b>
<b>16:10-18:30</b>	<b>Special session (3 – Vallfarten)</b> Design and Analysis of Experiments Chair: Augustyn Markiewicz
16:10-16:50	Anthony Atkinson <i>Optimum Designs for Two Treatments with Unequal Variances</i>
16:50-17:10	Barbara Bogacka <i>Optimum Dose Regimen in Clinical Trials</i>
<b>17:10-17:30</b>	<b>Coffee break (University Club)</b>
17:30-17:50	Augustyn Markiewicz <i>On sufficiency of quadraticly sufficient statistics in possibly mixed model</i>
17:50-18:10	Kazuhiro Ozawa <i>Repeated Latin and Youden squares with whole plot and subplot control treatments</i>
<b>18:15-20:00</b>	<b>Get-together and cash bar (University Club)</b>

---

---

Program

---

Monday August 25 (9:00-14.00)	
<b>9:00-10:30</b>	<b>Invited speaker session (4 – Vallfarten)</b>
	Chair: Stephen Haslett
9:00-9:45	Yongge Tian <i>Formulas, equalities and inequalities for covariance matrices of estimations under a general linear model</i>
9:45-10:30	Alan Agresti <i>Some Perspectives about Generalized Linear Modeling</i>
<b>10:30-11:10</b>	<b>Coffee break (University Club)</b>
<b>11:10-12:30</b>	<b>Special session (5a – Nobel)</b>
	Computation-Intensive Methods in Regression Models – I
	Chair: Mijgan Tez
11:10-11:30	Deniz Inan <i>A Comparison of compound Poisson class distributions</i>
11:30-11:50	Secil Yalaz Toprak <i>Semiparametric Regression in the Presence of Measurement Error</i>
11:50-12:10	Esra Emiroglu <i>An alternative approach to solve the LAD-LASSO problem</i>
12:10-12:30	Öyküm Esra Askin <i>Gamma frailty for the exponential-zero truncated Poisson distribution</i>
<b>11:10-12:30</b>	<b>Special session (5b – Planck)</b>
	The use of Kronecker Product in Statistical Modeling
	Chair: Anuradha Roy
11:10-11:30	Katarzyna Filipiak <i>Hypothesis testing for a separable covariance structure with AR(1) under the two-level multivariate model</i>
11:30-11:50	Wan-Lun Wang <i>Estimation in Multivariate t Nonlinear Mixed-effects Models with Missing Outcomes</i>
11:50-12:10	Tsung-I Lin <i>Multivariate skew-normal linear mixed models for multi-outcome longitudinal data</i>
12:10-12:30	Daniel Klein <i>Multivariate data with block compound symmetry covariance structure</i>
<b>12:30-14:00</b>	<b>Lunch (University Club)</b>



Program

---

Monday August 25 (14:00-19:00)	
<b>14:00-15:30</b>	<b>Invited speaker session (6 – Vallfarten)</b>
	Chair: Sat Gupta
14:00-14:45	S. Ejaz Ahmed <i>Big data, big bias, small surprise!</i>
14:45-15:30	Julia Volaufova <i>Generalized <math>R^2</math> in Linear Mixed Models</i>
<b>15:30-16:10</b>	<b>Coffee break (University Club)</b>
<b>16:10-17:30</b>	<b>Special session (7a – Nobel)</b>
	Computation-Intensive Methods in Regression Models – II
	Chair: Müjgan Tez
16:10-16:30	Birsen Eygi Erdogan <i>Prediction of Turkish Bank Failures using Panel Factor Logistic Regression</i>
16:30-16:50	Özlem Türksen <i>Optimization of Compartment Models by Using Metaheuristic Approaches</i>
16:50-17:10	Esra Akdeniz-Duran <i>Efficiency of the Generalized Difference-based Liu Estimator in Partially Linear Models with Correlated Errors</i>
17:10-17:30	Sellahattin Aydogdu <i>Bootstrap approach in parallel analysis to determine non-zero canonical correlations</i>
<b>16:10-17:30</b>	<b>Session (7b – Planck)</b>
	Chair: Daniel Klein
16:10-16:30	Moudud Alam <i>Fitting spatial hurdle models using a hierarchical likelihood method</i>
16:30-16:50	Deliang Dai <i>Mahalanobis distances of factor structure data</i>
16:50-17:10	Su-Yun Huang <i>Statistical Properties for Multilinear Principal Component Analysis</i>
17:10-17:30	Johannes Forkman <i>A simple parametric bootstrap method for testing principal components in normally distributed data</i>
<b>17:30-18:00</b>	<b>Coffee break (Planck)</b>
<b>18:00-19:00</b>	<b>Session (8a – Nobel)</b>
	Chair: Allan Agresti
18:00-18:20	Kamila Facevicova <i>Coordinate representation of compositional tables</i>
18:20-18:40	Eva Fiserova <i>Regression with Compositional Response</i>
18:40-19:00	Klara Hruzova <i>Orthogonal regression among parts of compositional data</i>
<b>18:00-19:00</b>	<b>Session (8b – Planck)</b>
	Chair: Michael Goldstein
18:00-18:20	Emil Karlsson <i>Explicit Estimators for a Banded Covariance Matrix in a Multivariate Normal Distribution</i>
18:20-18:40	Anuradha Roy <i>Model of Fatigue failure due to Multiple Cracks using Extended Birnbaum-Saunders Distribution</i>
18:40-19:00	Dennis Dobler <i>Resampling Aalen-Johansen Processes: Why Efron's Bootstrap and Permutation Techniques fail and Multiplier Approaches succeed</i>

Program

---

Tuesday August 26 (9:00-14:00)	
<b>9:00-10:30</b>	<b>Invited speaker session (9 – Vallfarten)</b> Chair: Anthony Atkinson
9:00-9:45	Lynn LaMotte <i>Factor-effect formulations such that ANOVA effects can be tested with deleted-terms sums of squares</i>
9:45-10:30	Emmanuel Lesaffre <i>Joint modeling of a multilevel factor analytic model and a multilevel covariance regression model</i>
<hr/>	
<b>10:30-11:10</b>	<b>Coffee break (University Club)</b>
<hr/>	
<b>11:10-12:30</b>	<b>Special session (10a – Nobel)</b> Statistics in Life Sciences Chair: Miguel Fonseca
11:10-11:30	Miguel Fonseca <i>Inference in k-Exchangeable Multivariate Models</i>
11:30-11:50	Paulo Rodrigues <i>Robust singular value decomposition with application to multi-location plant breeding trials</i>
11:50-12:10	Francico Carvalho <i>Mean Driven Balance and Uniformly Best Linear Unbiased Estimators</i>
12:10-12:30	Marcin Przystalski <i>Hypothesis testing in variance components with constraints</i>
<hr/>	
<b>11:00-12:40</b>	<b>Special session (10b – Planck)</b> Statistical Methods for Complex, High-Dimensional Data Chair: Tatjana Pavlenko
11:00-11:20	Natalia Stepanova <i>A new class of goodness-of-fit tests with applications to the problem of detecting sparse heterogeneous mixtures</i>
11:20-11:40	José Sánchez <i>Joint estimation of transcription networks and expression with applications to classification</i>
11:40-12:00	Rebecka Jörnsten <i>Network Component Resolution and Cancer Landscapes</i>
12:00-12:20	Patrik Rydén <i>Structure learning for improved classification accuracy for high-dimensional omics data</i>
12:20-12:40	Annika Tillander <i>Higher criticism for estimating proportion of non-null effect in high-dimensional multiple comparison</i>
<hr/>	
<b>12:30-14:00</b>	<b>Lunch (University Club)</b>

Program

---

Tuesday August 26 (14:00-20:00)	
<b>14:00-15:30</b>	<b>Invited speaker session (11 – Vallfarten)</b> Chair: Jamal Najim
14:00-14:45	Yonghui Liu <i>Influence diagnostics in a vector autoregressive model</i>
14:45-15:30	Sat Gupta <i>Ratio and Regression Estimation of Finite Population Mean Using Optional Randomized Response Models</i>
<hr/>	
<b>15:30-16:10</b>	<b>Coffee break (University Club)</b>
<b>16:10-17:10</b>	<b>Session (12a – Nobel)</b> Chair: Muni S. Srivastava
16:10-16:30	Thomas Holgersson <i>High-dimensional CLTs for individual Mahalanobis distances</i>
16:30-16:50	Shinpei Imori <i>Covariance Components Selection in High-Dimensional Growth Curve Model with Random Coefficients</i>
16:50-17:10	Marju Valge <i>Testing for Covariance Structure in High-Dimensional Setup: Comparison of Score Test and Likelihood Ratio Test</i>
17:10-17:30	Sudhir Paul <i>Testing Equality of Scale Parameters of Two Weibull Distributions in the Presence of Unequal Shape Parameters</i>
<hr/>	
<b>16:10-17:30</b>	<b>Special session (12b – Planck)</b> Bayesian Inference for Complex Problems Chair: Mattias Villani
16:10-16:30	Michael Goldstein <i>Bayes linear uncertainty analysis for complex computer models</i>
16:30-16:50	Camila Caiado <i>Bayesian uncertainty analysis for tipping points modelled by computer simulators</i>
16:50-17:10	Patrik Waldmann <i>SIS-BEN: an efficient two-stage approach for high dimensional analysis of correlated data</i>
17:10-17:30	Måns Magnusson <i>A Supervised Topic Models for Textual Data using a Multivariate probit model</i>
<hr/>	
<b>17:30-18:00</b>	<b>Refreshments (University Club)</b>
<b>18:00-18:30</b>	<b>Poster session (13 – University Club)</b>
<b>18:30-19:30</b>	<b>Special session (14 – University Club)</b> It's 30 years... Chair: Simo Puntanen
<hr/>	
<b>18:30-20:00</b>	<b>Refreshments and cash bar</b>

Program

---

Wednesday August 27 (9:00-22:00)	
<b>9:00-10:30</b>	<b>Invited speaker session (15 – Vallfarten)</b> Chair: Lynn LaMotte
9:00-9:45	Thomas Mathew <i>Small Sample Asymptotics: Two Applications</i>
9:45-10:30	Jamal Najim <i>Spiked Models in Large Random Matrices and two statistical applications</i>
<b>10:30-11:00 Coffee break (University Club)</b>	
<b>11:00-13:00 Invited speaker session (16 – Vallfarten)</b> Ericsson Session - Young Scientists Awards Chair: Julia Volaufova	
11:00-11:30	Maryna Prus <i>Linear Aspects in Random Coefficient Regression Models: Optimal Designs for Different Design Criteria</i>
11:30-12:00	Jolanta Pielaszkiewicz <i>Free Probability approach to Random Matrices – An alternative Free Moment - Cumulant relation formula</i>
12:00-12:30	Fatma Sevinc Kurnaz <i>A Robust Liu Regression Estimator</i>
12:30-13:00	Alena Bachrata <i>Experimental Designs under Resource Constraints: Algorithmic Construction</i>
<b>13:00-14:00 Lunch (University Club)</b>	
<b>14:30-17:00 Excursion</b>	
<b>19:00-22:00 Conference dinner (University Club)</b>	
Thursday August 28 (8:30-10:20)	
<b>8:30-9:50 Session (17a – Vallfarten)</b> Chair: Dennis Cook	
8:30-8:50	Joseph Nzabanita <i>Estimation in the multivariate linear normal models with linearly structured covariance matrices</i>
8:50-9:10	Jan Hauke <i>On some applications of saddle-point matrices</i>
9:10-9:30	Hilmar Drygas <i>Gauss-algorithm and Gram-Schmidt orthogonalization</i>
9:30-9:50	Ying Li <i>A two-step model for linear prediction with connections to PLS</i>
<b>8:30-9:50 Session (17b – Planck)</b> Chair: Johannes Forkman	
8:30-8:50	Martin Ricker <i>A new method to compare statistical growth curves: The PL-GMANOVA model and its application</i>
8:50-9:10	Lars Rönnegård <i>Rotating a Linear Mixed Model for Multiple Testing in Genome-wide Association Studies having Repeated Measurements</i>
9:10-9:30	Majbritt Felleki <i>A Hierarchical Model with a Correlation between the Levels: An Example from Animal Breeding</i>
9:30-9:50	M. Revan Özkale <i>Diagnostic statistics for binary logistic regression in the presence of multicollinearity</i>
<b>9:50-10:20 Coffee break (Planck)</b>	

Program

---

Thursday August 28 (10:20-14:30)	
<b>10:20-11:20</b>	<b>Session (18a – Nobel)</b> Chair: Silvelyn Zwanzig Vicenç Torra <i>New results on the Choquet integral based distributions</i> Mahendran Shitan <i>First-Order Random Coefficient Autoregressive (RCA(1))</i> <i>Model: Joint Whittle Estimation and Information</i>
<b>10:20-11:20</b>	<b>Session (18b – Planck)</b> Chair: Francisco Carvalho Sergii Kulishov <i>Fractal and Antifractal Oxymorons, Moebius Strip Like Transformations of Biomedical Data as Basis for Exploratory Subgroup Analysis</i> Hiba Nassar <i>A Functional Hodrick Prescott Filter</i> Petr Koldanov <i>Statistical inferences for market network analysis</i>
<b>11:20-11:40 Coffee break (Planck)</b>	
<b>11:40-13:00</b>	<b>Special session (19a – Nobel)</b> Linear models applied to complex problems Chair: Thomas Holgersson Chengcheng Hao <i>Explicit influence analysis for count data under AB-BA crossover trials</i> Yuli Liang <i>Hypothesis testing in multilevel models with block circular covariance structures</i> Matias Quiroz <i>Speeding Up MCMC by Efficient Data Subsampling</i> Bertil Wegmann <i>Bayesian inference for heteroscedastic Rician time series with applications to fMRI</i>
<b>11:40-12:40</b>	<b>Session (19b – Planck)</b> Chair: Timo Schmid Ufuk Beyaztas <i>Some Notes on Sufficient Bootstrap</i> Beste Hamiye Sertdemir <i>New Block Bootstrap Methods: Sufficient and/or Ordered</i> Silvelyn Zwanzig <i>SimSel - a method for variable selection</i>
<b>13:00-14:30</b>	<b>Lunch (University Club)</b>

## Program

---

---

---

<b>Thursday August 28 (14:30-16:30)</b>	
<b>14:30-15:50</b>	<b>Session (20a – Planck)</b>
	Chair: Eva Fiserova
14:30-14:50	Sandra Donevska <i>Variable selection stepwise procedure for compositional data</i>
14:50-15:10	Alzbeta Kalivodova <i>Partial least squares discriminant analysis for compositional data: logratio methodology and its application to metabolomics</i>
15:10-15:30	Fikriye Kurtoglu <i>Restricted ridge estimation of generalized linear models</i>
15:30-15:50	Huseyin Guler <i>Nonlinear Inequality Constrained Ridge Regression Estimator</i>
<b>14:30-15:50</b>	<b>Special session (20b – Vallfarten)</b>
	Small Area Estimation
	Chair: Stephen Haslett
14:30-14:50	Stephen Haslett <i>Small area estimation methods and their links to linear and linear mixed models</i>
14:50-15:10	Geoffrey Jones <i>"Hard" versus "soft" predictions from unit-level models for small area estimation of proportions.</i>
15:10-15:30	Innocent Ngaruye <i>Small Area Estimation for multivariate repeated measures data</i>
15:30-15:50	Timo Schmid <i>Bias correction of robust small area estimators under spatial correlation</i>
<b>15:50-16:10</b>	<b>Coffee break (University Club)</b>
<b>16:10-17:00</b>	<b>Conclusion and Awards (Vallfarten)</b>

---

## 2 | Abstract - Invited Speakers

### Some Perspectives about Generalized Linear Modeling

**Alan Agresti**

University of Florida, Gainesville, Florida, USA

#### Abstract

This talk discusses several topics pertaining to generalized linear modeling. With focus on categorical data, the topics include (1) the large-sample behavior of residuals, (2) cautions in using Wald inference when effects are large or near the boundary of the parameter space, (3) bias in using ordinary linear models with ordinal categorical response data, (4) an improved way to use the generalized estimating equations (GEE) method for marginal modeling of a multinomial response, (5) a Bayesian challenge for high-dimensional data. I will not present any new research results, but the topics seem relevant for a conference devoted to perspectives in linear statistical inference. These topics got my attention while I was recently writing a book, *Foundations of Linear and Generalized Linear Models* (to appear, Wiley 2015).

**Keywords:** Bayesian methods, categorical data, discrete data, GEE, multinomial distribution, ordinal data, residuals, Wald inference, zero-inflated data

### Big Data, Big Bias, Small Surprise!

**S. Ejaz Ahmed**

Brock University, St. Catharines, Canada

#### Abstract

In high-dimensional data settings where number of variables is greater than observations, or when number of variables are increasing with the sample size many penalized regularization approaches were studied for simultaneous variable selection and estimation. However, with the existence of covariates with weak effect, many existing variable selection methods may not distinguish covariates with weak signals and no signal. In this case, the prediction based on a selected submodel may not be desirable. In this talk, we propose a high-dimensional shrinkage estimation strategy to improve the prediction performance of a given submodel. Such a high-dimensional shrinkage estimator (HDSE) is constructed by shrinking a ridge estimator in the direction of a predefined candidate submodel. Under an asymptotic distributional quadratic risk criterion, its prediction performance is toughly investigated. We reveal that the proposed HDSE performs better than the full model estimator. More importantly, it significantly improves the prediction

performance of any candidate submodel generated from most existing variable selection schemes. The relative performance of the proposed HDSE strategy is appraised by both simulation and the real data analysis.

## Envelopes: Methods for Efficient Estimation in Multivariate Statistics

**R. Dennis Cook**

University of Minnesota, Minneapolis, USA

### Abstract

An envelope is a nascent construct for increasing efficiency in multivariate statistics without altering the traditional goals. For instance, an envelope estimator of the coefficient matrix in a multivariate linear model has the potential to be substantially less variable than standard estimators, sometimes equivalent to taking thousands of additional observations [1, 2]. Improvements in efficiency are made possible by recognizing that the responses and predictors may contain information that is immaterial to the purpose of estimating the coefficients, while still introducing substantial extraneous variation into estimation. This leads to a general construct – an envelope – for enveloping the material information and thereby reducing estimative variation.

Envelopes also link with some standard multivariate methodology. In particular, partial least squares regression depends fundamentally on an envelope and this envelope can be used as a well-defined parameter that characterizes partial least squares. The establishment of an envelope as the nucleus of partial least squares then opens the door to pursuing the same goals but using envelope estimators that can significantly improve partial least squares predictions [3].

We will begin with a mostly intuitive introduction to response envelopes in the multivariate linear model and then briefly describe some of the asymptotic results and inner workings of envelopes. This will be followed by a discussion of predictor envelopes and their connection to partial least squares. We will also describe how to extend the scope of envelope methods beyond linear models. The discussion will include several examples for illustration. Emphasis will be placed on concept and their potential impact on data analysis. Papers on envelopes can be found at <http://users.stat.umn.edu/~rdcook/RecentArticles>, and MATLAB code for envelopes is available at <http://code.google.com/p/envlp>.

**Keywords:** Dimension reduction, Grassmannians, Partial least squares, Reducing subspaces, SIMPLS algorithm

### References

- [1] Cook, R.D., Li, B, and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate parameter estimation (with discussion). *Statistica Sinica*, 20:927–1010.
- [2] Cook, R.D., Helland, I. and Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society B*, 75:851–877.
- [3] Su, Z. and Cook, R.D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika*, 98:133–146.



## **Ratio and Regression Estimation of Finite Population Mean Using Optional Randomized Response Models**

**Sat Gupta**

University of North Carolina, Greensboro, USA

### **Abstract**

There is a vast body of literature on ratio and regression estimation of finite population mean when the primary variable of interest is non-sensitive. However, much less work has been done in this area when the primary variable is sensitive in nature. Gupta et al. (2012, *Communications in Statistics – Theory and Methods*) presented improved ratio and regression estimators of finite population mean of sensitive variables. In the present study, we improve these estimators further using optional randomized response methods.

## **Factor-effect formulations such that ANOVA effects can be tested with deleted-terms sums of squares**

**Lynn Roy LaMotte**

LSU Health, New Orleans, USA

### **Abstract**

In any linear model involving effects of categorical factors (e.g., A main effects, AB interaction effects), it is possible to identify the estimable parts and to describe the numerator sum of squares for a test statistic. However, it is simpler, both conceptually and computationally, to obtain the sum of squares by deleting the terms for the effect in question from the full model. This does not work generally for dummy-variable formulations; it has been observed to work in some settings with some other schemes.

The objective for this talk is to characterize factor-effect formulations for which this deleted-terms scheme produces the appropriate numerator sum of squares for the effect in question.

**Keywords:** ANOVA effects, unbalanced models, estimable effects

## **Joint modeling of a multilevel factor analytic model and a multilevel covariance regression model**

**Emmanuel Lesaffre**

Erasmus MC, Rotterdam, the Netherlands, and L-Biostat, KULeuven, Leuven, Belgium

### **Abstract**

We propose a novel modeling approach that can model both the mean structure and the covariance structure with a mixed effects model in a multivariate context. We called this the multilevel covariance regression (MCR) model. When the dimension of the response is high, a joint model of a multilevel factor analytic (MFA) model and an MCR model (MHOF model) is then proposed. We applied the MCR model to data from the

RN4CAST (Sermeus et al. 2011) FP7 project which involves 33,731 registered nurses in 2,169 nursing units in 486 hospitals in 12 European countries. The MHOF model was applied to the Belgium part of the project. As response we have taken in the first analysis the historically derived three burnout dimensions (Maslach and Jackson, 1981), while the MHOF model is based on the raw data, i.e. the responses to the 22-item questionnaire. The three burnout dimensions are emotional exhaustion (EE), depersonalization (DP) and personal accomplishment (PA). Applying the MHOF model to burnout could address the following questions simultaneously: 1) is the burnout structure the same as the commonly used structure by Maslach and Jackson? 2) how much variation of burnout could be explained by the level-specific fixed and random effects? 3) do the variances and correlations among burnout stay constant across level-specific characteristics and units at each level? We opted for the Bayesian approach as our estimating method for the MCR and MHOF models. The JAGS (just another Gibbs sampler) MCMC (Markov chain Monte Carlo) program was used through the R package rjags. Most parameters were assigned a non-informative prior except for the fixed and random effects in the factor loadings in the MCR part. These parameters were assigned a mixture prior respectively to overcome the "flipping states" issue in Bayesian context. Model comparison was done using the pseudo Bayes factor (PSBF).

## **Influence diagnostics in a vector autoregressive model**

**Yonghui Liu**

Shanghai University of International Business and Economics, China

### **Abstract**

In this talk, we use a likelihood approach and the local influence method introduced by Cook (1986) to study a vector autoregressive (VAR) model. We present the maximum likelihood estimators and the information matrix. We establish the normal curvature and slope diagnostics for the VAR model under several perturbation schemes and use the Monte Carlo method to obtain benchmark values for determining the influence of directional diagnostics and possible influential observations. An empirical study using the VAR model to fit real data of monthly returns of IBM and S&P500 index illustrates the effectiveness of the proposed diagnostics.

## **Small Sample Asymptotics: Two Applications**

**Thomas Mathew**

University of Maryland Baltimore County, Baltimore, USA

### **Abstract**

Standard likelihood based methods usually used to analyze data arising from a parametric model are typically accurate to the first order. Small sample asymptotic procedures provide major improvements in accuracy, and are available for discrete as well as for continuous data. In the talk, small sample asymptotics will be used for two applications. The first application is on the computation of tolerance limits under the logistic regression model for binary data. The data consist of binary responses, and upper tolerance limits are to be constructed for the number of positive responses in future trials

corresponding to a fixed level of the covariates. The problem has been motivated by an application of interest to the U.S. Army, dealing with the testing of ballistic armor plates for protecting soldiers from projectiles and shrapnel, where the probability of penetration of the armor plate depends on covariates such as the projectile velocity, size of the armor plate, etc. The second application is on a multivariate bioassay problem: several independent multivariate bioassays are performed at different laboratories or locations, and the problem of interest is to test the homogeneity of the relative potencies, assuming the usual slope-ratio or parallel line assay model. The problem has been investigated in the literature using likelihood based methods, under the assumption of a common covariance matrix across the different studies. This assumption is relaxed in this investigation.

Numerical results show that for both of the above applications, usual likelihood based procedures can be inaccurate in terms of providing satisfactory coverage probabilities or type I error probabilities. Furthermore, methodology based on small sample asymptotics results in significantly more accurate results in the small sample scenario. The first application will be illustrated using data from the U.S. Army dealing with the testing of ballistic armor plates. The bioassay application will be illustrated using data from a dental study, where pain intensity scores based on a standard treatment and a test treatment are analyzed.

**Keywords:** Combining bioassays, Logistic regression, Multivariate bioassay, Small sample asymptotics, Tolerance limit.

## Spiked Models in Large Random Matrices and two statistical applications

**Jamal Najim**

Université Paris-Est Marne La Vallée and CNRS (Centre National de la Recherche Scientifique, Paris, France

### Abstract

Large Random Matrix (LRM) theory goes back to Wigner's pioneering work in the late 40's and has witnessed a tremendous activity these last 20 years. Recently, LRM theory has proved to be well-adapted to handle statistical problems where in a series of observations, the dimension  $N$  of each observation is of the same order as the size  $n$  of the sample; this assumption is often formally stated as

$$\frac{N}{n} \rightarrow c \in (0, \infty) .$$

Spiked models [1] have been introduced in the early 2000 to account for data where a significant part of the information lies in the largest eigenvalues of associated empirical covariance matrices. In this talk, we will review important results associated to spiked models and also present two applications to statistical signal processing. The first application is devoted to signal detection [2]; the second one is concerned with direction of arrival estimation [3].

**Keywords:** Large Random Matrices, Spiked Models, Statistical Signal Processing

**References**

- [1] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *The Annals of Statistics*, 29(2):295–327.
- [2] Bianchi, P. and Debbah, M. and Maida, M. and Najim, J. (2011). Performance of Statistical Tests for Single Source Detection using Random Matrix Theory. *IEEE Trans. Inf. Th.*, 57(4):2400–2419.
- [3] Hachem, W. and Loubaton, P. and Mestre, X. and Najim, J. and Vallet, P. (2013) A subspace estimator for fixed rank perturbations of large random matrices. *Journal of Multivariate Analysis*, 114:427–447.

**Test for the mean in a Growth Curve model for high dimensions**

Muni S. Srivastava<sup>1</sup> and Martin Singull<sup>2</sup>

<sup>1</sup>University of Toronto, Canada

<sup>2</sup>Linköping University, Sweden

**Abstract**

In this paper, we consider the problem of estimating and testing a general linear hypothesis in a general multivariate linear model, the so called Growth Curve Model, when the  $p \times N$  observation matrix is normally distributed with an unknown covariance matrix.

The maximum likelihood estimator (MLE) for the mean is a weighted estimator with the inverse of the sample covariance matrix which is unstable for large  $p$  close to  $N$  and singular for  $p$  larger than  $N$ . We modify the MLE to an unweighted estimator and propose a new test which we compare with the previous likelihood ratio test (LRT) based on the weighted estimator, i.e., the MLE. We show that the performance of the LRT and the new test based on the unweighted estimator are similar.

For the high-dimensional case, when  $p$  is larger than  $N$ , we construct two new tests based on the trace of the variation matrices due to the hypothesis (between sum of squares) and the error (within sum of squares).

To compare the performance of these four tests we compute the attained significance level and the empirical power.

**Keywords:** Asymptotic distribution; High dimension; GMANOVA; Growth Curve Model; Estimation; Hypothesis testing; Power comparison.

## Formulas, equalities and inequalities for covariance matrices of estimations under a general linear model

**Yongge Tian**

Central University of Finance and Economics, Beijing, China

### Abstract

I'll introduce in this talk how to establish expansion formulas for calculating ranks and inertias of covariance matrices of estimations and their operations under a general linear model, and show to derive various algebraic and statistical properties of the estimations from the formulas, including establishing various possible equalities and inequalities in the Löwner sense for some well-known estimations under a general linear model.

## Generalized $R^2$ in Linear Mixed Models

**Julia Volaufova, Lynn R. LaMotte, and Ondrej Blaha**

LSU Health, School of Public Health, New Orleans, Louisiana, USA

### Abstract

The  $R^2$  statistic in a fixed-effects regression setting is routinely interpreted as a measure of proportion of variability explained by the model. During the recent decades there is an attempt to come up with a similar measure for linear mixed models. Here we present a brief overview of possible generalizations and different definitions of  $R^2$  for linear mixed effects models. We suggest a new definition based on a relation between the linear fixed and mixed models. We try also to address questions such as “Do the different  $R^2$  measures coincide in a given specific model?” “What do they measure?”, etc..

**Keywords:**  $R^2$  statistic, fixed effects model, random effects model.

## Experimental Designs under Resource Constraints: Algorithmic Construction

Alena Bachratá and Radoslav Harman  
Comenius University in Bratislava, Bratislava, Slovakia

### Abstract

We developed a heuristic method related to the Detmax procedure [4], for constructing efficient designs of experiments under a broad class of resource constraints. The resource constraints can represent restrictions on various types of limits associated with the experiment.

To illustrate the performance of the heuristic, we computed D-efficient experimental designs for estimating a set of treatment contrasts in the case of block size-two experiments with upper limits on the number of replications of each non-control treatment and an upper limit on the total number of treatments.

**Keywords:** Block design, Design of experiments, Heuristic optimization, Resource constraints

### References

- [1] Atkinson, A.C., Donev, A.N., Tobias, R.D. (2007). Optimum Experimental Designs, With SAS. *Oxford University Press*.
- [2] Cheng, C.S. (1981). Maximizing the total number of spanning trees in a graph: two related problems in graph theory and optimum design theory. *J. Combin. Theor. B* 31, 240-248.
- [3] Harman, R., Bachratá, A., Filová, L. (2014). Heuristic construction of exact experimental designs under multiple resource constraints. *e-print arXiv*: 1402.7263.
- [4] Mitchell, T.J. (1974). An Algorithm for the Construction of "D- Optimal" Experimental Designs. *Technometrics* 16(2), 203-210.

## A Robust Liu Regression Estimator

Fatma Sevinc Kurnaz<sup>1</sup> and Peter Filzmoser<sup>2</sup>  
<sup>1</sup>Yildiz Technical University, Istanbul, Turkey  
<sup>2</sup>Vienna University of Technology, Vienna, Austria

### Abstract

The least squares regression estimator is the best linear unbiased estimator under the normality assumption of the errors. However, it can be very sensitive in presence of multicollinearity and outliers in the data. We introduce new robust Liu type estimates to circumvent these problems. Depending on the data size, different proposals are made: If the number of observations  $n$  is larger than the number of explanatory variables  $p$ , the proposed estimator is based on the MM regression estimator [5]. When  $p > n$ , the proposed estimates are based on a robust version of the Partial Least Squares estimator,

the so-called Partial Robust M estimator [4]. In both cases, we propose methods for a robust estimation of the biasing parameter; one of them is using the minimization of the mean squared error according to biasing parameter, and the other one is based on cross-validation. The superiority of the proposed estimates to other estimators [1, 2, 3] is illustrated through both simulation studies and real data sets.

**Keywords:** Robust estimator; Liu estimator; MM estimates; partial least squares, partial robust M estimator.

### References

- [1] Kan B., Alpu O., Yazıcı B. (2013). Robust ridge and robust Liu estimator for regression based on the LTS estimator. *Journal of Applied Statistics*, 1:799-821.
- [2] Liu K. (1993). A new class of biased estimates in linear regression. *Communications in Statistics: Theory and Methods*, 22:393-402.
- [3] Maronna R. A. (2011). Robust Ridge Regression for High-Dimensional Data *Technometrics*, 53:44-53.
- [4] Serneels S., Croux C., Filzmoser P., Espen, P. J. V. (2005). Partial Robust M-Regression *Chemometrics and Intelligent Laboratory Systems*, 79:55-64.
- [5] Yohai V. J. (1987). High breakdown point and high efficiency robust estimates for regression *The Annals of Statistics*, 15:642-656.

## Free Probability approach to Random Matrices – An alternative Free Moment–Cumulant relation formula

Jolanta Pielaszkiwicz<sup>1</sup>, Dietrich von Rosen<sup>1,2</sup> and Matrin Singull<sup>1</sup>

<sup>1</sup>Linköping University, Linköping, Sweden

<sup>2</sup>Swedish University of Agricultural Sciences, Uppsala, Sweden

### Abstract

Talk introduces audience to the Free Probability theory as a source of tools to approach some questions related to random matrices. The goal is to present a new free cumulant–moment recurrent relation formula, motivate its importance and compare it to already existing results. We put our formula in relation to results widely studied and developed by Speicher and Nica [3] as well as to some non–recursive formulation, see [2]. Comparative example will be given.

**Keywords:** free cumulants, free moments, free probability, non–commutative probability space, R–transform, random matrices, Stieltjes transform

### References

- [1] Marčenko, V. A., Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, **72**(114):4:507–536.

- [2] Mottelson, I. W. (2012). Introduction to non-commutative probability. *http : //www.math.ku.dk/~musat/Free%20probability%20project\_final.pdf*. Accessed 25 July 2014.
- [3] Nica, A., Speicher, R. (2006). Lectures on the Combinatorics of Free Probability, Cambridge University Press, Cambridge.
- [4] Pielaszkiewicz, J., von Rosen, D., Singull, M. (2014). On Free Moments and Free Cumulants. *LiTH-MAT-R*, 2014:05 .

## Linear Aspects in Random Coefficient Regression Models: Optimal Designs for Different Design Criteria

Maryna Prus and Rainer Schwabe  
Otto-von-Guericke-University, Magdeburg, Germany

### Abstract

In random coefficient regression models the prediction of linear aspects of the individual effects are often of prior interest. In the case of given population parameters of location, the theory for the determination of optimal designs for prediction of the random effects has been developed by Gladitz and Pilz [1]. If the population parameters are unknown, these designs remain optimal for the prediction of individual deviations from the mean response (see Prus and Schwabe [3]), but they lose optimality and may become worthless for the prediction of the individual parameters themselves. The latter situation was considered in detail by Prus and Schwabe [2].

This talk provides an analytical approach for determining optimal designs for the prediction of some linear combinations of the random effects for different design criteria. The theoretical results will be illustrated by a simple example of straight line regression.

**Keywords:** optimal design for prediction, random coefficient regression, estimable aspects

### References

- [1] Gladitz, J., Pilz, J. (1982). Construction of optimal designs in random coefficient regression models. *Statistics* 13: 371-385.
- [2] Prus, M., Schwabe, R. (2011). Optimal Designs for Individual Prediction in Random Coefficient Regression Models. In: V. Melas, G. Nachtmann and D. Rasch (Eds.). *Optimal Design of Experiments - Theory and Application: Proceedings of the International Conference in Honor of the late Jagdish Srivastava*: 122-129. Center of Experimental Design, University of Natural Resources and Life Sciences, Vienna.
- [3] Prus, M., Schwabe, R. (2013). Optimal designs for the prediction of individual effects in random coefficient regression. In: D. Uciniski, A. C. Atkinson and M. Patan (Eds.) *mODa 10 - Advances in Model-Oriented Design and Analysis*: 211-218. Springer, Cham.



## 3 | Abstract - Contributed Speakers

### Efficiency of the Generalized Difference-based Liu Estimator in Partially Linear Models with Correlated Errors

**Esra Akdeniz-Duran**

Istanbul Medeniyet University, İstanbul, Turkey

#### Abstract

A generalized difference-based estimator is introduced for the parameter vector  $\beta$  in a partially linear model with correlated errors. The generalized difference-based LIU-estimator is defined for the vector parameter in a semiparametric regression model. Under the linear nonstochastic constraint,  $R\beta = r$ , a generalized restricted difference-based LIU estimator is given. The performance of the new estimator is evaluated by using a simulated data set. Finally, a numerical example is presented showing the advantages of the proposed estimator, with respect to MSE.

**Keywords:** Differencing matrix, Generalized restricted difference-based Liu estimator, Multicollinearity

#### References

- [1] Akdeniz-Duran, E., Haerdle, W. and Osipenko, M. (2012). Difference based ridge and Liu type estimators in semiparametric regression models. *Journal of Multivariate Analysis*, 105: 164-175.
- [2] Akdeniz, F., Akdeniz-Duran, E., Roozbeh, M. and Arashi, M. (2013). New difference-based estimator of parameters in semiparametric regression models. *Journal of Statistical Computation and Simulation*, 83(5):810-824.
- [3] Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics: Theory and Methods*, 22:393-402.
- [4] Green, P.J. and Silverman, B.W. (1994). Nonparametric Regression and Generalized Linear Models. *Chapman and Hall*.
- [5] Härdle, W.K., Müller, M., Sperlich, M.S. and Werwatz, A. (2004). Nonparametric and Semiparametric Models, *Springer Verlag, Heidelberg*.

## Fitting spatial hurdle models using a hierarchical likelihood method

Shengqiang Guo and **Md. Moudud Alam**  
Dalarna University, Borlänge, Sweden

### Abstract

The hurdle model is widely used for analysing count response with excessive zeros. For spatially observed counts, one also has to take care of possible spatial correlation along with excessive zero counts. Computational procedure for fitting hurdle mixed model with independent random effects are already available (see e.g. [2]). However, in absence of any statistical software package for fitting spatial hurdle model in a non-Bayesian way, MCMC based Bayesian approaches are often used by the applied researchers, but they are computationally too slow. In this paper we show that a hurdle mixed model with spatially correlated random effects can be fitted via the hierarchical likelihood method [1] by using an iterative weighted least square algorithm. We use conditional autoregressive random effects to deal with spatial correlation in the hurdle model and present a simplified computational procedure by using eigenvalue decomposition of the covariance matrix of the random effects. We demonstrate the finite sample properties of the proposed estimator in a simulation study. We also present a real data example by analysing reindeer pellet group counts obtained from a survey conducted at Storliden mountain area in the north of Sweden.

**Keywords:** CAR normal random effect, spatial correlation, eigenvalue decomposition, EQL, reindeer distribution.

### References

- [1] Lee, Y. Nelder, J. A. (1996). Hierarchical Generalized Linear Models. *Journal of the Royal Statistical Society B*, 58:619–678.
- [2] Molas, M., and Lesaffre, E. (2010). Hurdle models for multilevel zero-inflated data via h-likelihood. *Statistics in Medicine*, 29:3294–3310.

## Gamma frailty for the exponential-zero truncated Poisson distribution

Öykum Esra Askin<sup>1</sup> and Deniz Inan<sup>1,2</sup>

<sup>1</sup>Yildiz Technical University, Istanbul, Turkey

<sup>2</sup>Marmara University, Istanbul, Turkey

### Abstract

In standard time to failure models, it is assumed that all individuals in the study group are subject to the same risk. However, in general, study population is not homogeneous and unobservable sources of heterogeneity have different effect (frailty) on each individual's hazard. In this study, a model with gamma distributed frailty for the exponential-zero truncated Poisson baseline hazard is introduced. Unconditional survival and hazard

functions are derived and model parameters are estimated via maximum likelihood by using the Expectation and Maximization (EM) algorithm.

**Keywords:** Exponential-Poisson distribution, Gamma frailty, EM algorithm

### References

- [1] Adamidis, K. and Loukas, S., (1998). A life time distribution with decreasing failure rate. *Statist. Probab. Lett.* 39, 35–42.
- [2] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist.Soc. Ser. B* 39, pp. 1–38.
- [3] Kus C.(2007). A new lifetime distribution, *Comput. Statist. Data Anal.* 51, pp. 4497–4509.
- [4] Ng, H.K.T., Chan, P.S. and Balakrishnan, N. (2002). Estimation of parameters from progressively censored data using EM algorithm. *Comput. Statist. Data Anal.* 39, pp. 371–386.

## Optimum Designs for Two Treatments with Unequal Variances

**Anthony C. Atkinson**

London School of Economics, United Kingdom

### Abstract

Let the means of the two treatments be  $\mu_1$  and  $\mu_2$ . If estimation of both parameters is of interest, the D-optimum design is that for a model in which the number of support points of the design (allocation to one or the other treatment) is the same as the number of parameters. Thus subjects should be allocated equally to the two treatments, regardless of the ratio of the variances. If, on the other hand, interest is in estimation of the treatment difference  $\mu_1 - \mu_2$ , "Neyman" allocation is appropriate, in which allocation is proportional to the standard deviation of the measurements.

Now suppose there are covariates. In clinical trials these are usually treated as nuisance parameters, over which balance is obtained by randomization. The two designs given above are therefore appropriate, depending on the objectives of the experimenter. However, in personalised medicine, the coefficients of the covariates will also be of importance in deciding on treatment. [1] finds D-optimum designs when there is an arbitrary number of covariates and describes how the optimum design varies with the number of covariates.

The talk will describe this work and extend it to design when treatment difference is of interest in the presence of covariates. The approach taken to this problem of "generalized Neyman allocation" is  $D_A$ -optimality. Surprisingly general results are obtained with a minimum of algebra (although with the use of a computer).

**Keywords:**  $D_A$ -optimality, "Generalized Neyman allocation", Personalized medicine.

## References

- [1] Atkinson, A.C. (2015). Optimum designs for two treatments with unequal variances in the presence of covariates. (Submitted).

## Bootstrap approach in parallel analysis to determine non-zero canonical correlations

Selahattin Aydogdu and Gulhayat Golbasi Simsek  
Yildiz Technical University, Istanbul, Turkey

### Abstract

Canonical Correlation Analysis (CCA) is a multivariate statistical method investigating the relationships between two sets of variables. Canonical correlations are the maximized correlation coefficients between two new sets which are linear transformations of the original variables from both sets and not correlated within the sets. The number of nonzero canonical correlations in population is called dimensionality. Parallel Analysis (PA) was developed for determining the number of principal components or factors to retain in the factor analysis (Horn, 1965). In PA, the first step is similar to a Monte Carlo simulation process of generate random samples from hypothesized model of zero dimensionality. The second main step is the comparison of random eigenvalues to the from original sample. In this study, the contribution of PA to the dimensionality issue of CCA is twofold. The first is to investigate the applicability of classical PA as a dimensionality determination method in CCA. The second, we propose a bootstrap modification to classical PA in which simulated eigenvalues are compared to the bootstrap distribution of ordered eigenvalues from the original sample. The performances of classical PA and the bootstrap modified PA in terms of true, under and over determination of dimensionality are compared to ones from the classical test statistics with an extensive simulation study in which the canonical correlations of designed populations are various in number and magnitude.

**Keywords:** Canonical correlations, dimensionality, parallel analysis, Bootstrap

## References

- [1] Buja, A. and Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, 27(4), 509-540.
- [2] Calinski T., Krzysko M. and Wolynski W. (2006). A comparison of some tests for determining the number of nonzero canonical correlations. *Communication in Statistics-Simulation and Computation*, 35:727-749.
- [3] Horn, J. L. (1965). A rationale and test fort he number of factors in factor analysis. *Psychometrika*, 30:179-185.
- [4] O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, Computers*, 32(3): 396-402.

## Some Notes on Sufficient Bootstrap

Ufuk Beyaztas<sup>1</sup> and Aylin Alin<sup>1,2</sup>

<sup>1</sup>Dokuz Eylul university, Izmir, Turkey

<sup>2</sup>Michigan State university, East Lansing, USA

### Abstract

The idea of nonparametric bootstrap was introduced by Bradley Efron in the late 1970s, and it has become a general framework in almost all area of statistics. The main idea of bootstrap is basically based on the plug-in principle. Let  $\chi_n = (X_1, X_2, \dots, X_n)$  be independently and identically distributed random variables from an unknown distribution function  $F$ , whose parameter  $\theta_0$  is of our interest. The parameter  $\theta_0$  can be expressed as some functional,  $T$ , of the unknown distribution function  $F$  i.e.,  $\theta_0 = T(F)$ . Then, according to plug-in principle, the bootstrap estimator of  $\theta_0$  can be expressed applying the same functional to the some empirical estimate of  $F$ ,  $\hat{F}$ , so that  $\hat{\theta}_n = T(\hat{F})$ . Generally,  $\hat{F}$  is considered as *empirical distribution function*,  $F_n$ , which put mass  $1/n$  at each data point  $X_i$  in the original data set of size  $n$ .

In recent years, different types of bootstrap method have been studied including the m-out-of-n ( $m/n$ ) and sufficient bootstrapping.  $m/n$  bootstrap proposed as re-sampling fewer than  $n$  observations from  $F_n$ , say  $m$  observations, to fix the inconsistency of conventional bootstrap under certain cases. On the other hand, sufficient bootstrapping was introduced to reduce the computational burden of conventional Bootstrap. The main idea of this method is to use only distinct units in re-samples. That is, the sufficient bootstrap re-sample is obtained after discarding the repeated units from the bootstrapped sample. In this study, we show that the sufficient bootstrap distribution is consistent with true distribution function as long as conventional bootstrap is consistent. However, it does not cure the problems as  $m/n$  bootstrap in non-regular cases for which we propose sufficient  $m/n$  bootstrap.

**Keywords:** Asymptotic, Sufficient bootstrap, m-out-of-n bootstrap.

## Optimum Dose Regimen in Clinical Trials

Kabir Soeny<sup>1</sup>, Barbara Bogacka<sup>1</sup>, Byron Jones<sup>2</sup> and Thomas Bouillon<sup>2</sup>

<sup>1</sup>Queen Mary, University of London , United Kingdom

<sup>2</sup>Novartis Pharma AG, Basel, Switzerland

### Abstract

In this talk we will present optimum experimental designs for dose regimen selection in clinical trials, where the criteria of optimality meet the requirements that the drug concentration be close to the target drug exposure needed to cure the disease and ensure a target reduction in viral load as well as minimizing the drug exposure once the target is met. Furthermore, in cases when the medication is a combination of drugs, we find an optimum ratio of the components as well as an optimum dosing regimen. We define new criteria of optimality, which as described above, come from the practical aspects of drug development. We present their properties as well as the new optimization algorithm. This methodology is particularly applicable to infectious diseases, such as malaria.

**Keywords:** pharmacokinetic optimality criteria, pharmacodynamic optimality criteria.

## Bayesian uncertainty analysis for tipping points modelled by computer simulators

**Camila Caiado**

Durham University, United Kingdom

### Abstract

Physical systems are often modelled by computer simulators that may contain one or more equilibrium states which may appear as discontinuities on the model's input space. We use multiple emulation and a probabilistic classifier to investigate the system's behaviour in different states and estimate boundaries between such regions. We illustrate this approach with a four-box ocean circulation model.

## Mean Driven Balance and Uniformly Best Linear Unbiased Estimators

**Francisco Carvalho**<sup>1,2</sup>, Roman Zmyślony<sup>3,4</sup> and João T. Mexia<sup>2,5</sup>

<sup>1</sup>CMA, Universidade Nova de Lisboa, Caparica, Portugal

<sup>2</sup>Instituto Politécnico de Tomar, Tomar, Portugal

<sup>3</sup>University of Zielona Góra, Zielona Góra, Poland

<sup>4</sup>University of Opole, Opole, Poland

<sup>5</sup>Universidade Nova de Lisboa, Caparica, Portugal

### Abstract

The equivalence of Ordinary Least Squares Estimators, OLSE, and Gauss-Markov Estimators, GME, for models with variance-covariance matrix  $\sigma^2\mathbf{M}$  is extended to derive a necessary and sufficient balance condition for mixed models with mean vector  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ , with  $\mathbf{X}$  an incidence matrix, having OLSE for  $\boldsymbol{\beta}$  that are Best Linear Unbiased Estimator, BLUE, whatever the variance components. This approach leads to least squares like estimators for variance components.

To illustrate the range of applications for the balance condition, interesting special models are considered.

**Keywords:** OLSE, Orthogonal block structure models, UBLUE, Kruskal condition

### References

- [1] Caliński, T. & Kageyama, S. (2000) *Block Designs. A Randomization Approach. Volume I: Analysis*. Lecture Notes in Statistics. Springer.
- [2] Caliński, T. & Kageyama, S. (2003). *Block Designs. A Randomization Approach. Volume II: Design*. Lecture Notes in Statistics. Springer.
- [3] Gotu, B. (2001). The equality of OLS and GLS estimators in the linear regression model when the disturbances are spatially correlated. *Statistical Papers*, **42**, 253–263.

- [4] Isotalo, J.; Puntanen, S. (2009). A note on the equality of the OLSE and the BLUE of the parametric function in the general Gauss-Markov model. *Statistical Papers*, **50**, 185–193.
- [5] Kruskal, W. (1968). When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. *The Annals of Mathematical Statistics*, **39** 1, 70–75.
- [6] Zmysłony, R. (1980). *A characterization of Best Linear Unbiased Estimators in the general linear model*. Lecture Notes in Statistics. Springer. **2**, 365–373.

## Mahalanobis distances of factor structure data

**Deliang Dai**

Linnaeus University, Växjö, Sweden

### Abstract

In this paper, we implement the factor model as an option to derive the covariance matrix that is used for building a Mahalanobis distance. The distribution properties of the new Mahalanobis distances are derived. The contamination effects of outliers on Mahalanobis distances from separate parts of the factor model are also investigated. The results from an empirical example show the difference between detecting the outliers by using the separated Mahalanobis distances and the original Mahalanobis distance.

**Keywords:** factor model, Mahalanobis distance, outlier detection, systematic error

## Resampling Aalen-Johansen Processes: Why Efron’s Bootstrap and Permutation Techniques fail and Multiplier Approaches succeed.

**Dennis Dobler** and Markus Pauly

University of Duesseldorf, Germany

### Abstract

Especially in medical applications the analysis of transition probability matrices of non-homogeneous Markov processes is of great importance. While observations may be incomplete, e.g. due to random left-truncation and right-censoring, estimation of these matrices is conducted by employing the Aalen-Johansen estimator which is based on counting processes. However, results of weak convergence towards a Gaussian process cannot be utilized straightforwardly since its complicated limiting covariance structure depends on unknown quantities. For many null hypotheses of interest we show that classical resampling procedures, such as Efron’s bootstrap [1] or permutation techniques, fail to mimic the correct asymptotic covariances; see also [2] and [3]. This fact leads to the loss of control over the probabilities of type I errors, which is why these techniques need to be handled with caution. A solution to this problem is given by inserting bootstrap multipliers into a martingale representation of the Aalen-Johansen estimator. We suggest a new procedure, generalizing the wild bootstrap (cf. [4]) for a larger class of

possible distributions. While gaining the desired limiting distributions, consistent tests and confidence bands are easily developed. For illustrating the small sample performance simulation studies are presented.

**Keywords:** Aalen-Johansen Estimator, Efron's Bootstrap, Markov Process, Permutation, Resampling, Survival Analysis, Wild Bootstrap

## References

- [1] Efron, B. 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, vol. 7(1):1–26.
- [2] Pauly, M. 2011. Weighted resampling of martingale difference arrays with applications. *Electronic Journal of Statistics*, vol. 5:41–52.
- [3] Dobler, D. and Pauly, M. 2013. Bootstrapping Aalen-Johansen Processes for Competing Risks: Handicaps, Solutions, and Limitations. *arXiv:1401.7801*.
- [4] Beyersmann, J., Pauly, M. and Di Termini, S. 2013. Weak convergence of the wild bootstrap for the Aalen-Johansen estimator. *Scandinavian Journal of Statistics*, vol. 40(3):387–402.

## Variable Selection Stepwise Procedure for Compositional Data

Sandra Donevska<sup>1</sup>, Eva Fišerová<sup>1</sup>, Peter Filzmoser<sup>1,2</sup> and Karel Hron<sup>1</sup>

<sup>1</sup>Palacký University, Olomouc, Czech Republic

<sup>2</sup>Vienna University of Technology, Wien, Austria

### Abstract

Compositional data or compositions for short are multivariate observations, which components are called parts carrying only relative information. Naturally their sample space is the simplex with the Aitchison geometry [1]. Therefore for the purpose of performing standard statistical methods we have to transform the compositions to standard multivariate observations on the real space. Three log-ratio transformations will serve us for this intention: additive log-ratio (alr), centered log-ratio (clr) and isometric log-ratio (ilr) transformation.

In the contribution the aim will be to present a stepwise procedure with which the reduction of parts of the initial composition results in a subcomposition conveying all the important information from the original composition [1]. Thus the final subcomposition is much easier to handle i.e., it is simpler to perform multivariate statistical analysis and simultaneously to interpret their results. The usefulness of this procedure is shown on real world examples from geochemistry.

**Keywords:** Aitchison geometry on the simplex, Centered log-ratio transformation, Isometric log-ratio transformation, Variable selection



## References

- [1] Aitchison, J. (1986). The statistical analysis of compositional data. *Monographs on Statistics and Applied Probability*. Chapman & Hall Ltd., London (UK), 416 p.
- [2] Hron, K., Filzmoser, P., Donevska, S., and Fišerová, E. (2013). Covariance-based variable selection for compositional data. *Mathematical Geosciences*, 45:487-498.

## Gauss-algorithm and Gram-Schmidt orthogonalization

**Hilmar Drygas**

University of Kassel, Germany

### Abstract

The method of least squares leads to the normal equations for obtaining the estimator of the parameter vector of the linear model. These normal equations can be solved by the Gauss-elimination algorithm. This algorithm leads to the Gram-Schmidt orthogonalizers. The proof of this fact is rather simple. The Gauss-algorithm also leads to an upper triangular matrix. From this the covariance-matrix of the least squares estimator, the inverse of the design-matrix, can be derived. Other possibilities from the Gram-Schmidt orthogonalizers are: (i) Easy computation of the main determinants of the design-matrix (ii) Two methods of inversion by using the inverse of partitioned matrices (iii) Individual computation of variances and covariances.

## An alternative approach to solve the LAD-LASSO problem

**Esra Emiroglu** and Kadri Ulas Akay

Istanbul University, Turkey

### Abstract

The least absolute deviation (LAD) regression is more robust alternative to the popular least squares (LS) regression whenever there are outliers in the response variable, or the errors follow a heavy-tailed distribution. The least absolute shrinkage and selection operator (LASSO) is a popular choice for shrinkage estimation and variable selection. By combining these two classical ideas, LAD-LASSO is an estimator which is able to perform shrinkage estimation while at the same time selecting the variables and is resistant to heavy-tailed distributions and outliers. The aim of this article is to reformulate LAD-LASSO problem to solve with the Simplex Algorithm, which is an area of Mathematical Programming.

**Keywords:** LAD, LASSO, LAD-LASSO, Simplex Algorithm.

## References

- [1] Arthanari, T. S., Dodge, Y., (1993) *Mathematical Programming in Statistics*, John Wiley and Sons Inc., New York, USA.
- [2] Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society*, Series B Vol. 58, No. 1, 267-288.

- [3] Wang, H., Li, G. and Jiang, G. (2007). Robust Regression Shrinkage and Consistent Variable Selection Through The LAD-LASSO. *Journal of Business and Economic Statistics* 25, 347-355.

## Prediction of Turkish Bank Failures Using Panel Factor Logistic Regression

**Birsen Eygi Erdogan**

Marmara University, Istanbul, Turkey

### Abstract

Crisis in financial sector in over the last two decades have shown the importance of early warning systems especially for the bank failures. In this study it is aimed to develop an early warning system for Turkish commercial banks failure using panel data for the 2002-2012 periods. The data is analyzed using pooled logistic regression versus fixed, random and mixed panel logistic regression. The return on assets ratio was used for the definition of the bank failure that is the dependent variable. To avoid a multicollinearity problem, factors constructed by using a factor analysis on financial ratios were used as independent variables.

**Keywords:** Bank failure, Logistic regression, Panel data

### References

- [1] Altman E.I., (1968), Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Finance*, c.XXIII:589-609.
- [2] Canbas, S., Cabuk, A., Kilic, SB.(2005), Prediction of commercial bank failure via multivariate statistical analysis of financial structures: *Eur. J. Oper. Res.* 166:528-546., DOI: 10.1016/j.ejor.2004.03.023
- [3] Demyanyk, Y., Hasan, I., (2010), Financial crises and bank failures: A review of prediction methods , *OMEGA-INT J MANAGE SCI* , vol. 38, no. 5, pp. 315-324, DOI: 10.1016/j.omega.2009.09.007
- [4] Eygi Erdogan, B., (2013), Prediction of bankruptcy using support vector machines: an application to bank bankruptcy, *Journal of Statistical Computation and Simulation*, Vol. 83, Iss. 8, DOI: 10.1080/00949655.2012.666550
- [5] Ilk, O., Pekkurnaz, D., Cinko, M. (2013), Modeling company failure: a longitudinal study of Turkish banks, *Optimization: A Journal of Mathematical Programming and Operations Research*, DOI: 10.1080/02331934.2013.855762
- [6] Kumar, P. R. , Ravi, V., (2007), Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review, *European Journal of Operational Research* - *EJOR* , vol. 180, no. 1, pp. 1-28, DOI: 10.1016/j.ejor.2006.08.043

## Coordinate representation of compositional tables

**Kamila Fačevicová** and Karel Hron  
Palacký University, Olomouc, Czech Republic

### Abstract

Compositional tables represent a special case of compositional data [1] as positive vectors and continuous counterpart of contingency tables, which carry relative information about relationship between row and column factors [4]. Consequently, cells of a compositional table contain quantitatively expressed relative contributions on a whole, thus for their statistical analysis only ratios between cells are informative. The standard Euclidean geometry is not appropriate for this purpose and should be replaced by the Aitchison geometry on the simplex that fully accounts for specific features of compositional tables. Moreover, the Aitchison geometry enables to decompose the compositional table into its independent and interactive parts; in particular, the latter one is crucial for further investigation of the relationship between row and column factors [3]. Furthermore, by assigning orthonormal coordinates to decomposition of compositional tables, standard methods for their statistical processing can be applied [2].

This contribution deals with problem, how to choose the system of ilr coordinates which are well-interpretable and easy to handle. The method, which expresses the interaction table in coordinates by dividing the interaction table onto smaller tables, and its generalization are presented. The resulting coordinates are closely linked to odds ratio between cells of the interaction table and thus represent a natural alternative to the coordinates obtained from the well-known sequential binary partition for construction of orthonormal coordinates in compositional data analysis.

**Keywords:** balances, compositional tables, isometric logratio transformation

### References

- [1] Aitchison, J. (1986). *The statistical analysis of compositional data.*, Chapman and Hall, London.
- [2] Egozcue, J. J., Pawłowsky-Glahn, V., Mateu-Figueras, G., and Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Math Geol*, 35:279-300.
- [3] Egozcue, J. J., Díaz-Barrero, J. L., and Pawłowsky-Glahn, V. (2008). Compositional analysis of bivariate discrete probabilities. In *Daunis-i-Estadella J, Martín-Fernández JA, eds, Proceedings of CODAWORK'08, The 3rd Compositional Data Analysis Workshop*. University of Girona, Spain.
- [4] Fačevicová, K., Hron, K., Todorov, V., Guo, D., and Templ, M. (2014). Logratio approach to statistical analysis of 2 x 2 compositional tables. *Journal of Applied Statistics*, 41:944-958.

## A Hierarchical Model with a Correlation between the Levels: An Example from Animal Breeding

Majbritt Felleki

Dalarna University, Borlänge, Sweden, and  
Swedish University of Agricultural Sciences, Uppsala, Sweden

### Abstract

The Double Hierarchical Generalized Linear Models [2] came with a set of tools for inference in Generalized Linear Models with random effects when the parameters in the distributions of random effects were also containing random effects. The tools have been found to be implementable and perform well for a range of distributions of both the observations and random effects.

A particular example of a Double Hierarchical Generalized Linear Model with the random effects correlated between the levels was studied; a linear mixed model in which the variances of the residuals were containing random effects. Observations  $y$  were modelled  $y = X\beta + Za + Wp + e$ , where  $X$ ,  $Z$ , and  $W$  were known matrices,  $\beta$  was the vector of fixed effects,  $a$  and  $p$  were random effects, and  $e$  was the vector of residuals.

The residuals were heteroscedastic with  $\text{Var}(e) = \Phi$ , where  $\Phi$  was a diagonal matrix with diagonal  $\phi$ , and  $\log \phi = X_v\beta_v + Z_v a_v + W_v p_v$ . The distributions of the random effects were  $p \sim N(0, \sigma_p^2 I)$ ,  $p_v \sim N(0, \sigma_{p_v}^2 I)$ , and

$$\begin{pmatrix} a \\ a_v \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_a^2 A & \rho \sigma_a \sigma_{a_v} A \\ \rho \sigma_a \sigma_{a_v} A & \sigma_{a_v}^2 A \end{pmatrix}\right)$$

where  $A$  was a correlation structure for  $a$  and  $a_v$ , while  $I$  was the identity matrix. In animal breeding  $Z$  and  $Z_v$  would for instance link the observation  $y$  to the subject itself, or to the father or the mother of the subject, and  $A$  would be the genetic correlations among individuals under the assumption of an additive genetic structure.

To include the correlation  $\rho$  between the levels is new within the Double Hierarchical Generalized Linear Models, and a tool for estimation was developed. Moreover, as the model was fitted to large complex data on traits of related animals, an approximation was done. The algorithm has been found to perform well for the model stated [1]. The model was fitted by iterative reweighted least squares; by iteratively updating and fitting a bivariate linear mixed model.

**Keywords:** Animal Breeding, Double Hierarchical Generalized Linear Models

### References

- [1] Felleki, M., Lee, D., Lee, Y., Gilmour, A.R., and Rönnegård, L. (2012). Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. *Genetics Research*, 94:307-317.
- [2] Lee, Y., and Nelder, J.A. (2006). Double Hierarchical Generalized Linear Models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 55:139-185.

## Hypothesis testing for a separable covariance structure with AR(1) under the two-level multivariate model

Katarzyna Filipiak<sup>1</sup>, Daniel Klein<sup>2</sup> and Anuradha Roy<sup>3</sup>

<sup>1</sup>Poznań University of Life Sciences, Poland

<sup>2</sup>P. J. Šafárik University, Košice, Slovakia

<sup>3</sup>University of Texas at San Antonio, USA

### Abstract

Hypotheses testing for separability of a covariance structure for two-level multivariate data using likelihood ratio test (LRT) are widely studied in the literature. In this paper we use Rao's score test (RST) statistics to test the separability of a covariance structure with the first component as autoregression of order 1 (AR(1)) correlation matrix under the assumption of multivariate normality. We compare these two test procedures, and show that RST, unlike LRT, works for even very small sample sizes, which are not greater than the total dimensionality of the variables. We conduct simulation studies for the purpose of sample size consideration, for the estimation of empirical percentiles of the null distribution of RST as well as LRT statistics. We also compare powers of LRT and RST. Both tests are implemented with real data examples.

**Keywords:** Likelihood ratio test, Rao's score test, Separable covariance structure, AR(1) correlation structure, Maximum likelihood estimates.

## Regression with Compositional Response Variables

Eva Fišerová

Palacký University, Olomouc, Czech Republic

### Abstract

When the response variables are compositional, i.e. multivariate observations carrying only relative information (proportions, percentages), a special treatment for their regression analysis is necessary. Compositional data are characterized by the simplex sample space with the Aitchison geometry that forms the Euclidean structure of the sample space [1]. Using proper log-ratio transformations, the data are moved isometrically to the real Euclidean space where it is possible to use standard regression tools.

Although after the log-ratio transformation regression with compositional response naturally leads to multivariate regression, Egozcue et. al [3] proposed a univariate approach with series of submodels. A multivariate approach to regression modeling (see, e.g. [2]) has several advantages in comparison to a series of univariate models. Specifically, multivariate models respect the association between outcomes, and thus, in general, procedures are more efficient. Further, they can evaluate the joint influence of predictors on all outcomes and avoid the issue of multiple testing. On the other hand, there are situations when the multivariate model can be decomposed to a series of simpler models, univariate or multivariate, depending on the issue.

The aim of the contribution is to present differences and equivalences between multivariate and univariate approaches to regression with compositional response variables. We will show either the situations, when it is possible to use univariate approach, or the

cases, when multivariate analysis is necessary. Theoretical results will be applied to a real-world example from geochemistry.

**Keywords:** Aitchison geometry on the simplex, Log-ratio transformations, Multivariate regression, Univariate regression.

### References

- [1] Aitchison, J. (1986). The statistical analysis of compositional data. *Monographs on Statistics and Applied Probability*. Chapman & Hall Ltd., London (UK).
- [2] Seber, G.A.F. (2004). Multivariate Observations. *Wiley*.
- [3] Egozcue, J.J., Daunis-i-Estadella, J., Pawlowsky-Glahn, V., Hron, K. and Filzmoser, P. (2012). Simplicial Regression. The Normal model. *Journal of Applied Probability and Statistics*, 6:87-108.

## Inference in $k$ -Exchangeable Multivariate Models

Miguel Fonseca<sup>1</sup> and Anuradha Roy<sup>2</sup>

<sup>1</sup>Universidade Nova de Lisboa, Lisboa, Portugal

<sup>2</sup>The University of Texas at San Antonio, San Antonio, USA.

### Abstract

In this paper,  $k$ -exchangeable multivariate models will be the main focus. The  $k$ -exchangeable linear model arises when the  $m$ -dimensional error vectors are exchangeable and jointly normally distributed, a generalization of the doubly exchangeable structure. Decomposing the covariance matrix using quadratic subspaces, estimation and hypothesis testing is derived.

**Keywords:** multivariate models,  $k$ -exchangeable linear models, inference

### References

- [1] Roy, A and Fonseca, M. (2012). *Linear Models with Doubly Exchangeable Distributed Errors*. *Communications in Statistics - Theory and Methods*, 41:2545-69

## A simple parametric bootstrap method for testing principal components in normally distributed data

**Johannes Forkman**

Swedish University of Agricultural Sciences, Uppsala, Sweden

### Abstract

Forkman and Piepho [1] proposed a resampling method for testing interaction terms in models for analysis of genotype-by-environment data. The "genotype main effects and genotype-by-environment interaction effects" model is closely related to principal component analysis (PCA). For this reason, the method proposed by [1], which is called the "simple parametric bootstrap method", can be used for testing principal components in PCA as well. The proposed bootstrap method is parametric in the sense that it assumes homoscedastic and normally distributed observations. The method is "simple", because it only involves repeated sampling of standard normal distributed values. Specifically, no parameters need to be estimated. The present paper describes how the simple parametric bootstrap method can be used in PCA.

**Keywords:** Dimensionality reduction, Principal component analysis, Resampling methods, Singular value decomposition

### References

- [1] Forkman, J., and Piepho H. P. (2014). Parametric bootstrap methods for testing multiplicative terms in GGE and AMMI models. *Biometrics*. In press.

## Bayes linear uncertainty analysis for complex computer models

**Micheal Goldstein**

Durham University, United Kingdom

### Abstract

This talk provides an overview of the Bayes linear approach to uncertainty analysis for complex physical systems modelled by computer simulators, with particular emphasis on emulation, model discrepancy, history matching and forecasting.

# Nonlinear Inequality Constrained Ridge Regression Estimator

Huseyin Guler and Cigdem Kosar  
Cukurova University, Adana, Turkiye

## Abstract

The linear regression model may have some constraints on the parameters. Some of these constraints are nonlinear. [1] defined inequality constrained least squares (ICLS) estimator and ICLS could be used if the model parameters are nonlinearly constrained. However, if the model suffers from multicollinearity, ICLS leads to unstable parameter estimates, estimates with inflated variances, and size distortion for the hypothesis tests about the parameters. In order to solve this problem, [2] defined inequality constrained ridge regression (ICRR) estimator and showed that ICRR is better than ICLS in the sense of mean square error (mse) under multicollinearity. However, ICRR doesn't consider the inequality constraints and it can be applied only if the constraints are linear. In this paper, we have defined a new estimator called nonlinear inequality constrained ridge regression (NICRR) estimator with a Bayesian framework. NICRR deals with multicollinearity while allowing nonlinear constraints on the parameters. We have defined the prior distributions for the dependent variable and model parameters and then obtained the posterior distributions. Since the posterior distribution of the model parameters is nonlinear and involves nonlinear inequality constraints, the expected value of the distribution is obtained with Markov Chain Monte Carlo and Metropolis-Hastings Algorithm. We have conducted a Monte Carlo experiment to compare NICRR with least squares, ordinary ridge regression, and nonlinear inequality constrained least squares estimator. With this experiment we have investigated the effect of number of observations, number of variables, error variance, the degree of multicollinearity and different nonlinear restrictions on parameter estimates. The results of the experiment suggest that NICRR have generally smaller mse values then the compared estimators and leads to better estimates.

**Keywords:** Nonlinear inequality restrictions, Nonlinear inequality restricted least squares, Nonlinear inequality restricted ridge regression, Markov Chain Monte Carlo

## References

- [1] Liew, C.K. (1976), "Inequality constrained least squares estimation", *Journal of the American Statistical Association*, 71(355), p746-751.
- [2] Toker, S., Ustundag Siray, G., and Kaciranlar, S. (2013), "Inequality constrained ridge regression estimator", *Statistics and Probability Letters*, 83, 2391-2398.



## Explicit influence analysis for count data under AB–BA crossover trials

Chengcheng Hao<sup>1</sup>, Dietrich von Rosen<sup>2,3</sup> and Tatjana von Rosen<sup>1</sup>

<sup>1</sup>Stockholm University, Stockholm, Sweden

<sup>2</sup>Swedish University of Agricultural Sciences, Uppsala, Sweden

<sup>3</sup>Linköping University, Linköping, Sweden

### Abstract

Generalized linear mixed models (GLMMs) are commonly utilized in crossover studies with count data. The authors describe a method to assess influence of subjects in maximum likelihood estimates under GLMM-type crossover models. Various perturbations are performed on the observed data. By decomposing the GLMM-type crossover model, the authors achieve explicit expressions of the influence on the estimated treatment effect. The influence measure turns to relate to raw residuals of the subjects in the unperturbed model.

**Keywords:** Crossover design, Generalized linear mixed model, Perturbation scheme, Statistical diagnostic

## Small area estimation methods and their links to linear and linear mixed models

Stephen Haslett

Massey University, New Zealand

### Abstract

Small area estimation (sae) is a set of techniques used to provide sufficiently accurate estimates at a finer level than is possible using the data for each small area alone. Indeed small area estimates can be obtained even in areas in which there is no sample. Models for sae come in a variety of forms that extend beyond linear and mixed linear models, but this talk (rather than attempting to be encyclopaedic) will instead consider common themes and differences between sae using linear and linear mixed models, and the usual framework used for linear and linear mixed models. The intent is to provide context for the other talks in this session. Common themes will include estimation of fixed and prediction of random parameters. Differences will include use of sample survey data in sae, and the interest in aggregates of combinations of fixed and random parameter estimates, rather than in prediction of individual observations or estimation and prediction of individual parameters.

## On some applications of saddle-point matrices

Jan Hauke<sup>1</sup>, Charles R. Johnson<sup>2</sup> and Tadeusz Ostrowski<sup>3</sup>

<sup>1</sup>Adam Mickiewicz University, Poznań, Poland

<sup>2</sup>William and Mary Colledge, Williamsburg, USA

<sup>3</sup>The Jacob of Paradyż University of Applied Sciences, Gorzów Wlkp, Poland

### Abstract

For a given square matrix  $A \in R^{n \times n}$  and vector  $e \in R^{n-1}$  of ones let us denote by  $(A, e)$  the matrix (often called the saddle point matrix) defined as the block matrix  $(A, e) = \begin{pmatrix} A & e \\ e^T & 0 \end{pmatrix}$ . If the matrix  $A$  is symmetric, then  $(A, e)$  may be interpreted as the bordered Hessian of a standard quadratic program over the standard simplex, and it is called the Karush-Kuhn-Tucker matrix of the program, which is known to have a large spectrum of applications (see e.g. [1]). The saddle point matrices play significant role as well in other branches of mathematics. In the paper we show some applications of them in: game theory and analysis. An application of specific saddle point matrices which are hollow, symmetric, and nonnegative is likewise shown in geometry as a generalization of Heron's formula for presentation of the formula of  $n$ -dimensional irregular tetrahedron volume (an extension of results of [2]).

**Keywords:** Saddle point matrix, Volume of hypertetrahedron

### References

- [1] Bomze I. M. (1998). On standard quadratic optimization problems, *Journal of Global Optimization*, 13: 369–387.
- [2] Griffiths M.(2005).  $n$ -dimensional enrichment for Further Mathematicians, *The Mathematical Gazette*, 89: 409–416.

## High-dimensional CLTs for individual Mahalanobis distances

Thomas Holgersson<sup>1,2</sup> and Deliang Dai<sup>1</sup>

<sup>1</sup>Linnaeus University, Växjö, Sweden

<sup>2</sup>Jönköping University, Jönköping, Sweden

### Abstract

In this paper we derive central limit theorems for two different types of Mahalanobis distances in situations where the dimension of the parent variable increases proportionally with the sample size. It is shown that although the two estimators are closely related and behave similarly in finite dimensions, they have different convergence rates and are also centred at two different points in high-dimensional settings. The limiting distributions are shown to be valid under some general moment conditions and hence available in a wide range of applications.

**Keywords:** Mahalanobis distance, increasing dimension, weak convergence, Marcenko-Pastur distribution, outliers, Pearson family.

## Orthogonal regression among parts of compositional data

Klára Hružová<sup>1</sup>, Karel Hron<sup>1</sup>, Peter Filzmoser<sup>1,2</sup> and Valentin Todorov<sup>3</sup>

<sup>1</sup>Palacký University, Olomouc, Czech Republic

<sup>2</sup>Vienna University of Technology, Vienna, Austria

<sup>3</sup>United Nations Industrial Development Organization (UNIDO) - Vienna International Centre, Vienna, Austria

### Abstract

Compositional data, multivariate observations carrying only relative information (specially proportions or percentages), follow the Aitchison geometry on the simplex [1]. Therefore, it is not possible to apply standard statistical methods, that rely mostly on Euclidean geometry in real space, directly to compositions. For this purpose, isometric logratio (ilr) coordinates were formed to express compositional data in coordinates with respect to an orthonormal basis on the simplex [2]. Due to reduced dimension of compositions (one less than the number of components), interpretable coordinates are of primary interest. The aim of this contribution is to develop a concise methodology for orthogonal regression of compositional data in ilr coordinates, following the approach of [3]. The orthogonal regression itself is solved with singular value decomposition, applied to both response and explanatory variables [4]. Theoretical developments are used to analyze the relation between relative contributions of activities forming gross value added.

**Keywords:** isometric logratio transformations, orthogonal regression, singular value decomposition

### References

- [1] Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London. (Reprinted in 2003 with additional material by The Blackburn Press).
- [2] Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barceló-Vidal, C. (2003): Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35:279–300.
- [3] Hron, K., Filzmoser, P. and Thompson, K. (2012): Linear regression with compositional explanatory variables. *Journal of Applied Statistics*, 39/5:1115–1128.
- [4] Markovsky, I. and Van Huffel, S. (2007): Overview of total least-squares methods. *Signal Processing*, 87:2283–2302.

# Statistical Properties for Multilinear Principal Component Analysis

Su-Yun Huang<sup>1</sup>, Ting-Li Chen<sup>1</sup>, Hung Hung<sup>2</sup>, I-Ping Tu<sup>1</sup> and Pei-Shien Wu<sup>3</sup>

<sup>1</sup>Academia Sinica, Taipei, Taiwan

<sup>2</sup>National Taiwan University, Taipei, Taiwan

<sup>3</sup>Duke University, Durham, United States

## Abstract

Principal component analysis (PCA) is a simple and very popular method in statistical data analysis for dimension reduction. It can be done by eigenvalue decomposition of covariance matrix. Traditional PCA deals with vector variables and each observation is represented in a vector form. When observations are tensor objects, such as images, videos, EEG signals over a spatial domain or gene-gene interactions (as symmetric random matrices), traditional PCA first vectorizes these tensor objects and then proceeds with the eigenvalue decomposition of a large covariance matrix. This vectorized PCA for tensor data can be difficult and inefficient. The main reason is that the estimation process of PCA is unstable when the sample size is small compared to the dimension of the vectorized data. Multilinear principal component analysis (MPCA) [3, 4] is a modification of PCA. It preserves the natural tensor structure of observations in searching for principal components. The main advantage of preserving the tensor structure is the parsimonious usage of parameters in specifying the principal component subspaces, which mitigates the adverse influence of high-dimensionality, and hence, leads to efficiency gain in estimation and prediction [2]. The rationale for the success of MPCA will be presented from the statistical point of view and based on a real data application [1].

**Keywords:** Asymptotic analysis, Asymptotic efficiency, Dimension reduction, High-order singular value decomposition, Multilinear principal component analysis, Tensor decomposition.

## References

- [1] Chen, T.L., Hsieh, D.N., Hung, H., Tu, I.P., Wu, P.S., Wu, Y.M., Chang, W.H. and Huang, S.Y. (2014).  $\gamma$ -SUP: a clustering algorithm for cryo-electron microscopy images of asymmetric particles. *Annals of Applied Statistics*, 8:259-285.
- [2] Hung, H., Wu, P.S., Tu, I.P. and Huang, S.Y. (2012). On multilinear principal component analysis of order-two tensors. *Biometrika*, 99:569-583.
- [3] Kolda, T.G. and Bader, B.W. (2009). Tensor decompositions and applications. *SIAM Review*, 51:455-500.
- [4] Lu, H., Plataniotis, K.N. and Venetsanopoulos, A.N. (2008). MPCA: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks*, 19:18-39.

## Covariance Components Selection in High-Dimensional Growth Curve Model with Random Coefficients

Shinpei Imori<sup>1</sup> and Dietrich von Rosen<sup>2,3</sup>

<sup>1</sup>Hiroshima University, Japan

<sup>2</sup>Linköping university, Linköping, Sweden

<sup>3</sup>Swedish University of Agricultural Sciences, Uppsala, Sweden

### Abstract

The growth curve model (GCM or GMANOVA) is a traditional and suitable model for analyzing longitudinal data. Recently, we often encounter high-dimensional data, where the dimension  $p$  of the response variable is large in relation to the number of independent observations  $n$ . For such data, we need to specify parameter structures on the covariance matrices, since there are too many parameters included in the unstructured covariance matrices. The GCM with random coefficients may lead to covariance matrices for longitudinal responses with a reasonable number of parameters.

From the perspective of a model selection, the Bayesian Information Criterion (BIC) [3] and its generalization (GIC) [2] are often preferable in order to identify the true model because of their consistency, i.e., the selection probability of the true covariance components by the proposed criterion goes to 1 under some adequate conditions. The main aim of this paper is to propose a GIC-type criterion for selecting the true covariance components in the GCM and show consistency in a high-dimensional setting, where  $n$  is smaller than  $p$ . Our criterion is based on the results of [1], which proposed a test statistic for random coefficients covariance components in extended GCM.

**Keywords:** High-dimensional data analysis, Model selection.

### References

- [1] Fujikoshi, Y. and von Rosen, D. (2000). LR Tests for Random-Coefficient Covariance Structures in an Extended Growth Curve Model. *Journal of Multivariate Analysis*, 75, 245–268.
- [2] Nishii, R. (1984). Asymptotic Properties of Criteria for Selection of Variables in Multiple Regression. *Annals of Statistics*, 12, 758–765.
- [3] Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

## A Comparison of compound Poisson class distributions

Deniz Inan<sup>1</sup>, Öyküm Esra Askin<sup>2</sup> and Ali Hakan Buyuklu<sup>1,2</sup>

<sup>1</sup>Marmara University, Istanbul, Turkey

<sup>2</sup>Yildiz Technical University, Istanbul, Turkey

### Abstract

The present study is focused on compound Poisson class of time to failure distribution. Under the concept of population heterogeneity, Kus (2007) obtained a decreasing failure rate distribution by compounding exponential and zero truncated Poisson distributions (EP). With the similar mixing procedure, various types of compound distributions were proposed such as Weibull–Poisson, Lindley–Poisson and Rayleigh–Poisson. In this study, a comparison of properties of compound Poisson class distributions is given and fitting performances are investigated for a real data set. The Expectation and Maximization (EM) algorithm is performed to estimate the maximum likelihood estimators of four different model parameters.

**Keywords:** Exponential-Poisson distribution, Compound Poisson Class, EM algorithm

### References

- [1] Adamidis, K. and Loukas, S., (1998). A life time distribution with decreasing failure rate. *Statist. Probab. Lett.* 39, 35–42.
- [2] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist.Soc. Ser. B* 39, pp. 1–38.
- [3] Kus C.(2007). A new lifetime distribution, *Comput. Statist. Data Anal.* 51, pp. 4497–4509.
- [4] Ghitany, M. E., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its application. *Math. Comput. Simul.* 78, pp. 493–506.
- [5] Erisoglu U., Erisoglu M. and Calis N. (2013). Heterogeneous data modeling with two-component Weibull–Poisson distribution, *Journal of Applied Statistics*, 40:11, pp. 2451–2461.
- [6] L. Wanbo and S. Daimin (2012). A new compounding life distribution: The Weibull–Poisson distribution, *J. Appl. Stat.* 39(1) , pp. 21–38.
- [7] Barreto-Souza, W. and Cribari-Neto, F., (2009). A generalization of the exponential–Poisson distribution. *Statistics and Probability Letters* 79, pp. 2493–2500.

## ‘Hard’ versus ‘Soft’ Predictions from Unit-level Models for Small Area Estimation of Proportions

Geoff Jones and Penny Bilton  
Massey University, Palmerston North, New Zealand

### Abstract

Sometimes the variable of interest in small area estimation is the proportion of units in a small area or domain that possess a particular attribute. Examples are the proportion of households in poverty, or the proportion of children who are underweight. Given a sample of unit-level data comprising the target variable  $Y$  and a set of auxiliary variables  $X$ , together with unit-level population data for  $X$ , there may be benefits in producing unit-level predictions for the census units from a unit-level model fitted to the survey data, and then amalgamating these to small area level, as in the World Bank’s (ELL) poverty mapping methodology. When the target variable is binary, the unit-level predictions can either be binary (0/1) or probabilities in  $[0,1]$ . We term these ‘hard’ and ‘soft’ estimates respectively. We discuss the relative performance of these alternatives, particularly in the case where the modelling is done using classification trees.

**Keywords:** Poverty mapping, Classification trees

## Network Component Resolution and Cancer Landscapes

Alexandra Jauhiainen<sup>1</sup>, José Sánchez<sup>2</sup> and Rebecka Jörnsten<sup>2</sup>

<sup>1</sup>Karolinska Institutet, Stockholm, Sweden

<sup>2</sup>Chalmers and the University of Gothenburg, Göteborg, Sweden

### Abstract

The study of cancer at the molecular level has increased our understanding of these diseases in recent years. In particular, network modelling has become a popular method to construct mechanistic hypotheses and generate predictions of potential drug targets. However, as estimation uncertainty is commonly ignored, there is a great risk of over-interpreting these network models.

In fact, it is non-trivial to exhaustively assess uncertainty in network estimation. Bootstrap is sometimes used, in which resampling based marginal statistics on whether an edge is present or absent are recorded. Thresholding the individual link statistics, i.e. keeping links which appear in e.g. 90% of the bootstrap networks, may generate a graph that is not a member of the set of bootstrap graphs. The structure of the thresholded graph may thus not be supported by the empirical data (Cf. multivariate mean and median).

We present a novel analysis pipeline, NetCoR, applicable to any network estimation method. NetCoR builds on; bootstrap, robust module identification, data depth based graph representation, and rate-distortion theory. We exploit resampling bootstrap graphs, not to directly generate the joint high-dimensional edge distribution in graph space (which is complex), but rather to identify a bounded set of bootstrap graphs that capture a large proportion of variation over all observed graphs.

Our final result comprises a set of local network regions with a corresponding set of candidate graph structures. High-resolution network regions are characterized by one or a

small number of candidate graphs, while low-resolution network regions are characterized by a large number of alternative network structures. The allocation of a certain number of candidate graphs to each network region draws upon results from rate-distortion theory.

This paradigm has multiple benefits; (i) The local number of candidate networks captures the confidence in the estimated structure; (ii) This method provides a fair and efficient way to compare different estimation methods; and, (iii) We derive a global network resolution measure and show this is appropriate for selecting the sparsity level of the network, outperforming standard methods such as cross-validation and the Bayesian Information Criterion. We apply our method to cancer data from The Cancer Genome Atlas (TCGA) and discuss our findings.

We have developed an R package which implements the NetCoR paradigm from start to end and are currently working toward integrating NetCoR with CancerLandscapes.org. Time permitting, we will briefly discuss the CancerLandscapes portal and cancer comparative network models.

**Keywords:** Network estimation, Rate-distortion

## Partial least squares discriminant analysis for compositional data: logratio methodology and its application to metabolomics

Alžběta Kalivodová<sup>1</sup> and Karel Hron<sup>1</sup> and Peter Filzmoser<sup>2</sup>

<sup>1</sup>Palacký University Olomouc, Czech Republic

<sup>2</sup>Vienna University of Technology, Austria

### Abstract

Results of metabolological measurements are often expressed as observations carrying only relative information, called compositional data. They consist of positive vectors, where the only relevant information is contained in ratios between their parts [1]. The partial least squares discriminant analysis (PLS-DA) is popular method in chemometrics used for classification of high-dimensional data [2], based on singular value decomposition of both the response and covariates. In order to enable processing of compositional data using PLS-DA, logratio methodology, leading to new coordinates with respect to natural geometric features of compositions, needs to be applied. Theoretical considerations, forming the main aim of the contribution, will be applied to real-world data from metabolomics. The data set, resulting from targeted metabolomics, is used to analyze significance of the corresponding regression parameters (metabolites) in order to highlight possible markers of inherited metabolic disorders. The significance of metabolites is investigated by applying PLS-DA, adjusted for dealing with compositional data. The presented experiment confirms a more general experience that significance of important metabolites (markers of diseases), resulting from a bootstrap procedure, is usually better visible using the logratio approach [3].

**Keywords:** compositional data, metabolomics, partial least squares discriminant analysis, significance testing



## References

- [1] Pawlowsky-Glahn, V., Buccianti, A. (2011). *Compositional data analysis: Theory and applications*. Wiley, Chichester.
- [2] Varmuza, K., Filzmoser, P. (2009). *Introduction to multivariate statistical analysis in chemometrics*. Taylor & Francis, New York.
- [3] Kalivodová, A., Hron, K., Filzmoser, P. Najdekr, L., Janečková, H., Adam, T. (2014). PLS-DA for compositional data with application to metabolomics. *Journal of Chemometrics*, accepted for publication.

## Explicit Estimators for a Banded Covariance Matrix in a Multivariate Normal Distribution

**Emil Karlsson** and Martin Singull  
Linköping University, Linköping, Sweden

### Abstract

The problem of estimating mean and covariances of a multivariate normal distributed random vector has been studied in many forms. This talk focuses on the estimators proposed in [1] for a banded covariance structure with  $m$ -dependence. It presents the previous results of the estimator and rewrites the estimator when  $m = 1$ , thus making it easier to analyze. This leads to an adjustment, and a proposition for an unbiased estimator can be presented. A new and easier proof of consistency is then presented.

**Keywords:** Banded covariance matrices, Covariance matrix estimation, Explicit estimators, Multivariate normal distribution

### References

- [1] Ohlson, M., Andrushchenko, Z. and von Rosen, D. (2011). Explicit estimators under  $m$ -dependence for a multivariate normal distribution. *Annals of the Institute of Statistical Mathematics*, 63(1):29–42.

## Multivariate data with block compound symmetry covariance structure

Daniel Klein<sup>1</sup>, Anuradha Roy<sup>2</sup>, Ricardo Leiva<sup>3</sup> and Ivan Žežula<sup>1</sup>

<sup>1</sup>P. J. Šafárik University, Košice, Slovakia

<sup>2</sup>The University of Texas at San Antonio, San Antonio, USA

<sup>3</sup>Universidad Nacional de Cuyo, Mendoza, Argentina

### Abstract

Hotelling's  $T^2$  test is the conventional method to test the equality of mean vectors in two populations. However, Hotelling's  $T^2$  statistic is based on the unbiased estimate of the unstructured variance-covariance matrix. Nevertheless, the variance-covariance matrix may have some structure, and one should use an unbiased estimate of that structure to test the equality of mean vectors. A natural extension of the Hotelling's  $T^2$  statistic is obtained for testing the mean vectors for doubly multivariate observations with  $q$  response variables and  $p$  sites in blocked compound symmetric covariance matrix setting.

**Keywords:** Blocked compound symmetric covariance structure, Doubly multivariate data, Hotelling's  $T^2$  statistic.

### References

- [1] Geisser, S. (1963). Multivariate Analysis of Variance for a Special Covariance Case. *Journal of the American Statistical Association* 58(303): 660-669.
- [2] Roy, A., and Fonseca, M. (2012), Linear Models with Doubly Exchangeable Distributed Errors. *Communications in Statistics - Theory and Methods*, 41(13): 2545-2569.
- [3] Szatrowski, T. H. (1982). Testing and Estimation in the Block Compound Symmetry Problem. *Journal of Educational Statistics* 7(1): 3-18.

## Statistical inferences for market network analysis

Petr Koldanov

National Research University Higher School of Economics, Nizhny Novgorod, Russia

### Abstract

Network model of financial market attracts a growing attention last decades. In this model each stock corresponds to a vertex and a weighted link between two vertices is given by a measure of similarity of corresponding returns. The obtained network is a complete weighted graph. Market network analysis is an investigation of different characteristics of this graph. Most popular characteristics are minimum spanning tree, planar maximally filtered graph [1], market graph, maximum cliques and maximum independent sets [2].

Existing publications in the field does not pay attention to stochastic nature of financial data. As a consequence the statistical uncertainty of obtained results is out of control. In the present paper a general approach to handle this problem is proposed.

This approach is based on statistical multiple decision theory. It is shown that network characteristics in Pearson correlation network are not robust. In contrast stable measure of similarity in the class of elliptically contoured distribution is proposed. This measure is based on probability of signs coincidence of random variables. Optimal statistical procedures for network characteristic identification are proposed. The statistical uncertainty of some popular network structures is investigated and compared for different financial markets. Some results in this direction are presented in [3, 4].

**Keywords:** market network analysis, statistical inference.

## References

- [1] Tumminello M., Aste T., Matteo T.D., Mantegna R.N.(2005) A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*. 102 (30), 10421–10426.
- [2] Boginski V., Butenko S., Pardalos P.M.(2005) Statistical analysis of financial networks. *J. Computational Statistics and Data Analysis*. 48 (2), 431–443.
- [3] Koldanov A.P., Koldanov P.A., Kalyagin V.A., Pardalos P.M.(2013) Statistical procedures for the market graph construction.: *Computational Statistics and Data Analysis* 68 17–29.
- [4] Bautin G.A., Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M.(2013) Simple measure of similarity for the market graph construction *Computational Management Science*, 10, 105-124.

## Fractal and Antifractal Oxymorons, Moebius Strip Like Transformations of Biomedical Data as Basis for Exploratory Subgroup Analysis

Sergii Kulishov<sup>1</sup> and Olexandr Iakovenko<sup>2</sup>

<sup>1</sup>Higher medical educational institution of Ukraine "Ukrainian medical stomatological academy", Poltava, Ukraine

<sup>2</sup>V.P. Komissarenko Institute of Endocrinology and Metabolism AMS Ukraine, Kyiv, Ukraine

## Abstract

Proposed and tested an algorithm of using fractal and antifractal oxymorons, Moebius strip like transformations of biomedical data for exploratory subgroup analysis. The algorithm is reduced to initialization of study objects with fractal and antifractal data, Moebius strip like structure; formation of categorical variabilities that consist from informative numeric variabilities as sum of progradient and antigradient data, including similar to "superior and inferior surfaces of strip", by iteration process as for receiving fractal and antifractal sets; statistical analysis of categorical variabilities and dependent numeric variabilities, using parametric and nonparametric methods; formulation of the conclusion. Our algorithm of using fractal and antifractal oxymorons, Moebius strip

like transformations of biomedical data for exploratory analysis will help to uncover “re-entry” mechanisms of pathology, principles of prophylaxis and treatment chronic aging-dependent diseases.

**Keywords:** Antifractal, Fractal, Exploratory analysis, Moebius strip

## Restricted ridge estimation of generalized linear models

**Fikriye Kurtoğlu** and M. Revan Özkale  
Çukurova University, Adana, Turkey

### Abstract

Multicollinearity has long been recognized as a problem in estimating the parameters in linear regression models producing least squares estimates that are too large in absolute value and resulting in large variance-covariance matrix for the least squares estimator. The adverse effects of multicollinearity on parameter estimation in generalized linear models are also explored by various authors [1,4] in the case of maximum likelihood estimation. Segerstedt [4] proposed the ordinary ridge regression estimator and Nyquist [2] considered the restricted estimator of parameters in generalized linear models in the presence of multicollinearity.

In this study, we introduce restricted ridge estimator by unifying the ordinary ridge regression estimator and the restricted estimator in generalized linear models. The superiority of the restricted ridge estimator over the maximum likelihood and restricted estimators is considered according to the mean squared error matrix criterion. The results are illustrated by conducting a simulation study.

**Keywords:** Restricted ridge, Generalized linear model, Multicollinearity, Mean squared error

### References

- [1] Mackinnon, M.J. and Puterman, M.L. (1989). Collinearity in generalized linear models. *Commun. Statist. Theory Meth.*, 18(9):3463-3472.
- [2] Nyquist, H. (1991). Restricted estimation of generalized linear models. *Journal of the Royal Statistical Society Series C(Applied Statistics)*, 40(1):133-141.
- [3] Özkale, M.R. (2014). The relative efficiency of the restricted estimators in linear regression models. *Journal of Applied Statistics*, 41(5):998-1027.
- [4] Segerstedt, B. (1992). On ordinary ridge regression in generalized linear models. *Commun. Statist. Theory Meth.*, 21(8):2227-2246.

## A two-step model for linear prediction with connections to PLS

Ying Li<sup>1</sup> and Dietrich von Rosen<sup>1,2</sup>

<sup>1</sup>Swedish University of Agricultural Sciences, Uppsala, Sweden

<sup>2</sup>Linköping university, Linköping, Sweden

### Abstract

In this article, we consider the regression of a univariate response on some random predictors with the general goal of inferring a future observation. The data have a near-collinear structure and additionally group effects are assumed to exist. A two-step estimation procedure is proposed. The first step is to summarize the information in the predictors via a bilinear model. The bilinear model has a Krylov structured within individual design matrix, which is the link to classical partial least squares (PLS) analysis and a between individual design matrix which handles the group effects. The second step is the prediction step using conditional expectation approach. The two step approach gives us new insight in understanding PLS. Explicit maximum likelihood estimator of the dispersion matrix and mean for the predictors are derived under the assumption that the covariance between the response and explanatory variable is known, it is shown that the mean square error of the two step approach is always smaller than PLS.

**Keywords:** A two-step method, PLS, MLE, Krylov space, Growth curve model

### References

- [1] Ying Li and Dietrich von Rosen, (2012), Maximum likelihood estimators in a two step model for PLS. *Communications in Statistics - Theory and Methods*, **41**, 2503-2511.
- [2] Ying Li, Peter Udén and Dietrich von Rosen, (2013), A two-step PLS-inspired method for linear prediction with group effect. *Sankhyā A*, **75**, 96-117.

## Hypothesis testing in multilevel models with block circular covariance structures

Yuli Liang<sup>1</sup>, Dietrich von Rosen<sup>2,3</sup> and Tatjana von Rosen<sup>1</sup>

<sup>1</sup>Stockholm University, Stockholm, Sweden

<sup>2</sup>Swedish University of Agricultural Science, Uppsala, Sweden

<sup>3</sup>Linköping university, Linköping, Sweden

### Abstract

The results concerning hypothesis testing in multilevel models with block circular covariance structures will be presented.

Hypotheses about a general block structure (external tests) of the covariance matrix and specific variance components (internal tests) are of interest. The corresponding likelihood ratio statistics are derived and their distributions are studied. Equivalent hypotheses under specific restricted models are also discussed.

**Keywords:** canonical reduction, equivalent hypotheses, patterned blocks, variance components

## References

- [1] Crainiceanu, C. M. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66, 165–185.
- [2] Liang, Y. L., von Rosen, D., and von Rosen, T. (2014). On estimation in hierarchical models with block circular covariance structures. *Annals of the Institute of Statistical Mathematics*, accepted.
- [3] Olkin, I. (1973). Testing and estimation for structures which are circularly symmetric in blocks. In *Proceedings of the Symposium on Multivariate Analysis*. 183–195. Dalhousie, Nova Scotia.
- [4] Szatrowski, T. H. (1980). Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *The Annals of Statistics*, 8, 802–810.
- [5] Szatrowski, T. H., Miller, J. J. (1980). Explicit maximum likelihood estimates from balanced data in the mixed model of the analysis of variance. *The Annals of Statistics*, 8, 811–819.

## Multivariate skew-normal linear mixed models for multi-outcome longitudinal data

Tsung-I Lin<sup>1</sup> and Wan-Lun Wang<sup>2</sup>

<sup>1</sup>National Chung Hsing University, Taichung, Taiwan

<sup>2</sup>Feng Chia University, Taichung, Taiwan

### Abstract

More than one series of longitudinal data frequently encountered in biomedical, psychological and clinical research are routinely analyzed under a multivariate linear mixed model framework with underlying multivariate normality assumptions of the random effects and within-subject errors. However, such a normality assumption might not offer robust inference if the data, even after being transformed, particularly exhibit skewness. In this paper, we propose a multivariate skew-normal linear mixed model constructed by assuming a multivariate skew-normal distribution for the random effects and a multivariate normal distribution for the random errors. A damped exponential correlation structure is adopted to address the within-subject autocorrelation possibly existing among irregularly observed measures. We present an efficient alternating expectation-conditional maximization (AECM) algorithm for maximum likelihood estimation of parameters. The techniques for estimation of random effects and prediction of future outcomes in this model are discussed. Our proposed model is motivated by, and used for, the analysis of AIDS clinical trials in which we investigate the “association-of-the-evolutions” and the “evolution-of-the-association” of HIV-1 RAN copies and CD4<sup>+</sup> T cell counts during antiviral therapies.

**Keywords:** AECM algorithm, AIDS study, Asymmetric random effects, Damped exponential correlation, Skewness

## A Supervised Topic Models for Textual Data using a Multivariate probit model

Måns Magnusson, Leif Jonsson and Mattias Villani  
Linköping university, Linköping, Sweden

### Abstract

We propose a new topic model for text documents with a categorical label. Each document in a topic model is seen as a mixture of topics, and each topic is a probability distribution over a vocabulary of words. The topic structure of a document together with other features of the document determines a probability distribution over a categorical label. We assume that the categories are dependent through a multinomial probit model with a potentially sparse covariance structure. The motivation for this model comes from the problem of identifying a faulty program module in a computer systems using users' textual bug reports and prior knowledge about relationships between the program modules. We develop a MCMC scheme for sampling from the joint posterior distribution of the parameters in the topic model and the multinomial probit model.

**Keywords:** Topic model, Multinomial probit, Latent dirichlet allocation

## On sufficiency of quadraticly sufficient statistics in possibly mixed model

Augustyn Markiewicz<sup>1,2,\*</sup> and Barbora Arendacká<sup>3</sup>

<sup>1</sup>Poznań University of Life Sciences, Poznań, Poland

<sup>2</sup>Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland

<sup>3</sup>Physikalisch-Technische Bundesanstalt, Berlin, Germany

### Abstract

The notion of quadratic sufficiency was introduced and characterized in [1] in the context of fixed linear model. It was proved there that under normality quadraticly sufficient statistic is also sufficient. In the paper it is studied the problem of estimation in possibly misspecified model; i.e. when some effects assumed to be fixed are random. It is shown that quadraticly sufficient statistic under fixed model it is sufficient under respective mixed linear normal model. The results are applied to provide data reduction in factorial experiment.

**Keywords:** Linear sufficiency, quadratic sufficiency, mixed model

### References

- [1] Mueller, J. (1987). Sufficiency and completeness in the linear model. *Journal of Multivariate Analysis*, 21:312-323.

\*This research has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 283496, project transPLANT (<http://www.transplantdb.eu>), coordinated by EMBLEBI.

## A Functional Hodrick Prescott Filter

Hiba Nassar<sup>1</sup>, Boualem Djehiche<sup>2</sup> and Astrid Hilbert<sup>1</sup>

<sup>1</sup>Linnaeus University, Växjö, Sweden

<sup>2</sup>The Royal Institute of Technology, Stockholm, Sweden

### Abstract

We propose a functional version of the Hodrick-Prescott filter for functional data which take values in an infinite dimensional separable Hilbert space. We further characterize the associated optimal smoothing parameter when the associated linear operator is compact and the underlying distribution of the data is Gaussian. Moreover, we study a version of the functional Hodrick-Prescott filter where the associated operator is not necessarily compact, but merely closed and densely defined with closed range. We show that the associated optimal smoothing operator preserves the structure obtained in the compact case, when the underlying distribution of the data is Gaussian.

**Keywords:** Inverse problems, adaptive estimation, Hodrick-Prescott filter, smoothing, trend extraction, Gaussian measures on a Hilbert space.

## Small Area Estimation for multivariate repeated measures data

Innocent Ngaruye

Linköping University, Linköping, Sweden

### Abstract

The problem of how to produce reliable estimates of characteristics for small areas based on small samples taken from these areas has been of concern of researchers in recent years.

In this presentation, we consider a multivariate linear regression model for repeated measures data such as the Growth Curve Model in Small Area Estimation settings to get a model which borrows strength across small areas and over time, by incorporating simultaneously the effects of areas and time interaction. This model accounts for repeated surveys, group individuals and random effects variations. The model allows to find the small area means at each time point, per group units and particularly the pattern of changes or mean growth profiles over time. The estimation of model parameters is discussed with a likelihood based approach and explicit maximum estimators are obtained. Finally, a simulation study is conducted.



## Estimation in the multivariate linear normal models with linearly structured covariance matrices

**Joseph Nzabanita**

University of Rwanda, College of Science and Technology, Kigali, Rwanda and  
Linköping University, Linköping, Sweden

### Abstract

This paper considers the multivariate linear normal model (also known as the extended growth curve model) with a linearly structured covariance matrix. The main theme is to generalize the results in [2] and [1]. The idea is first used in [2] and is based on the study of residuals. A proper decomposition of the residual space, the orthogonal complement to the design space, into  $m + 1$  orthogonal subspaces and a study of residuals obtained from projections of observations on these subspaces yields explicit consistent estimators of the covariance matrix. An explicit consistent estimator of the mean is also proposed.

**Keywords:** multivariate linear normal model, extended growth curve model, estimation, linearly structured covariance matrix, residuals.

### References

- [1] Nzabanita, J., von Rosen, D. and Singull, M. (2012). Estimation of parameters in the extended growth curve model with a linearly structured covariance matrix. *Acta et Commentationes Universitatis Tartuensis de Mathematica* Volume 16(1):13–32
- [2] Ohlson, M. and von Rosen, D. (2010). Explicit estimators of parameters in the Growth Curve model with linearly structured covariance matrices. *Journal of Multivariate Analysis* 101:1284–1295.

## Repeated Latin and Youden squares with whole plot and subplot control treatments

**Kazuhiro Ozawa**<sup>1</sup>, Iwona Mejza<sup>2</sup>, Stanisław Mejza<sup>2</sup> and Shinji Kuriki<sup>3</sup>

<sup>1</sup>Gifu College of Nursing, Gifu, Japan

<sup>2</sup>Poznań University of Life Sciences, Poznań, Poland

<sup>3</sup>Osaka Prefecture University, Osaka, Japan

### Abstract

We consider a nested row-column design with split units for a two-factor experiment. The experimental material is divided into  $b$  blocks. Each block constitutes a row-column design with  $k_1$  rows and  $k_2$  columns. The units of the row-column design are treated as the whole plots, and the levels of a factor  $A$  (called whole plot treatments) are arranged on the whole plots. Additionally, each whole plot is divided into  $k_3$  subplots and the levels of the second factor  $B$  (called subplot treatments) are arranged on the subplots. We assume that the factors  $A$  and  $B$  have a whole plot control treatment and a subplot control treatment, respectively, and the remaining levels are test treatments.

Kachlicka and Mejza [1] presented a mixed linear model for the observations with fixed treatment combination effects and random block, row, column, whole plot and subplot effects, and the mixed model results from a four-step randomization. This kind of randomization leads us to an experiment with an orthogonal block structure, and the multistratum analysis can be applied to the analysis of data in the experiment.

Kachlicka and Mejza [2] and Mejza and Kuriki [3] used Youden squares supplemented by control treatments repeatedly for the whole plot treatments. In this talk, we consider a nested row-column design with split units in which the whole plot control treatment occurs the equal numbers of times in each row and in each column within each block, and the subplot control treatment occurs the equal number of times in each whole plot. A Latin square and a Youden square are used repeatedly for the whole plot test treatments, and a proper block design is used for the subplot test treatments. We give the stratum efficiency factors for such a nested row-column design, which has the general balance property.

**Keywords:** Control treatment, General balance, Latin square, Stratum efficiency factor, Test treatment, Youden square

### References

- [1] Kachlicka, D. and Mejza, S. (1996). Repeated row-column designs with split units. *Computational Statistics & Data Analysis*, 21:293–305.
- [2] Kachlicka, D. and Mejza, S. (2003). Whole-plot control treatments in Youden square with split units. *Colloquium Biometricum*, 33A:77–84.
- [3] Mejza, S. and Kuriki, S. (2013). Youden Square with split units, J.L. da Silva et al. (eds.), *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications*, Studies in Theoretical and Applied Statistics, Springer-Verlag, Berlin Heidelberg, 3-10.

## Testing Equality of Scale Parameters of Two Weibull Distributions in the Presence of Unequal Shape Parameters

Khurshid Alam and **Sudhir Paul**  
 University of Windsor, Canada

### Abstract

Data in the form of survival times arise in many fields of studies such as engineering, manufacturing, aeronautics and bio-medical sciences. A popular model for survival data is the two parameter Weibull distribution. Often lifetime or survival time data that are collected in the form of two samples are assumed to have come from two independent Weibull populations with different shape and scale parameters. In such a situation it may be of interest to test the equality of the scale parameters with the shape parameters being unspecified. This is equivalent to testing the equality of the location parameters with the shape parameters being unspecified in two extreme value distributions. Also, this is analogous to the traditional Behrens-Fisher problem of testing the equality of the means  $\mu_1$  and  $\mu_2$  of two normal populations where the variances  $\sigma_1^2$  and  $\sigma_2^2$  are unknown. We develop four test procedures, namely, a likelihood ratio test, a  $C(\alpha)$  test based on

the maximum likelihood estimates of the nuisance parameters, a  $C(\alpha)$  test based on the method of moments estimates of the nuisance parameters by Cran (1988) and a  $C(\alpha)$  test based on the method of moments estimates of the nuisance parameters by Teimouri and Gupta (2013). These test statistics are then compared, in terms of empirical size and power, using a simulation study.

## Hypothesis testing in variance components with constrains

**Marcin Przystalski**<sup>1</sup> and Miguel Fonseca<sup>2</sup>

<sup>1</sup>The Research Center for Cultivar Testing, Słupia Wielka, Poland

<sup>2</sup>Universidade Nova de Lisboa, Lisboa, Portugal

### Abstract

When testing hypothesis for variance components in random and mixed models using likelihood ratio tests, the positivity of these parameters is often overlooked. Using the results obtained in [1], hypothesis tests with positivity constrains based on approximate distributions and permutation procedures for random models with 2 and 3 variance components has been proposed. The performance of these procedures is investigated through simulation studies.

**Keywords:** inference with constraints, mixed models, permutation test procedure

### References

- [1] Self S. G., and Liang K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*,82:605-610.

## It's 30 years ...

**Simo Puntanen**

University of Tampere, Tampere, Finland

### Abstract



**Keywords:** Poles, Finns, Swedes

## Speeding Up MCMC by Efficient Data Subsampling

**Matias Quiroz**<sup>1</sup>, **Mattias Villani**<sup>2</sup> and **Robert Kohn**<sup>3</sup>

<sup>1</sup>Stockholm University and Sveriges Riksbank, Sweden

<sup>2</sup>Linköping University, Sweden

<sup>3</sup>University of New South Wales, Sydney, Australia

### Abstract

The computing time for Markov Chain Monte Carlo (MCMC) algorithms can be prohibitively large for datasets with many observations, especially when the data density for each observation is costly to evaluate. We propose a framework based on a Pseudo-marginal MCMC where the likelihood function is unbiasedly estimated from a random subset of the data, resulting in substantially fewer density evaluations. The subsets are selected using efficient sampling schemes, such as Probability Proportional-to-Size (PPS) sampling where the inclusion probability of an observation is proportional to an approximation of its contribution to the likelihood function. We illustrate the method on a large dataset of Swedish firms containing half a million observations.

**Keywords:** Bayesian inference, Markov Chain Monte Carlo, Pseudo-marginal MCMC, Big Data, Unbiased likelihood estimation, Probability Proportional-to-Size sampling.

## A New Method to Compare Statistical Growth Curves: The PL-GMANOVA Model and its Application

Martin Ricker<sup>1</sup>, Víctor Peña Ramírez<sup>2</sup> and Dietrich von Rosen<sup>3</sup>

<sup>1</sup>Instituto de Biología, Universidad Nacional Autónoma de México (UNAM), Mexico City

<sup>2</sup>Posgrado en Ciencias Biológicas, UNAM, Mexico City

<sup>3</sup>Swedish University of Agricultural Sciences, Uppsala, Sweden

### Abstract

Growth curves are monotonically increasing functions that measure repeatedly the same subjects over time. The classical growth curve model in the statistical literature is the Generalized Multivariate Analysis of Variance (GMANOVA) model. This model is combined here with the linear regression model  $\hat{Q} = \hat{A} \cdot T + E$ , where  $\hat{Q} = Ei(-\hat{b} \cdot r) - Ei(-\hat{b} \cdot \hat{r}_1)$ ,  $\hat{\cdot}$  = estimated parameter,  $Ei(x) = \int e^x/x dx$ ,  $b = -1/(\text{turning point radius in a sigmoid curve})$ ,  $r = \text{tree trunk radius}$ ,  $\hat{A} = \text{estimated initial relative growth}$ ,  $T = (t - t_1)$ ,  $t = \text{time}$ ,  $\hat{r}_1$  at  $t_1$  is an estimated calibrating time-radius point, and  $E$  is an error term for each tree and time point. Advantages of the approach are that growth rates can be compared among growth curves with different turning point quantities and different starting points, hidden outliers are easily detectable, the method is statistically robust, and autocorrelation and heteroscedasticity of the residuals are allowed. The model was implemented with dendrochronological data of 235 *Pinus montezumae* trees on ten Mexican volcano sites to calculate comparison intervals for  $\hat{A}$ , as well as variance components: 97% of the growth variation was found among trees within sites, and only 3% among sites. One site (at the Popocatépetl volcano) stood out with  $\hat{A}$  being 3.5 times the value of the site with the slowest-growing trees.

**Keywords:** GMANOVA, growth curves

### References

- [1] Kollo, T., von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*. Springer, Dordrecht, The Netherlands. 489 p.
- [2] Potthoff, R.F., and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51(3 and 4): 313-326.
- [3] Ricker, M., and del Río, R. (2004). Projecting diameter growth in tropical trees: A new modeling approach. *Forest Science*, 50(2):213-224.

## Robust singular value decomposition with application to multi-location plant breeding trials

Paulo Canas Rodrigues<sup>1,2</sup>, Andreia Monteiro<sup>3</sup> and Vanda Lourenço<sup>2,3</sup>

<sup>1</sup>Federal University of Bahia, Salvador, Brazil

<sup>2</sup>CMA-FCT-UNL, Caparica, Portugal

<sup>3</sup>Nova University of Lisbon, Caparica, Portugal

### Abstract

The distribution of continuous variables is usually not normal, often showing heavy tails. Therefore, in such scenarios, the classical approach whose likelihood-based inference leans on the normality assumption may be inappropriate, having low statistical efficiency.

Robust statistical methods [2] are designed to accommodate for certain data deficiencies, allowing for reliable results under various conditions. They ought to be resistant to influent factors as outlying observations, non-normality and other model misspecifications. Moreover, if the model verifies the classical assumptions, robust methods provide results close to the classical ones.

Rodrigues et al. (2014) [3] proposed a generalization of the additive main effects and multiplicative interaction (AMMI) model [1] which is able to model two-way data tables where different columns and/or rows and/or cells might have different weights, with application to statistical genetics. In this paper a new methodology where robust statistical methods replace the classic ones to model, structure and analyse genotype-by-environment interactions, in the context of multi-location plant breeding trials, is presented. In particular, interest resides in the development of a robust version of the AMMI model and the comparison between its performance and the performance of the classic AMMI model. This is achieved through Monte Carlo simulations where various contamination schemes are considered.

**Keywords:** AMMI model, Robust Statistics, Singular Value Decomposition, Statistical Genetics

### References

- [1] Gauch, H.G. (1992). Statistical analysis of regional yield trials: AMMI analysis of factorial designs, Amsterdam, Elsevier.
- [2] Maronna, R.A., Martin, R.D., and Yohai, V.J. (2006). Robust statistics, theory and methods, Chichester, Wiley.
- [3] Rodrigues, P.C., Malosetti, M., Gauch, H.G., and van Eeuwijk, F.A. (2014). Weighted AMMI to study genotype-by-environment interaction and QTL-by-environment interaction. *Crop Science* (in press)

## Model of Fatigue Failure Due to Multiple Cracks using Extended Birnbaum-Saunders Distribution

Anuradha Roy<sup>1</sup>, Ricardo Leiva<sup>2</sup>, Rubén Bageta<sup>3</sup> and Juan Carlos Pina<sup>4,5</sup>

<sup>1</sup>The University of Texas at San Antonio, San Antonio, USA

<sup>2</sup>F.C.E., Universidad Nacional de Cuyo, Mendoza, Argentina

<sup>3</sup>F.C.A. and I.C.B., Universidad Nacional de Cuyo, Mendoza, Argentina

<sup>4</sup>Materials innovation institute M2i, The Netherlands

<sup>5</sup>Eindhoven University of Technology, The Netherlands

### Abstract

In this article we propose an extension of the Birnbaum-Saunders (BS) distribution to model the life time of a material due to the growth of  $m$  cracks. See [1] for Extended Birnbaum-Saunders (EBS) distribution. Our new EBS distribution for  $m$  number of cracks has parameters  $\mathbf{A}$  and  $\boldsymbol{\beta}$  as the shape matrix and the scale vector respectively, as compared to the scalar quantities  $\alpha$  and  $\beta$  as the shape and the scale parameters of the classical BS distribution. We show through many remarks that for  $m = 1$  our new EBS distribution reduces to the conventional BS distribution. We also show with the help of simulation study that our EBS is a better model than the BS model in the sense of parameter estimation (more precision and less standard deviation) when we have multiple crack information in the model. It is shown that the hazard function of the EBS distribution can be expressed as a sum of  $m$  hazard functions, and that each of these  $m$  hazard functions is an upside down function for all shape matrices  $\mathbf{A}$  and scale vectors  $\boldsymbol{\beta}$ . Finally, we observe an interesting connection between our EBS distribution and Balakrishnan skew normal distribution (proposed by Balakrishnan as a discussant of [2]).

**Keywords:** Extended Birnbaum-Saunders distribution, Hazard functions, Skew normal distribution

### References

- [1] Leiva R., Roy A., Bageta R., and Pina J.C. (2014). An Extension of the Birnbaum-Saunders Distribution as a Model for Fatigue Failure due to Multiple Cracks, *Journal of Statistical Theory and Practice*, DOI:10.1080/15598608.2014.910149.
- [2] Arnold, B.C. and Beaver, R.J. (2002). Skewed multivariate models related to hidden truncation and / or selective reporting, *Test*, 11(1), 7-54.

## Structure learning for improved classification accuracy for high-dimensional omics data

Patrik Rydén<sup>1</sup>, Rafael Björk<sup>1</sup>, and Tatjana Pavlenko<sup>2</sup>

<sup>1</sup> Umeå University, Sweden

<sup>2</sup> Royal Institute of Technology (KTH), Sweden

### Abstract

High throughput omics technologies in life science such as next generation sequencing, array technologies and mass spectrometry have allowed genome-wide measurements of complex cellular responses for a broad range of treatments and diseases. The technologies are powerful, but in order to utilize their full potential new statistical tools are required. We consider a biological system that can be represented by a graph with disjoint modules, where nodes represent variables, and edges represent variable interactions. Commonly, the system is observed under several conditions (e.g. T-cells are observed on diseased and healthy individuals) and the aim with the study may be threefold: network analysis (i.e. estimate the complete graph), cluster analysis (i.e. determine the modules of the graph) and to build a classifier that enables us to predict the class membership of an individual. For the problem when the data are multivariate normal distributed with few connections between the modules the precision matrix can be estimated using the graphical lasso. Here we introduce techniques that based on the estimated precision matrix identifies the modules of the graph and show how the estimated modular structure can be used to improve the accuracy of the linear discriminant analysis (LDA) classifier. Our method combines graphical lasso, bootstrap, distance sensitive ordering and dynamical programming. The suggested algorithms were evaluated in a large scale simulation study and on real microarray data. Compared to the usual LDA-classifier the suggested block-LDA classifier had considerably lower misclassification rate. The algorithms are fully atomized and demands a minimum input from the user witch make them very useful and user friendly. Furthermore, we believe that our approach is general and can be used to improve a wide range of classifiers, not just the LDA-classifier.

**Keywords:** High-dimensional data, Classification

## Rotating a Linear Mixed Model for Multiple Testing in Genome-wide Association Studies having Repeated Measurements

Lars Rönnegård

Swedish University of Agricultural Sciences, Uppsala, Sweden, and  
Dalarna University, Falun, Sweden

### Abstract

In genome-wide association studies (GWAS) a large number of multiple tests are performed. Typically a linear model is fitted to each genetic marker in 10,000 or more positions along the genome where the response vector includes the values of the individuals' phenotype  $\mathbf{y}$  and the genotype of the marker (a single-nucleotide polymorphism, SNP) is included as a covariate  $\mathbf{x}_{snp}$ . The covariate values changes at each position, whereas  $\mathbf{y}$  does not change, and a Wald test is used to produce a P-value for the SNP



effect at each position. The significance level is adjusted through Bonferroni correction. A complication, however, is that the observations are often correlated because related individuals have been sampled. If the dependency is not modelled, the P-values will be inflated throughout the entire genome producing a large number of false positives.

Modelling the dependency between individuals in GWAS using a linear mixed model (LMM) has previously been thoroughly investigated. However, in the application motivating the current study, there are both related individuals and repeated observations on each individual. Here the phenotype  $\mathbf{y}$  is clutch size from 311 female flycatchers. Both the dependency from repeated observations and individuals being related can be modelled using a LMM, but fitting 1000s of LMM would be too time consuming. Hence, a fast method with a minor approximation is suggested.

The LMM to be fitted at each position is  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{x}_{snp}\boldsymbol{\beta}_{snp} + \mathbf{Z}\mathbf{g} + \mathbf{W}\mathbf{p} + \mathbf{e}$  where  $\mathbf{X}$  is the design matrix for any non-genetic fixed effects (age and year) and  $\boldsymbol{\beta}$  are the corresponding fixed effects,  $\mathbf{x}_{snp}$  is the SNP-covariate (coded 0, 1, 2) and  $\boldsymbol{\beta}_{snp}$  is the SNP effect. The model includes a random polygenic effect  $\mathbf{g}$  and a permanent environmental effect  $\mathbf{p}$  for each individual due to repeated measurements. The random effects and the residuals  $\mathbf{e}$  are assumed Gaussian:  $\mathbf{g} \sim N(0, \mathbf{K}\sigma_g^2)$ ,  $\mathbf{p} \sim N(0, \mathbf{I}\sigma_p^2)$  and  $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ . Here  $\mathbf{K}$  is a correlation matrix computed using all available markers and  $\mathbf{I}$  is the identity matrix. The variance matrix for  $\mathbf{y}$  is:  $\mathbf{V} = \mathbf{Z}\mathbf{K}\mathbf{Z}'\sigma_g^2 + \mathbf{W}\mathbf{W}'\sigma_p^2 + \mathbf{I}\sigma_e^2$ .

Letting the ratios  $\frac{\sigma_g^2}{\sigma_e^2}$  and  $\frac{\sigma_p^2}{\sigma_e^2}$  be equal to their estimates at  $\boldsymbol{\beta}_{snp} = 0$ , a computationally efficient algorithm can be used. An ordinary least squares problem is obtained by eigen-decomposing  $\mathbf{V}/\sigma_e^2$  as  $\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$  and premultiplying the entire LMM with  $\boldsymbol{\Lambda}^{-0.5}\boldsymbol{\Gamma}'$ . Since the eigendecomposition only needs to be performed once for all the 1000s of models the computations become feasible. The LMM was compared to the rotated model with fixed variance ratios for a subsample of the markers and the difference in P-values were hardly distinguishable.

**Keywords:** Linear Mixed Model, Eigen-decomposition, Multiple Testing

## Joint estimation of transcription networks and expression with applications to classification

José Sánchez<sup>1</sup>, Tatjana Pavlenko<sup>2</sup> and Rebecka Jörnsten<sup>1</sup>

<sup>1</sup>Chalmers and University of Gothenburg, Gothenburg, Sweden

<sup>2</sup>Royal Institute of Technology, Stockholm, Sweden

### Abstract

We propose to integrate and extend network modeling and expression estimation methods into a joint framework [2]. Our goal is to obtain estimates for transcription networks (given by the inverse covariance matrix) as well as expression levels (given by the mean vectors) for different cancer classes. To this end we assume the expression data for each class to come from a multivariate normal distribution  $N(\boldsymbol{\mu}^k, \boldsymbol{\Sigma}^k)$ , and propose to estimate  $\{\boldsymbol{\mu}^k\}$  and  $\{\boldsymbol{\Omega}^k\}$ , the set of mean vectors and precision matrices across all classes respectively, by optimizing the corresponding profile log-likelihood functions. In practice, given some initial values, we alternate between these two estimation procedures for  $\{\boldsymbol{\mu}^k\}$  and  $\{\boldsymbol{\Omega}^k\}$  until convergence. Each profile log-likelihood is maximized using methods that build on; (i) class specific sample size corrections and (ii) a novel bootstrap procedure for estimating robust sparse and fused structures.

Since our framework is that of Gaussian graphical models, with a minor additional constraint on the precision matrices to be block diagonal, we can recast the problem into an ensemble classifier, where each block takes on the format of a linear or a quadratic component of the corresponding discriminant function. Following [3], the contribution of each model component to the discriminatory power provides insight into the differences between cancer classes.

Finally, we apply our method to expression data for glioblastoma, breast and ovarian cancer from the Cancer Genome Atlas (TCGA). Results are incorporated into our analysis web tool Cancer Landscapes (available at [www.cancerlandscapes.org](http://www.cancerlandscapes.org)) which allows biologists and other scientists to further examine the properties of the estimates and their linear and quadratic components, i.e. model components that distinguish between cancer classes based on differential expression levels only, versus components differential in network structure.

**Keywords:** Cancer, discriminant analysis, expression, fused lasso, high-dimension, inverse covariance matrix, lasso.

## References

- [1] Danaher, P., Wang, P., and Witten, D. (2013) The joint graphical lasso for inverse covariance estimation across multiple classes. *Royal Stat Soc* 76(2) 373–397.
- [2] Pavlenko, T. Björkström, A., and Tillander, A. (2012) Covariance structure approximation via gLasso in high-dimensional supervised classification. *Applied Stat* 39(8) 1643–1666.

## Bias-corrected outlier robust small domain predictors under spatial correlation

Timo Schmid<sup>1</sup>, Ray Chambers<sup>2</sup>, Ralf Münnich<sup>3</sup> and Nikos Tzavidis<sup>4</sup>

<sup>1</sup>Freie Universität Berlin, Germany

<sup>2</sup>University of Wollongong, Australia

<sup>3</sup>University of Trier, Germany

<sup>4</sup>University of Southampton, UK

## Abstract

The presence of outliers in conjunction with small domain-specific sample sizes can cause problems when using model-based estimation of domain finite population parameters. In those cases, outlier robust estimation methods offer protection against the influence of outliers. Conventionally, model-based domain predictors are built by using random effects models that assume independent domain random effects. With real data, however, spatial dependencies in the data often occur. Outlier robust domain prediction in the presence of spatial correlation was recently considered by [1]. Conventionally, outlier robust predictors are plug-ins leading in some cases to bias. Recent literature in outlier robust domain estimation has considered bias-corrected predictors that employ both local and global bias-correction terms (cf. [2] or [3]).

In this talk we propose local and global bias-corrected small domain predictors under spatial correlation. In the case of a global bias-adjustment the correction term includes

the potential spatial impact from other domains on the domain of interest. Inference is performed by using parametric bootstrap. Simulations based on outlier-contaminated data indicate that the proposed predictors can lead to more efficient results. The talk concludes with an application using business survey data for estimating average labor costs in Italian provinces.

**Keywords:** Business data, Fix-point algorithm, Outlier, Spatial patterns

### References

- [1] Schmid, T. and Münnich, R. (2014). Spatial robust small area estimation. *to appear in Statistical Papers*, doi: 10.1007/s00362-013-0517-y.
- [2] Chambers, R., Chandra, H., Salvati, N. and Tzavidis, N. (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society: Series B*, 76:47-69.
- [3] Dongmo-Jiongo, V., Haziza, D. and Duchesne, P. (2014). Controlling the bias of robust small area estimators. *to appear in Biometrika*, doi: 10.1093/biomet/ast030.

## New Block Bootstrap Methods: Sufficient and/or Ordered

**Beste H. Sertdemir** and Esin Firuzan  
Dokuz Eylul university, Izmir, Turkey

### Abstract

In this study, we propose sufficient bootstrap method for time series data to obtain at least the same results with less computing time and less standard error for  $\mu$  than conventional non-overlapping block bootstrap methods. Also, we propose using ordered bootstrapped blocks to preserve more dependency structure of the original data set than that obtained from the blocks. The performances of the proposed methods in terms of estimation and prediction have been compared by a simulation study for MA(2) and AR(2) processes and a real-world example. The results showed that suggested methods are good competitors or they have better performances than conventional methods.

**Keywords:** Block bootstrap, Sufficient bootstrap, Time series.

## First-Order Random Coefficient Autoregressive (RCA(1)) Model: Joint Whittle Estimation and Information

**Mahendran Shitan**<sup>1</sup>, A. Thavaneswaran<sup>2</sup> and Turaj Vazifedan<sup>1</sup>

<sup>1</sup>University Putra Malaysia, Malaysia

<sup>2</sup>University of Manitoba, Canada

### Abstract

Random Coefficient Autoregressive model, (RCA) [1], has been discussed widely as a suitable model for nonlinear time series. The conditional least squares and likelihood parameter estimation of RCA ( $p$ ) model has also been discussed in [1]. The statistical inference of RCA (1) model has been presented in [2] while the conditional least

square estimates for nonstationary processes is studied in [3]. The optimal estimation for nonlinear time series using estimating equations has been investigated in [4].

Recently there has been interest in joint prediction based on spectral density of popular nonlinear time series models such as RCA models. Another way of estimating the parameters of the RCA(1) model is to do Whittle's estimation. In this paper the Whittle estimates of the parameters of an RCA model are studied. It is shown that the Whittle information of the autoregressive parameter in an RCA model is larger than the corresponding information in an Autoregressive (AR) model. Numerical computations of the Whittle's estimates for the parameters of the RCA(1) model were carried out and in particular the bias of the Whittle's estimators are discussed in this paper.

**Keywords:** Bias, Non-linear time series, RCA model, Whittle estimation and information

### References

- [1] Nicholls, D. F., and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction*. Springer-Verlag Inc., Berlin, New York.
- [2] Robinson, P. M. (1978). Statistical Inference for Random Coefficient Autoregressive Model. *Scand. J. Statist.*, 5:163–168.
- [3] Tjøstheim, D. (1986). Estimation in Nonlinear Time Series Models. *Stoch. Proc. Appl*, 21:251–273.
- [4] Thavaneswaran, A., and Abraham, B. (1988). Estimation for Non-linear Time Series Models using Estimating Equations. *J. Time Ser. Anal.*, 9:99–108.

## A new class of goodness-of-fit tests with applications to the problem of detecting sparse heterogeneous mixtures

Natalia Stepanova<sup>1</sup> and Tatjana Pavlenko<sup>2</sup>

<sup>1</sup>Carleton University, Ottawa, Canada

<sup>2</sup>KTH Royal Institute of Technology, Stockholm, Sweden

### 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. Based on the result of Csörgő et al. [1] 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. The results obtained, together with a new result on the convergence in distribution of the higher criticism statistic introduced in Donoho and Jin [2] demonstrate the advantage of our approach over a common approach that utilizes a family of regularly varying weight functions. We also show that, in various subtle problems of detecting sparse heterogeneous mixtures, the proposed test statistics achieve the detection boundary found by Ingster [3] 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, Sparse mixture models

**References**

- [1] Csörgő, M., Csörgő, S., Horváth, L., and Mason, D. (1986). Weighted empirical and quantile processes. *The Annals of Probability*, 14:31–85.
- [2] Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, 32:962–994.
- [3] Ingster, Yu. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distribution. *Mathematical Methods of Statistics*, 6:47–69.

**Higher criticism for estimating proportion of non-null effect in high-dimensional multiple comparison**

Annika Tillander<sup>1</sup> and Tatjana Pavlenko<sup>2</sup>

<sup>1</sup>Karolinska Institutet, Stockholm, Sweden

<sup>2</sup>KTH Royal Institute of Technology, Stockholm, Sweden

**Abstract**

Driven by modern applications in e.g. genetics and metabolomics, the selection of a small subset of features that are likely to be informative for a specific project becomes a challenging problem. This issue is crucial for success of supervised classification in very high-dimensional setting with sparsity patterns. It has been shown that a block-diagonal approximation of the inverse covariance matrix lead to an additive classifier and accounting for the dependence structure yield better performance accuracy [1]. Here we derive an asymptotic framework that represents sparse and weak blocks model and suggest a technique for block-wise feature selection by thresholding. Our procedure extends the standard Higher Criticism (HC) thresholding [2] to the case where dependence structure underlying the data can be taken into account and is shown to be optimally adaptive, i. e. performs well without knowledge of the sparsity and weakness parameters. We empirically investigate the detection boundary of our HC procedure and performance properties of some estimators of sparsity parameter. The relevance and benefits of our approach in high-dimensional classification is demonstrated using both simulation and real data.

**Keywords:** Detection boundary, higher criticism, high dimensionality, separation strength, supervised classification.

**References**

- [1] Pavlenko, T., Björkström, A., and Tillander, A. (2012). Covariance structure approximation via gLasso in high-dimensional supervised classification.. *Journal of Applied Statistics*, 39:1643-1666.
- [2] Donoho, D., and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, 32:962-994.

## Semiparametric Regression in the Presence of Measurement Error

Secil Yalaz Toprak<sup>1</sup>, Müjgan Tez<sup>2</sup> and H. İlhan Tutarlar<sup>1</sup>

<sup>1</sup>Dicle University, Diyarbakir, Turkey

<sup>2</sup>Marmara University, Istanbul, Turkey

### Abstract

The studies about measurement error models extend with the publication of papers on various topics. In literature, semiparametric partially linear model has been mostly studied in case of the measurement error has a known distribution [1, 2]. This study presents more detailed answer to the question that how the predictions of regression functions and densities can be obtained if the measurement error has an unknown distribution in a semiparametric regression model. The identification of the density of an unobserved random variable is possible when the joint density of two error-contaminated measurements of that variable is known [3]. The availability of two error-contaminated measurements of the independent variable is used to achieve the identification. It is shown in the application that the resulting rates are comparable to kernel deconvolution estimators, which provide consistent estimation under the much stronger assumption that the density of the measurement error is known.

**Keywords:** Semiparametric models, Partially linear model, Errors in variables, Measurement error

### References

- [1] Fan, J. and Truong, Y.K. (1993). Nonparametric regression with errors in variables. *Annals of Statistics*, 21: 1900,1925.
- [2] Liang, H. (2000). Asymptotic normality of parametric part in partially linear model with measurement error in the non-parametric part. *Journal of Statistical Planning and Inference*, 86: 51,62.
- [3] Schennach, S.M. (2004). Nonparametric regression in the presence of measurement error. *Econometric Theory*, 20: 1046,1093.

## New results on the Choquet integral based distributions

Vicenç Torra

IIIA-CSIC, Campus UAB, Bellaterra, Spain

### Abstract

We recently introduced multivariate probability distributions based on the Choquet integral. The Choquet integral is an integral for functions with respect to non-additive measures. It reduces to the Lebesgue integral when the measure is additive. Given a vector  $m$  and a non-additive measure  $\mu$ , we denote by  $C(m, \mu)$  the multivariate probability distribution based on the Choquet integral with respect to  $\mu$ .

The distribution based on the Choquet integral is a generalization of the multinormal distribution  $N(0, I)$ . When the measure is additive, we have the following result.

The family of distributions  $N(m, \Sigma)$  in  $\mathbb{R}^n$  with a diagonal matrix  $\Sigma$  of rank  $n$ , and the family of distributions  $C(m, \mu)$  with an additive measure  $\mu$  with all  $\mu(\{x_i\}) \neq 0$  are equivalent.

Nevertheless, in general, for an arbitrary  $\Sigma$  and an arbitrary  $\mu$ , the two families are different. So, they are two different generalizations of  $N(0, I)$ .

In this work we will present our last results related to this type of distributions. See e.g., [1, 2] for some of our previous results.

**Keywords:** Choquet integral, non-additive measures, Choquet integral based distributions.

**Acknowledgements.** Partial support by the Spanish MEC (projects ARES – CONSOLIDER INGENIO 2010 CSD2007-00004, and co-privacy – TIN2011-27076-C03-03) is acknowledged.

## References

- [1] Torra, V. (2014) Distributions based on the Choquet integral and non-additive measures, RIMS Kokyuroku, in press.
- [2] Torra, V., Narukawa, Y. (2012) On a comparison between Mahalanobis distance and Choquet integral: the Choquet-Mahalanobis operator, Information Sciences 190: 56-63.

## Optimization of Compartment Models by Using Metaheuristic Approaches

Ozlem Turksen<sup>1</sup> and Müjgan Tez<sup>2</sup>

<sup>1</sup> Ankara University, Ankara, Turkey

<sup>2</sup> Marmara University, Istanbul, Turkey

### Abstract

Modeling of dynamical systems plays a very important role in applied science, and compartment models are among the most important tools used for analyzing dynamical systems. The compartment models are often used to describe transport of material in biological systems and contain a number of compartments, each containing well mixed material. Most compartment models have more than one compartment and equations for such a model are obtained by describing a conservation law for each compartment. The equations are integrated to form a standard model that defines the response. Generally, the standard model is considered as nonlinear response model. The parameter estimation task can be achieved by using different kind of optimization tools. In this study, metaheuristic approaches are preferred to estimate the model parameters since the model has nonlinear structure. Genetic Algorithm (GA), one of the well known metaheuristic method, and hybrid approaches are applied for optimization of compartment model.

**Keywords:** Compartment Model, Metaheuristic Approaches, Genetic Algorithm, Hybrid Methods, Parameter Estimation

## References

- [1] Blomhøj, M., Kjeldsen, T.H., Ottesen, J., (2014), *Compartment Models*, 1-47.
- [2] Deb, K., (2004), *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley and Sons, New York.
- [3] Dreo, J., Petrowski, A., Siarry, P., Taillard, E. (2006), *Metaheuristics for Hard Optimization*, Springer-Verlag Berlin, Heidelberg.
- [4] Motulsky, H., Christopoulos, A., (2003), *Fitting Models to Biological Data Using Linear and Nonlinear Regression*, GrandPad Software.
- [5] Nelder, J.A., Mead, R. A., (1965), Simplex method for function minimization, *Computer Journal*, 7, 308-313.
- [6] Turksen, O., (2011), *Fuzzy and Heuristic Approach to the Solution of Multi-Response Surface Problems*, Ph.D. thesis, Ankara University.

## Testing Covariance Structure in High-Dimensional Setup: Comparison of Score Test and Likelihood Ratio Test

Marju Valge<sup>1</sup>, Tonu Kollo<sup>1</sup> and Dietrich von Rosen<sup>2,3</sup>

<sup>1</sup>Univeristy of Tartu, Tartu, Estonia

<sup>2</sup>Linköping university, Linköping, Sweden

<sup>3</sup>Swedish University of Agricultural Sciences, Uppsala, Sweden

### Abstract

In data analysis and modeling one is often interested in testing presence of certain covariance structure. In a simple case the hypothesized covariance matrix is an identity matrix, in more complex situations (e.g. when analyzing spacial-temporal data) the covariance matrix could have Kronecker product structure, for instance. In any case, one is interested in using a most suitable statistical test. Probably the most commonly used test is the likelihood ratio test (LRT). However, it has been shown that when the dimension of the covariance matrix increases the likelihood ratio test is not applicable any more, as it will almost always reject the null hypothesis. In order to solve this problem, corrections to the test have been made so that it could be used in high-dimensional case as well (see Bai, Jiang, Yao and Zheng (2009), for example). We carry out a simulation experiment to show that instead of the likelihood ratio test Rao's score test (RST) should be used to test for covariance structure when the dimension of the covariance matrix is high.

First we will introduce likelihood ratio test and Rao's score test and derive formulas for testing covariance structure in the multivariate Normal case. Then we compare the behaviour of both tests in high-dimensional setup using simulations. Finally we summarize the results and write down the conclusions. [1]

**Keywords:** Covariance structure, high dimension, hypothesis testing, likelihood ratio test, score test, simulation.

### References

- [1] Bai, Z., Jiang, D., Yao, J., Zheng, S. (2009). Corrections to LRT on Large Dimensional Covariance Matrix by RMT. *Annals of Statistics 2009*, Vol. 37, No. 6B, 3822-3840.



## SIS-BEN: an efficient two-stage approach for high dimensional analysis of correlated data

**Patrik Waldmann**

Linköping University, Sweden

### Abstract

High dimensional data analysis has become a major issue in a range of different scientific fields. Wide big data is often characterized by a huge number of predictor variables measured on a limited number of individuals. Moreover, variables are often correlated following complicated patterns. In these situations, it turns out to be difficult to find a method that provides an optimum between few computational operations and high statistical reliability. Tuning of hyper-parameters in regularized regression tends to be difficult in the frequentist paradigm, and Bayesian methods are often computationally intractable because of the need to iteratively cycle through a large number of variables. A two-stage approach that is computationally efficient and provides very low error rates regarding variable selection is here presented. The first stage is based on the Sure Independence Screening (SIS) technique, which is both fast and well-suited for correlated data. The second stage consist of the Bayesian Elastic Net (BEN) with adaptive regularization parameters. The method is tested on large scale simulated data, and real data from a genome-wide association study.

## Estimation in Multivariate $t$ Nonlinear Mixed-effects Models with Missing Outcomes

**Wan-Lun Wang**<sup>1</sup> and Tsung-I Lin<sup>2</sup>

<sup>1</sup>Feng Chia University, Taichung, Taiwan

<sup>2</sup>National Chung Hsing University, Taichung, Taiwan

### Abstract

The multivariate nonlinear mixed-effects model (MNLMM) has emerged as an effective tool for modelling multi-outcome longitudinal data following nonlinear growth patterns. In the framework of MNLMM, the random effects and within-subject errors are routinely assumed to be normally distributed for mathematical tractability and computational simplicity. However, a serious departure from normality may cause lack of robustness and subsequently make invalid inference. In this work, we introduce a robust extension of the MNLMM by considering a joint multivariate  $t$  distribution for the random effects and within-subject errors, called the multivariate  $t$  nonlinear mixed-effects model (MtNLMM). Moreover, a damped exponential correlation structure is employed to capture the extra serial correlation among irregularly observed multiple repeated measures. An ECM procedure coupled with the first-order Taylor approximation is developed for estimating model parameters. The techniques for estimation of random effects, imputation of missing responses and identification of potential outliers are also investigated. The methodology is motivated by, and applied to a real data set concerning 161 pregnant women from a study in a private fertilization obstetrics clinic in Santiago, Chile.

**Keywords:** Damped exponential correlation, ECM algorithm, Imputation, Multivariate longitudinal data, Outlier detection

## Bayesian inference for heteroscedastic Rician time series with applications to fMRI

Bertil Wegmann<sup>1</sup>, Anders Eklund<sup>2</sup>, Stephen LaConte<sup>2</sup>, Mattias Villani<sup>1</sup>

<sup>1</sup>Linköping University, Sweden

<sup>2</sup>Virginia Tech Carilion Research Institute, Virginia Tech, Roanoke, VA, USA

### Abstract

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive neuroimaging technique that measures brain activity. An active brain area consumes more oxygenated blood to support neural activity and the blood-oxygen-level dependent (BOLD) signal measures the change in blood flows. Data from an fMRI experiment is a time series of BOLD signals within each of hundreds of thousands of voxels in 3D brain images.

The common practice in fMRI studies is to assume that the magnitude of the BOLD signal follows a linear model with Gaussian noise. However, such a variable is well known to follow the Rician distribution (Gudbjartsson and Patz, 1995). The widespread use of the Gaussian distribution in fMRI comes from its simpler statistical properties and that the Rician distribution is well approximated by a Gaussian distribution when the Signal-to-Noise Ratio (SNR) is high. However, the push toward higher spatial resolutions in fMRI studies (Feinberg and Yacoub, 2012) may reduce the SNRs (Macovski, 1996) to a level where inaccurate Gaussian approximations lead to severely distorted activation maps.

We propose a new Bayesian model for heteroscedastic time series with Rician noise for fMRI analysis and a highly efficient Markov Chain Monte Carlo (MCMC) algorithm for inference based on a general approach in Villani et al. (2009).

**Keywords:** Bayesian inference, fMRI, Heteroscedastic time series, Markov chain Monte Carlo, Rician noise.

### References

- [1] Feinberg, D. et al. (2012). The rapid development of high speed, resolution and precision in fMRI. *NeuroImage*, vol. 62, 720-725.
- [2] Gudbjartsson, H. et al. (1995). The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, vol. 34, 910-914.
- [3] Macovski, A. (1996). Noise in MRI. *Magnetic Resonance in Medicine*, vol. 36, 494-497.
- [4] Villani, M. et al. (2009). Regression density estimation using smooth adaptive Gaussian mixtures. *Journal of Econometrics*, vol. 153, 155-173.

## SimSel - a Method for Variablen Selection

**Silvelyn Zwanzig**

Uppsala University, Sweden

### Abstract

The principle of SimSel is to study the influence of stepwise perturbation of the variable under consideration on the quality of a model fit. The main idea is that the disturbance of an unimportant variable will have no effect. Furthermore an independent pseudo variable is implemented to the data set which serves as an untreated control group and is explored for estimating the null distribution in the test step. The SimSel method introduced in Eklund and Zwanzig (2012) is based on an approximative quadratic model and the least squares criterion. Now generalizations in two extensions are considered.

- The design matrix is singular.
- Several variables are disturbed simultaneously.
- Instead of least squares other criteria for the model fit are applied, as a generalized Ridge or Lasso.

The goal is to detect unimportant variables as well as variables, which can be replaced by others.

**Keywords:** Variablen Selection, Ridge, Generalized Inverse, Pseudo Errors

### References

- [1] Eklund, M. and Zwanzig, S. (2012). SimSel: a new simulation method for variable selection. *Journal of Statistical Computation & Simulation*, 82:515-527.
- [2] Eklund, M. and Zwanzig, S. (2009). Ridge-SimSel - a generalization of the variable selection method SimSel to multicollinear data sets. U.U.D.M. 2009:12, ISSN 1101-3591.

## Diagnostic statistics for binary logistic regression in the presence of multicollinearity

**M. Revan Özkale**

Çukurova University, Adana, Turkey

### Abstract

Logistic regression has been widely used in many areas of applications to model binary data. When the explanatory variables are highly correlated in logistic regression, it is recognized that the maximum likelihood estimates are unstable or infinite and the variances of the maximum likelihood estimators are inflated. Hence, biased estimators in logistic regression are proposed by various authors such as [1], [2], [4]. Nevertheless, the presence of outliers and influential observations is an unavoidable phenomenon in data analysis. Hence, an important aspect of regression analysis is the identification of outlying

and influential observations. Although diagnostic statistics based on maximum likelihood estimates in logistic regression are available to assist this identification problem, there is no recommendation for the use of these diagnostic statistics with biased estimators in the existence of multicollinearity.

In this study, using one-step approximation method introduced by [3], change in parameter estimates and change in chi-square statistic due to case deletion is developed as well as residual and leverages under the ridge and principal components regression estimators for detecting influential and outlying observations in logistic regression with multicollinear data. The methods are illustrated using a data set and a simulation study is conducted to see the effect of the sample size, the number of continuous explanatory variables and the degree of the multicollinearity.

**Keywords:** Logistic regression, Diagnostic measures, Principal components logistic regression estimator, Ridge logistic estimator

### References

- [1] LeCessie, S. and VanHouwelingen, J.C. (1992). Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191-201.
- [2] Marx, B.D. and Smith, E.P. (1990). Principal component estimation for generalized linear regression. *Biometrika*, 77(1):23-31.
- [3] Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, 9(4):705-724.
- [4] Smith, E.P. and Marx, B.D. (1990). Ill-conditioned information matrices, generalized linear models and estimation of the effects of acid rain, *Environments*, 1(1):57-71.

## 4 | Abstract - Posters

### Simulation Study of the Proportional Hazards Model with Delayed entry

Silvie Bělašková and Eva Fišerová  
Palacký University, Olomouc, Czech Republic

#### Abstract

In the survival analysis, studied subject sometimes entries during the observed time period. In the proportional hazards model [3], these subjects are referred as delayed entry or left truncated [1]. With real data, there is a possibility that the delayed entries approach may have extremely different hazards of the covariates, and statistical significance of their estimated coefficients as well, compared to the standard approach, see e.g. [2].

The aim of this contribution is to demonstrate hazard changes with different censored proportions and delayed entry. We constructed a model that includes covariates which are normally distributed and for the effect of delayed entry bias, having five different time specification, with respect to time recovery after the surgery. The first delayed time is 24 hours after the surgery as the first 24 hours postoperatively seems to be a potential period of the unexpected surgical complications. The simulations are considered for finite sample size when the failure time follows Weibull distribution and the censoring time follows the Poisson distribution.

**Keywords:** Delayed entry, Proportional hazards model, Simulation study.

#### References

- [1] Su, Y.; Wang, J. (2012). Modeling left-truncated and right-censored survival data with longitudinal covariates. *The Annals of Statistics*, 3:1465-1488.
- [2] Bělašková, S.; Fišerová, E. (2013). Study of Bootstrap Estimates in Cox Regression Model with Delayed Entry. *Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica*, 52:21-30.
- [3] Cox, D.R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society, Series B*, 20:187–220.

## Numerical reconstruction of the support of a source in an elliptic equation

Bandar Bin-Mohsin<sup>1</sup> and Daniel Lesnic<sup>2</sup>

<sup>1</sup>King Saud University, Riyadh, Saudi Arabia

<sup>2</sup>University of Leeds, Leeds, UK

### Abstract

In this paper, the inverse problem which consists of reconstructing an unknown inner boundary of a two-dimensional domain from a single pair of boundary Cauchy data associated to an elliptic equation is solved numerically using the meshless method of fundamental solutions. A nonlinear minimisation of the objective function is regularised when noise is added into the input boundary data. The stability of the numerical results is investigated for several test examples with respect to noise in the input data and various values of the regularisation parameters.

**Keywords:** Elliptic equation, Inverse problem, Inhomogeneity, Method of fundamental solutions, Regularisation.

### References

- [1] Fang, W., and Cumberbatch, E. (1992). Inverse problems for metal oxide semiconductor field-effect transistor contact resistivity. *SIAM Journal on Applied Mathematics*, 52:699–709.
- [2] Isakov, V. (2009). Inverse obstacle problems. *Inverse Problems*, 25:123002.
- [3] Kang, H., Kwon, K., and Yun, K. (2001) Recovery of an inhomogeneity in an elliptic equation. *Inverse Problems*, 17:25–44.
- [4] Hettlich, F., and Rundell, W. (1997). Recovery of the support of a source term in an elliptic differential equation. *Inverse Problems*, 13:959–976.
- [5] Kim, S., and Yamamoto, M. (2003). Uniqueness in identification of the support of a source term in an elliptic equation. *SIAM Journal on Mathematical Analysis*, 35:148–159.
- [6] Kim, S. (2002). Uniqueness determination of inhomogeneity in an elliptic equation. *Inverse Problems*, 18:1325–1332.
- [7] Bin-Mohsin, B., and Lesnic, D. (2011). Determination of inner boundaries in modified Helmholtz inverse geometric problems using the method of fundamental solutions. *Mathematics and Computers in Simulation*, 82:1445–1458.
- [8] Lesnic, D., and Bin-Mohsin, B. (2012). Inverse shape and surface heat transfer coefficient identification. *Journal of Computational and Applied Mathematics*, 236:1876–1891.
- [9] Karageorghis, A., Lesnic, D., and Marin, L. (2011). Survey of applications of the MFS to inverse problems. *Inverse Problems in Science and Engineering*, 19:309–336.

## The comparative log–linear analysis of unemployment in Poland in R

**Justyna Brzezińska**

University of Economics in Katowice, Poland

### Abstract

In categorical data analysis we can analyze categorical variables simultaneously in multi-way tables. This paper presents the use of log–linear models which allow to analyze the independence between any number of categorical variables. Different types of association can be provided: for nominal variables we have conditional independence, homogeneous association or conditional independence model, and for ordinal we have row or column effects model or uniform association. There are several criteria for testing the goodness–of–fit of the model: the chi-square statistic, the likelihood ratio, information criteria (AIC, BIC) [1], [2], [3].

With the rising unemployment rate in recent years, unemployment is one of the most important economic and social problem in Poland. A strong differentiation is observed in the unemployment rates for various regions of Poland, especially for young and university graduates, as well as for males as females. The log–linear analysis will be presented on the example from the Central Statistical Office of Poland. The comparative log–linear analysis will be conducted for multi-way tables on unemployment in 2004–2013. All calculations will be conducted in R with the use of `loglm` function in MASS library.

The project is funded by a grant from the National Science Centre based on the decision DEC-2012/05/N/HS4/00174

**Keywords:** log–linear models, analysis of multi-way tables, unemployment in Poland.

### References

- [1] Agresti A. (2002). Categorical data analysis. 2nd ed. Hoboken, NJ: Wiley.
- [2] Christensen, R. (1997). Log-linear Models and Logistic Regression. Springer–Verlag, New York.
- [3] Knoke D., Burke P. (1980). Log-linear Models. Sage University Paper Series on Quantitative Applications in the Social Science, series no. 07-020, Beverly Hills and London Sage.

## Confidence Bands for Percentile line in the Linear Regression Model

**Martina Chvosteková**

Palacký University Olomouc, Olomouc, Czech Republic

### Abstract

We consider the general linear regression model  $Y = \beta_0 \vec{1} + X\beta + \epsilon$ , where  $Y$  is an  $n \times 1$  vector of observations,  $X$  is an  $n \times p$  matrix of known constants,  $\beta_0$  is an unknown intercept,  $\beta$  is a  $q \times 1$  vector of unknown regression coefficients,  $\epsilon$  is an  $n \times 1$  vector of

independent identically normally distributed errors, each with zero mean and a common unknown variance  $\sigma^2$ . Let

$$p_{\gamma, T(X)} = \beta_0 + x^T \beta + \sigma z_\gamma$$

denotes the 100 $\gamma$ th percentile line for all  $x \in T(X)$ , where  $z_\gamma$  denotes the 100 $\gamma$ th percentile of the standard normal distribution,  $T(X) = \cup_{d_{min} \leq d \leq d_{max}} T^*(X, d)$ ,  $T^*(X, d) = \{x^T = \xi^T X : \xi \in R^q, 1/n + x^T (X^T X)^{-1} x = d^2\}$ .

The problem is to find a  $(1 - \alpha)$ -confidence band for  $p_{\gamma, T(X)}$ . The  $(1 - \alpha, \gamma)$ -simultaneous one-sided tolerance intervals can be used as the one-sided confidence band for  $p_{\gamma, T(X)}$ . We extended the one-sided confidence bands suggested in [1] and in [2] to the case  $x \in T(X)$  and we compared them with the simultaneous one-sided tolerance intervals suggested by Odeh and Mee [3]. Our results suggest the optimal form of the one-sided confidence band has not been found yet. We provided similar comparison of all the two-sided  $(1 - \alpha)$ -confidence bands for  $p_{\gamma, T(X)}$ .

**Keywords:** Confidence band, Percentile, Simultaneous inference

**Acknowledgment:** The work was supported by the project of the European Social Fund of the Ministry of Education, Youth and Sports of the Czech Republic, CZ.1.07/2.3.00/30.0041.

## References

- [1] Turner, D. L., Bowden, D. C. (1979). Sharp confidence bands for percentile lines and tolerance bands for the simple linear model. *Journal of the American Statistical Association*, 74(368), 885-888.
- [2] Thomas, D. L., Thomas, D. R. (1986). Confidence bands for percentiles in the linear regression model. *Journal of the American Statistical Association*, 81(395), 705-708.
- [3] Odeh, R. E., Mee, R. W. (1990). Ones-sided simultaneous tolerance limits for regression. *Communication in Statistics - Simulation and Computation*, 19(2), 663-680.

## IPM Method for the BLUEs of Subparameters under Multiple-Partitioned Linear Model

Nesrin Güler<sup>1</sup> and Simo Puntanen<sup>2</sup>

<sup>1</sup>Sakarya University, Sakarya, Turkey

<sup>2</sup>University of Tampere, Tampere, Finland

### Abstract

It is well-known that  $Gy$  is the best linear unbiased estimator (BLUE) of  $X\beta$  under the general linear model  $M = \{y, X\beta, V\}$  if and only if the matrix  $G$  satisfies the fundamental BLUE equation  $G(X : VX^\perp) = (X : 0)$ , see, e.g., [2]. This equation can be rewritten in the following form:

$$\Gamma \begin{pmatrix} G' \\ L \end{pmatrix} = \begin{pmatrix} 0 \\ X' \end{pmatrix}, \quad (4.1)$$

where  $\Gamma = \begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}$ . Rao [1] shows that the problem of inference from a linear model can be completely solved when one has obtained an arbitrary generalized inverse of



the partitioned matrix  $\Gamma$ . This approach based on the numerical evaluation of an inverse of the partitioned matrix  $\Gamma$  is known as the inverse partitioned matrix (IPM) method. In this study, we consider the general multiple-partitioned linear model  $M = \{y, X_1\beta_1 + \cdots + X_k\beta_k, V\}$  to characterise the BLUEs of subparameters. Using the IPM method, we give some results related to the BLUEs of  $X_i\beta_i$ ,  $i = 1, \dots, k$ , under the general multiple-partitioned linear model  $M$ .

**Keywords:** BLUE, general multiple-partitioned model, generalized inverse.

### References

- [1] Rao, C. R. (1971). Unified theory of linear estimation. *Sankhyā*, A33, 371-394. [Corrigendum (1972), 34, 194, 477.]
- [2] Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *The Annals of Mathematical Statistics*, 38, 1092-1109.

## Model Selection Information Criteria for Inflated Data

Ting Hsiang Lin and Min-Hsiao Tsai  
National Taipei University, Taiwan

### Abstract

Inflated data are prevalent in many situations. A variety of inflated models with extensions have been derived to fit data with excessive counts of some particular responses, and the family of Information Criteria (IC) have been used to compare the fit of models for selection purpose. Yet despite the common use in statistical applications, the information criteria have received little attention in inflated models.

A simulation study was conducted to compare the performance of three information criteria for inflated data and three inflated models are compared: Poisson regression (POI), Zero-inflated Poisson (ZIP), and a new proposed Zero- and K-inflated Poisson regression model (ZKIP). The effect of sample sizes, proportions of zero and K observations towards selection performance are also examined.

The results suggest that BIC and CAIC are more accurate than AIC in terms of model selection when the true model is simple (Poisson regression). For more complex models, such as ZIP and ZKIP, AIC was consistently better than BIC and CAIC, although it did not reach high levels of accuracy when sample size and proportion of zero observations are small. AIC also tended to over-fit the data for Poisson regression, whereas BIC and CAIC tended to under-parameterize the data for ZIP and ZKIP. Therefore, it is desirable to study other model selection criteria for small sample size for inflated data.

**Keywords:** Zero-inflated data, Model Selection, Information Criteria

### References

- [1] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34: 1-14.

- [2] Welsh, A.H., Cunningham, R.B., Donnelly, C.F. and Lindenmayer, D.B. (1996). Modelling the abundance of rare species: Statistical models for counts with extra zeros. *Ecological Modelling*, 88:297-308.
- [3] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *B.N. Petrov and F. Csake (eds.), Second International Symposium on Information Theory. Budapest: Akademiai Kiado*, 267-281.
- [4] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:4.
- [5] Bozdogan, H. (1987). Model-selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52:345-370.

## A comparison of R and Stata in sample surveys

**Ramidha Srihera**, Alongkorn Chamsranoi, Manutchanak Noiyanon and Pakamas Chaiyawan  
Thammasat University, Pathum Thani, Thailand

### Abstract

We compared two statistical packages (R and Stata) in estimating mean of cash and marketable securities reported in the balance sheet from 102 countries in the world (in millions of US\$) under simple random sampling and stratified random sampling methods using data set from NYU Stern school of business. The efficiency of the estimated mean from both packages is expressed in terms of bias and mean squared error. The results present that both statistical packages are found to be practiced and flexible in estimation of mean. However, it depends on the types of criteria. Besides, the effect of using different sampling schemes does not appear. At the end, there is some useful comment on determination of sample size in each stratum from both packages.

**Keywords:** Estimated mean, simple random sampling, stratified random sampling

### References

- [1] [http://pages.stern.nyu.edu/~adamodar/New\\_Home\\_Page/data.html](http://pages.stern.nyu.edu/~adamodar/New_Home_Page/data.html)
- [2] Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. New York: Wiley.
- [3] Damico, A. (2009). Transitioning to R: Replicating SAS, Stata, and SUDAAN Analysis Techniques in Health Policy Data. *The R Journal*, 1/2:37-44.
- [4] Oyeyemi, G.M., Adewara, A.A. and Adeyemi, R.A. (2010). Complex Survey Data Analysis: A Comparison of SAS, SPSS and STATA. *Asian Journal of Mathematics and Statistics*, 3(1):33-39.

## On some sampling distributions for skew normal population

Nan-Cheng Su<sup>1</sup> and Arjun K. Gupta<sup>2</sup>

<sup>1</sup>National Taipei University, Taiwan

<sup>2</sup>Bowling Green State University, USA

### Abstract

The distribution of weighted function of independent skew normal random variables, which includes the sample mean, is useful in many applications. In this paper we derive this distribution and study the null distribution of a linear form and a quadratic form. Finally, we discuss some of its applications in control charts, in which the skew normal model plays a key role.

**Keywords:** independence; moment generating function; quadratic form; skew normal distribution; skew normal sample mean

### References

- [1] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, 12: 171-178.
- [2] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. Roy. Soc. Ser. B*, 61:579-602.
- [3] Chen, J.T., Gupta, A.K. and Nguyen, T.T. (2004). The density of the skew normal sample mean and its applications. *J. Statist. Comput. Simulation*, 74:487-494.
- [4] Gupta, A.K. and Huang, W.J. (2002). Quadratic forms in skew normal variates. *J. Math. Anal. Applications*, 273:558-564.
- [5] Tsai, T.R. (2007). Skew normal distribution and the design of control charts for averages. *International J. Reliability, Quality Safety Engineering*, 14:49-63.

## On the Curvature Measurements of the Nonlinear Errors In Variables Models (NEIVM)

Pakize Taylan<sup>1</sup> and Ersin Uysal<sup>1</sup> and Müjgan Tez<sup>2</sup>

<sup>1</sup>Dicle University, Diyarbakir, Turkey

<sup>2</sup>Marmara University, Istanbul, Turkey

### Abstract

In regression models when explanatory variable contains measurements error parameter estimation can be affected by this error. In our study, we considered parameter estimation for nonlinear errors in variables models and it was investigated how the parameter estimates and surface curvatures were affected by the error using the differential geometric approaches which contain curvature of the model function. Statistical inference was performed by comparing "Parameter Effects Curvature and Intrinsic Curvature" in cases where the independent variables contain error and do not contain an error.

**Keywords:** Curvature, Errors-In-Variables, Intrinsic Curvature, Maximum likelihood, Measurement Errors, Nonlinear model, Parameter Effect.

**References**

- [1] Bates, D.M. and Watts, D.G., (1980), Relative Curvature Measures of Nonlinearity, *Journal of the Royal Statistical Society*, Ser. B40, 1-25.
- [2] Bates, D.M. and Watts, D.G., (1981), Parameter Transformations for Improved Approximate Confidence Regions in Nonlinear Least Squares, *Annals of Statistics*, Vol.9, No.6, 1152-1161.
- [3] Bates, D.M. and Watts, D.G., (1988), *Nonlinear Regression Analysis Applications*, Kingston. Wiley.
- [4] Beale, E.M.L., (1960), Confidence Regions in Nonlinear Estimation (with discussion), *Journal of the Royal Statistical Society*, B22, 41-48.

## 5 | List of Participants

1. **Alan Agresti**  
University of Florida, USA  
*aa@stat.ufl.edu*
2. **S. Ejaz Ahmed**  
Brock University, Canada  
*sahmed5@brocku.ca*
3. **Kadri Ulas Akay**  
Istanbul University, Turkey  
*kulas@istanbul.edu.tr*
4. **Esra Akdeniz-Duran**  
Istanbul Medeniyet University, Turkey  
*esraakdeniz@gmail.com*
5. **Moudud Alam**  
Dalarna University, Sweden  
*maa@du.se*
6. **Barbora Arendacka**  
Germany  
*barendacka@gmail.com*
7. **Öyküm Esra Askin**  
Yildiz Technical University Turkey  
*oykumesra@gmail.com*
8. **Anthony Atkinson**  
London School of Economics, United Kingdom  
*a.c.atkinson@lse.ac.uk*
9. **Sellahattin Aydogdu**  
Yildiz Technical University Turkey  
*saydogdu@yildiz.edu.tr*
10. **Alena Bachratá**  
Comenius University in Bratislava, Slovakia  
*alena.bachrata@fmph.uniba.sk*
11. **Silvie Belaskova**  
Palacky University Olomouc, Czech Republic  
*belaskova@fai.utb.cz*
12. **Ufuk Beyaztas**  
Dokuz Eylul University, Turkey  
*ufuk.beyaztas@deu.edu.tr*
13. **Elaine Billberg**  
Linköping University, Sweden  
*elaine.billberg@liu.se*

14. **Bandar Bin-Mohsin**  
King Saud Universtiy, Saudi Arabia  
*balmohsen@ksu.edu.sa*
15. **Barbara Bogacka**  
Queen Mary, University of London, United Kingdom  
*b.bogacka@qmul.ac.uk*
16. **Fredrik Bonander**  
Inspektionen för arbetslöshetsförsäkringen, Sweden  
*fredrik.bonander@iaf.se*
17. **Justyna Brzezinska**  
University of Economics in Katowice, Poland  
*justyna.brzezinska@ue.katowice.pl*
18. **Camila Caiado**  
Durham University, United Kingdom  
*c.c.d.s.caiado@durham.ac.uk*
19. **Francico Carvalho**  
New University of Lisbon and Instituto Politecnico de Tomar, Portugal  
*fpcarvalho@ipt.pt*
20. **Jinwen Chen**  
Tsinghua University, China  
*chen@math.tsinghua.edu.cn*
21. **Martina Chvosteková**  
Palacky University Olomouc, Czech Republic  
*martina.chvostekova@upol.cz*
22. **Dennis Cook**  
University of Minnesota, USA  
*dennis@stat.umn.edu*
23. **Deliang Dai**  
Linnaeus University, Sweden  
*deliang.dai@lnu.se*
24. **Marc Ditzhaus**  
Heinrich Heine Universität Düsseldorf, Germany  
*marc.ditzhaus@uni-duesseldorf.de*
25. **Dennis Dobler**  
University of Duesseldorf, Germany  
*dedob100@hhu.de*
26. **Sandra Donevska**  
Palacky University Olomouc, Czech Republic  
*sdonevska@seznam.cz*
27. **Hilmar Drygas**  
University of Kassel, Germany  
*hilmar.drygas@onlinehome.de*
28. **Esra Emiroglu**  
Istanbul University, Turkey  
*eertan@istanbul.edu.tr*
29. **Jan-Eric Englund**  
Swedish University of Agricultural Sciences, Sweden  
*Jan-Eric.Englund@slu.se*

30. **Birsen Eygi Erdogan**  
Marmara University, Turkey  
*birsene@marmara.edu.tr*
31. **Kamila Facevicova**  
Palacky University Olomouc, Czech Republic  
*kamila.facevicova@gmail.com*
32. **Majbritt Felleki**  
Dalarna University, Sweden  
*mfl@du.se*
33. **Katarzyna Filipiak**  
Poznan University of Life Sciences, Poland  
*kasfil@up.poznan.pl*
34. **Esin Firuzan**  
Dokuz Eylul University, Turkey  
*esin.firuzan@deu.edu.tr*
35. **Eva Fiserova**  
Palacky University Olomouc, Czech Republic  
*eva.fiserova@upol.cz*
36. **Miguel Fonseca**  
New University of Lisbon, Portugal  
*fmig@fct.unl.pt*
37. **Johannes Forkman**  
Swedish University of Agricultural Sciences, Sweden  
*johannes.forkman@slu.se*
38. **Michael Goldstein**  
Durham University, United Kingdom  
*michael.goldstein@durham.ac.uk*
39. **Huseyin Guler**  
Cukurova University, Turkey  
*hguler@cu.edu.tr*
40. **Nesrin Güler**  
Sakarya University, Turkey  
*nesring@sakarya.edu.tr*
41. **Shengqiang Guo**  
Dalarna University, Sweden  
*wodeorient@163.com*
42. **Sat Gupta**  
University of North Carolina at Greensboro, USA  
*sngupta@uncg.edu*
43. **Chengcheng Hao**  
Stockholm University, Sweden  
*chengcheng.hao@stat.su.se*
44. **Stephen Haslett**  
Massey University, New Zealand  
*s.j.haslett@massey.ac.nz*

45. **Jan Hauke**  
Adam Mickiewicz University, Poland  
*jhauke@amu.edu.pl*
46. **Jean-Marie Helbling**  
Ecole Polytechnique Fédérale de ÉLausanne, Switzerland  
*nadia.kaiser@epfl.ch*
47. **Thomas Holgersson**  
Jönköping International Business School, Sweden  
*thomas.holgersson@ihh.hj.se*
48. **Klara Hruzova**  
Palacky University Olomouc, Czech Republic  
*klara.hruzova@gmail.com*
49. **Su-Yun Huang**  
Academia Sinica, Taiwan  
*syhuang@stat.sinica.edu.tw*
50. **Shinpei Imori**  
Hiroshima University, Japan  
*imori.stat@gmail.com*
51. **Deniz Inan**  
Marmara University, Turkey  
*denizlukuslu@marmara.edu.tr*
52. **Geoffrey Jones**  
Massey University, New Zealand  
*g.jones@massey.ac.nz*
53. **Rebecka Jörnsten**  
University of Gothenburg and Chalmers Institute of Technology, Sweden  
*jornsten@chalmers.se*
54. **Alzbeta Kalivodova**  
Palacky University Olomouc, Czech Republic  
*Kalivodovaa@gmail.com*
55. **Emil Karlsson**  
Linköping University, Sweden  
*emika583@student.liu.se*
56. **John Karlsson**  
Linköping University, Sweden  
*john.karlsson@liu.se*
57. **Daniel Klein**  
P.J. Safárik University, Slovakia  
*daniel.klein@upjs.sk*
58. **Petr Koldanov**  
NRU Higher School of Economics, Russia  
*pkoldanov@hse.ru*
59. **Tomasz Kossowski**  
Adam Mickiewicz University, Poland  
*tkoss@amu.edu.pl*



60. **Sergii Kulishov**  
Ukrainian medical stomatological academy, Ukraine  
*kulishov@meta.ua*
61. **Shinji Kuriki**  
Osaka Prefecture University, Japan  
*kuriki@ms.osakafu-u.ac.jp*
62. **Fatma Sevinc Kurnaz**  
Yildiz Technical University, Turkey  
*fskurnaz@gmail.com*
63. **Fikriye Kurtoglu**  
Cukurova University, Turkey  
*fkurtoglu@cu.edu.tr*
64. **Lynn LaMotte**  
LSU Health Sciences Center School of Public Health, New Orleans, USA  
*llamot@lsuhsc.edu*
65. **Eammanuel Lesaffre**  
Leuven Biostatistics and Statistical Bioinformatics Centre, Belgium  
*e.lesaffre@erasmusmc.nl*
66. **Ying Li**  
Swedish University of Agricultural Sciences, Sweden  
*ying.li@slu.se*
67. **Yuli Liang**  
Stockholm University, Sweden  
*yuli.liang@stat.su.se*
68. **Tsung-I Lin**  
National Chung Hsing University, Taiwan  
*tilin@nchu.edu.tw*
69. **Ting Hsiang Lin**  
National Taipei University, Taiwan  
*tinghlin@mail.ntpu.edu.tw*
70. **Yonghui Liu**  
Shanghai Finance University, China  
*liuyh@lsec.cc.ac.cn*
71. **Måns Magnusson**  
Linköping University, Sweden  
*mons.magnusson@gmail.com*
72. **Behrang Mahjani**  
Uppsala University, Sweden  
*behrang.mahjani@it.uu.se*
73. **Augustyn Markiewicz**  
Poznan University of Life Sciences, Poland  
*amark@up.poznan.pl*
74. **Thomas Mathew**  
University of Maryland Baltimore County, USA  
*mathew@umbc.edu*
75. **Jamal Najim**  
Université Paris Est and CNRS, France  
*najim@univ-mlv.fr*

76. **Hiba Nassar**  
Linnaeus University, Sweden  
*hiba.nassar@lnu.se*
77. **Innocent Ngaruye**  
Linköping University, Sweden  
*innocent.ngaruye@liu.se*
78. **Joseph Nzabanita**  
Linköping University, Sweden  
*joseph.nzabanita@liu.se*
79. **Kazuhiro Ozawa**  
Gifu college of nursing, Japan  
*k-ozawa@gifu-cn.ac.jp*
80. **Sudhir Paul**  
University of Windsor, Canada  
*smjp@uwindsor.ca*
81. **Per-Anders Paulsson**  
Inspektionen för arbetslöshetsförsäkringen, Sweden
82. **Tatjana Pavlenko**  
KTH Royal Institute of Technology, Sweden  
*pavlenko@math.kth.se*
83. **Jolanta Pielaszkiewicz**  
Linköping University, Sweden  
*jolanta.pielaszkiewicz@liu.se*
84. **Maryna Prus**  
Otto-von-Guericke University, Germany  
*marina\_magdeburg@yahoo.com*
85. **Marcin Przystalski**  
The Research Center for Cultivar Testing, Slupia Wielka, Poland  
*m.przystalski@coboru.pl*
86. **Simo Puntanen**  
University of Tampere, Finland  
*simo.puntanen@uta.fi*
87. **Matias Quiroz**  
Stockholm University and Sveriges Riksbank, Sweden  
*quiroz.matias@gmail.com*
88. **Nils Hassan Quttineh**  
Linköping University, Sweden  
*nils-hassan.quttineh@liu.se*
89. **Abdalla Ramadan Abdalla Khalid**  
Youth peer education network, Sudan  
*shagy555@live.com*
90. **Martin Ricker**  
Universidad Nacional Autónoma de México, Mexico  
*martin\_tuxtla@yahoo.com.mx*
91. **Paulo Rodrigues**  
New University of Lisbon, Portugal  
*paulocanas@gmail.com*

92. **Veronika Rosteck**  
Springer, Germany  
*Veronika.Rosteck@springer.com*
93. **Anuradha Roy**  
The University of Texas at San Antonio, USA  
*Anuradha.Roy@utsa.edu*
94. **Patrik Ryden**  
UmeåUniversity, Sweden  
*patrik.ryden@math.umu.se*
95. **Lars Rönnegård**  
Dalarna University, Sweden  
*lrn@du.se*
96. **José Sánchez**  
Chalmers Institute of Technology, Sweden  
*sanchezj@chalmers.se*
97. **Busenur Sarica**  
Marmara University, Turkey  
*busenur.sarica@marmara.edu.tr*
98. **Timo Schmid**  
Freie Universität Berlin, Germany  
*Timo.Schmid@fu-berlin.de*
99. **Beste Hamiye Sertdemir**  
Dokuz Eylul University, Turkey  
*beste.sertdemir@deu.edu.tr*
100. **Mahendran Shitan**  
Universiti Putra Malaysia, Malaysia  
*sarasmahen@gmail.com*
101. **Martin Singull**  
Linköping University, Sweden  
*martin.singull@liu.se*
102. **Ramidha Srihera**  
Thammasat University, Thailand  
*srihera\_r@yahoo.com*
103. **Muni S. Srivastava**  
University of Toronto, Canada  
*srivasta@utstat.toronto.edu*
104. **Natalia Stepanova**  
Carleton University, Canada  
*nstep@math.carleton.ca*
105. **Nan-Cheng Su**  
National Taipei University, Taiwan  
*sunc@gm.ntpu.edu.tw*
106. **Müjgan Tez**  
Marmara University, Turkey  
*mtez@marmara.edu.tr*
107. **Yongge Tian**  
Central University of Finance and Economics, Beijing, China  
*yongge.tian@gmail.com*

108. **Annika Tillander**  
Karolinska Institutet, Sweden  
*annika.tillander@ki.se*
109. **Vicenc Torra**  
IIIA-CSIC, Spain  
*vtorra@iiia.csic.es*
110. **Özlem Türksen**  
Ankara University, Turkey  
*turksen@ankara.edu.tr*
111. **Marju Valge**  
University of Tartu, Estonia  
*marju.valge@ut.ee*
112. **Mattias Villani**  
Linköping University, Sweden  
*mattias.villani@gmail.com*
113. **Julia Volaufova LSU**  
Health Sciences Center School of Public Health, New Orleans, USA  
*jvolau@lsuhsc.edu*
114. **Dietrich von Rosen**  
Swedish University of Agricultural Sciences and Linköping University, Sweden  
*dietrich.von.rosen@slu.se*
115. **Patrik Waldmann**  
Linköping University, Sweden  
*Patrik.Waldmann@liu.se*
116. **Wan-Lun Wang**  
Feng Chia University , Taiwan  
*wlunwang@fcu.edu.tw*
117. **Bertil Wegmann**  
Linköping University, Sweden  
*bertil.wegmann@gmail.com*
118. **Josef Wilzén**  
Linköping University, Sweden  
*josef.wilzen@liu.se*
119. **Secil Yalaz Toprak**  
Dicle University , Turkey  
*secilyalaz@gmail.com*
120. **Takayuki Yamada**  
Nihon University, Japan  
*yma801228@gmail.com*
121. **Xiangfeng Yang**  
Linköping University, Sweden  
*xiangfeng.yang@liu.se*
122. **Silvelyn Zwanzig**  
Uppsala University, Sweden  
*zwanzig@math.uu.se*
123. **M. Revan Özkale**  
Cukurova University, Turkey  
*mrevan@cu.edu.tr*

# Index

- Agresti, A., 23  
Ahmed, S. E., 23  
Akay, K. U., 41  
Akdeniz-Duran, E., 33  
Alam, K., 66  
Alam, M. M., 34  
Alin, A., 37  
Arendacká, B., 63  
Askin, Ö. E., 34, 54  
Atkinson, A. C., 35  
Aydogdu, S., 36
- Bělašková, S., 85  
Bachratá, A., 30  
Bageta, R., 71  
Beyaztas, U., 37  
Bilton, P., 55  
Bin-Mohsin, B., 86  
Björk, R., 72  
Blaha, O., 29  
Bogacka, B., 37  
Bouillon, T., 37  
Brzezińska J., 87  
Buyuklu, A. H., 54
- Caiado, C., 38  
Carlos Pina, J., 71  
Carvalho, F., 38  
Chaiyawan, P., 90  
Chambers, R., 74  
Chamsranoi, A., 90  
Chen, T. L., 52  
Chvosteková, M., 87  
Cook, R. D., 24
- Dai, D., 39, 50  
Djehiche, B., 64  
Dobler, D., 39  
Donevska, S., 40  
Drygas, H., 41
- Eklund, A., 82
- Emiroglu, E., 41  
Eygi Erdogan, B., 42
- Fačevicová, K., 43  
Felleki, M., 44  
Fišerová, E., 40, 45, 85  
Filipiak, K., 45  
Filzmoser, P., 30, 40, 51, 56  
Firuzan, E., 75  
Fonseca, M., 46, 67  
Forkman, J., 47
- Güler, N., 88  
Goldstein, M., 47  
Guler, H., 48  
Guo, S., 34  
Gupta, A. K., 91  
Gupta, S., 25
- Hao, C., 49  
Harman, R., 30  
Haslett, S., 49  
Hauke, J., 50  
Hilbert, A., 64  
Holgersson, T., 50  
Hrůzová, K., 51  
Hron, K., 40, 43, 51, 56  
Huang, S. Y., 52  
Hung, H., 52
- Iakovenko, O., 59  
Imori, S., 53  
Inan, D., 34, 54
- Jörnsten, R., 55, 73  
Jauhiainen, A., 55  
Johnson, C. R., 50  
Jones, B., 37  
Jones, G., 55  
Jonsson, L., 63
- Kalivodova, A., 56

---

Karlsson, E., 57  
 Klein, D., 45, 58  
 Kohn, R., 68  
 Koldanov, P. A., 58  
 Kollo, T., 80  
 Kosar, C., 48  
 Kulishov, S., 59  
 Kuriki, S., 65  
 Kurnaz, F. S., 30  
 Kurtoğlu, F., 60  
  
 LaConte, S., 82  
 LaMotte, L. R., 25, 29  
 Leiva, R., 58, 71  
 Lesaffre, E., 25  
 Lesnic, D., 86  
 Li, Y., 61  
 Liang, Y. L., 61  
 Lin, T. H., 89  
 Lin, T. I., 62, 81  
 Liu, Y., 26  
 Lourenço, V. M., 70  
  
 Münnich, R., 74  
 Magnusson, M., 63  
 Markiewicz, A., 63  
 Mathew, T., 26  
 Mejza, I., 65  
 Mejza, S., 65  
 Mexia, J. T., 38  
 Monteiro, A., 70  
  
 Najim, J., 27  
 Nassar, H., 64  
 Ngaruye, I., 64  
 Noiyo, M., 90  
 Nzabanita, J., 65  
  
 Ostrowski, T., 50  
 Ozawa, K., 65  
  
 Pauly, M., 39  
 Pavlenko, T., 72, 73, 76, 77  
 Pielaszkiewicz, J., 31  
 Prus, M., 32  
 Przystalski, M., 67  
 Puntanen, S., 68, 88  
  
 Quiroz, M., 68  
  
 Rönnegård, L., 72  
 Ramírez, V. P., 69  
  
 Ricker, M., 69  
 Rodrigues, P. C., 70  
 Roy, A., 45, 46, 58, 71  
 Rydén, P., 72  
  
 Sánchez, J., 55, 73  
 Schmid, T., 74  
 Schwabe, R., 32  
 Sertdemir, B. H., 75  
 Shitan, M., 75  
 Simsek, G. G., 36  
 Singull, M., 28, 31, 57  
 Soeny, K., 37  
 Srihera, R., 90  
 Srivastava, M. S., 28  
 Stepanova, N., 76  
 Su, N. C., 91  
 Sudhir, P., 66  
  
 Taylan, P., 91  
 Tez, M., 78, 79, 91  
 Thavaneswaran, A., 75  
 Tian, Y., 29  
 Tillander, A., 77  
 Todorov, V., 51  
 Toprak, S. Y., 78  
 Torra, V., 78  
 Tsai, M. H., 89  
 Tu, I. P., 52  
 Turksen, O., 79  
 Tutalar, H. I., 78  
 Tzavidis, N., 74  
  
 Uysal, E., 91  
  
 Valge, M., 80  
 Vazifedan, T., 75  
 Villani, M., 63, 68, 82  
 Volaufova, J., 29  
 von Rosen, D., 31, 49, 53, 61, 69, 80  
 von Rosen, T., 49, 61  
  
 Waldmann, P., 81  
 Wang, W. L., 62, 81  
 Wegmann, B., 82  
 Wu, P. S., 52  
  
 Žežula, I., 58  
 Zmysłony, R., 38  
 Zwanzig, S., 83  
  
 Özkale, M. R., 83







