



11TH ISC2019

INTERNATIONAL
STATISTICS
CONGRESS

4-8 OCTOBER
2019

VOGUE HOTEL BODRUM



TÜRK İSTATİSTİK DERNEĞİ



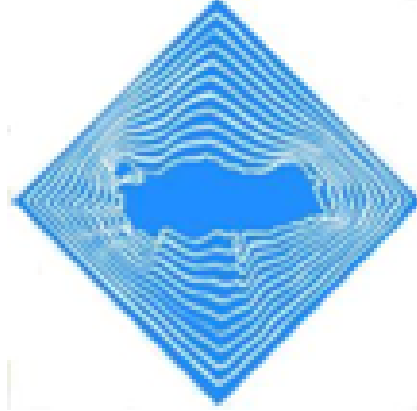
ABSTRACTS BOOK

istkon2019.giresun.edu.tr

The Congress Sponsored By



Giresun University



Turkish Statistical Association

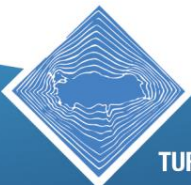


TÜBİTAK

**The Scientific and Technological
Research Council of Turkey**



**Central Bank of the Republic of
Turkey**



PREFACE

Dear Congress Participant and Dear Reader of this Book,

This preface intends to be personal. It presents how I see the way leading close to the time when this book just about goes to press

It was two years ago in the previous congress that was held at Ankara University when I first met and approached Prof. Ayşen APAYDIN to ask her for being a part of and somehow on top position of the organizing committee for the one we organized on between 4 October and 8 October 2019 in Bodrum-MUĞLA, TURKEY. With no hesitation she politely accepted me for such position. I gratefully thank her for that.

Thereafter I began to build a strong team for the congress by first calling my dearest friend Prof. Erol EĞRİOĞLU who was in fact a generous and a great man to establish a team who could handle every part of the event from Giresun University. The next day I approached a colleague of mine, Prof. Özge AKKUŞ, from next door to ask her support in such organization. She was absolutely positive and told me that she would love and enjoy to be in every part of the organization with no regret. The same day I needed another back and called one of my best colleagues, Assoc. Prof. Akay ÜNVAN who is still working at Yıldırım Beyazıt University from Ankara to ask for that. He also accepted my invitation thankfully with no hesitation as well. Prof. APAYDIN who was chairing the congress also built her team from some members of Turkish Statistical Association. That's how the organization team were established to be reality.

Since from 2011 on the meeting has been no more organized with the Turkish Statistical Association individually, so it should be backed by other institutions in Turkey. This one the 11th is the first to be organized with many experienced researchers from different variety of institutions in Turkey that have indeed made the congress success.

We had around 300 submissions and 255 of them including the keynote talks were happily presented during the congress so there was about 20% withdrawal for some reasons. We were also 293 participants at the congress including the organization committee members. Around 20% of the participants did attend from abroad from 27 different countries. Around 50% of the abstracts are from the Statistics field whereas the rest come from other field of disciplines.

We were working together with Prof. EĞRİOĞLU and his team gratefully until a month ago when he had to leave for visiting an institution in the UK to increase his level of academic career. Luckily, I and the other members of the organizing team have worked with the other committee members of Giresun University having no problem at all. Actually, I realized how his team from distance can be that success. They enormously did a great job. I would especially like to express my sincere thanks to Assoc. Prof. Ufuk YOLCU and Assoc. Prof. Eren BAŞ for their extraordinary work from the beginning to present. I sincerely thank Assist. Prof. Ali Zafer DALAR for taking his time to prepare the templates and design the official website of the congress. I have also been grateful to work with the staff members of Statistics Department, Giresun University. My great thanks go to Assist. Prof. Fatma Zehar DOĞRU who did her best to lead and organize her colleagues during the congress.

I greatly appreciate Prof. AKKUŞ for her endless support. I indeed owe her big time. I am lucky to be working with her in the same department at Mugla Sıtkı Koçman University. Many thanks go to Res. Assist. Selman Mermi who is a PhD student of Prof. AKKUŞ. He has worked hard in every step of the congress and still dealing with some other staff too.



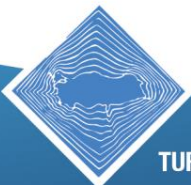
Gratefully thanks go to Prof. APAYDIN for chairing and letting us lead the congress with full authority. She has made the congress with full of success.

The organisation has taken us for about a year and we are still working on the publication process. During all this time I have had to ignore my family. At all I gratefully appreciate my wife Pinar and my daughter Alara Ecem for not complaining and being such patience. I owe them many many thanks.

We the organizing committee seriously and sincerely thank all participants for their invaluable contributions that have made the congress success in very high level.

We honestly have enjoyed every moment preparing and bringing this book to you. Hope sincerely to meeting and working with you in other events. Take care of yourself and have a nice day (24/10/2019). ☺

Atila GÖKTAŞ
Co-Chair of ISC2019
Muğla Sıtkı Koçman University
Department of Statistics
Muğla-TURKEY



HONORARY COMMITTEE

Prof. Dr. Yılmaz Can, Rector of Giresun University

Prof. Dr. Süleyman Günay, Hacettepe University

Prof. Dr. Mehmet Aydın Erar, Mimar Sinan University of Fine Arts

ADVISORY BOARD

Assoc. Prof. Dr. Sinan Saraçlı, Afyon
Kocatepe University

Prof. Dr. Mehmet Yılmaz, Ankara University

Asst. Prof. Dr. Mükerrerrem Bahar Başkır, Bartın
University

Assoc. Prof. Dr. Hamit Mirtagioğlu, Bitlis
Eren University

Prof. Dr. Güzin Yüksel, Çukurova University

Prof. Dr. Aylin Alın, Dokuz Eylül University

Prof. Dr. Kıvanç Aksoy, Eskişehir Osmangazi
University

Prof. Dr. Onur Köksoy, Ege University

Prof. Dr. Berna Yazıcı, Eskişehir Technical
University

Prof. Dr. Sinan Çalık, Fırat University

Prof. Dr. Hasan Bal, Gazi University

Prof. Dr. Erol Eğrioğlu, Giresun University

Prof. Dr. Özgür Yeniay, Hacettepe University

Prof. Dr. Münevver Turanlı, İstanbul Ticaret
University

Prof. Dr. Türkan Erbay Dalkılıç, Karadeniz
Technical University

Prof. Dr. Sevgi Yurt Öncel, Kırıkkale
University

Assoc. Prof. Dr. Birsen Eygi Erdoğan,
Marmara University

Prof. Dr. Ayşen Dener Akkaya, Middle East
Technical University

Prof. Dr. Gülay Başarır, Mimar Sinan Güzel
Sanatlar University

Prof. Dr. Dursun Aydın, Muğla Sıtkı Koçman
University

Asst. Prof. Dr. Ahmet Pekgör, Necmettin
Erbakan University

Prof. Dr. Yüksel Terzi, Ondokuz Mayıs
University

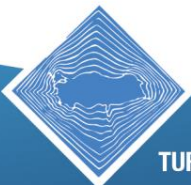
Prof. Dr. Mustafa Küçüköğüt, Selçuk
University

Assoc. Prof. Dr. Nesrin Alkan, Sinop
University

Assoc. Prof. Dr. Muhammet Bekçi, Sivas
Cumhuriyet University

Asst. Prof. Dr. Hüseyin Albayrak, Süleyman
Demirel University

Prof. Dr. Gülhayat Gölbaşı ŞİMŞEK, Yıldız
Technical University



SCIENTIFIC BOARD

Adil Baykasoğlu, Dokuz Eylül University,
TURKEY

Aladdin Shamilov, Eskişehir Technical
University, TURKEY

Ali Hakan Büyüklü, Yıldız Technical
University, TURKEY

Ali İhsan Genç, Çukurova University,
TURKEY

Asma Gul, Shaheed Benazir Bhutto Women
University, PAKISTAN

Atakan Gerger, Continuous Improvement &
Lean, PFW GmbH, GERMANY

Atalay Çağlar, Pamukkale University,
TURKEY

Birdal Şenoğlu, Ankara University, TURKEY

Brahma Deo, Indian Institute of Technology
Bhubaneswar, Odisha, INDIA

Buğra Saraçoğlu, Selçuk University, TURKEY

Burcu Hüdaverdi Üçer, Dokuz Eylül
University, TURKEY

Canan Hamurkaroğlu, Karabuk University,
TURKEY

Cem Kadılar, Hacettepe University, TURKEY

Cemal Atakan, Ankara University, TURKEY

Cemil Çolak, İnönü University, TURKEY

Dariush Akbarian, İslamic Azad University,
IRAN

Dragan Vukmirovic, Belgrade University,
Belgrade, SERBIA

Duru Karasoy, Hacettepe University,
TURKEY

Ebru Çağlayan Akay, Marmara University,
TURKEY

Efendi Nasiboğlu, Dokuz Eylül University,
TURKEY

Eleanor Stillman, Sheffield University, UK

Emel Çankaya, Sinop University, TURKEY

Ergun Karaağaoğlu, Hacettepe University,
TURKEY

Esin Avcı, Giresun University, TURKEY

Esin Firuzan, Dokuz Eylül University,
TURKEY

Fatma Noyan Tekeli, Yıldız Technical
University, TURKEY

Femin Yalçın, İzmir Katip Çelebi University,
TURKEY

Feridun Taşdan, Western Illinois University,
USA

Fikri Akdeniz, Çağ University, TURKEY

Fikri Gökpinar, Gazi University, TURKEY

Gökhan Silahtaroğlu, Medipol University,
TURKEY

Güçkan Yapar, Dokuz Eylül University,
TURKEY

Gül Ergün, Hacettepe University, TURKEY

Güneş Çetin Gerger, Celal Bayar University,
TURKEY



Hadi Alizadeh Noughabi, Birjand University,
IRAN

Halil Aydoğdu, Ankara University, TURKEY

Hamparsum Bozdoğan, Tennessee University,
USA

Hamza Erol, Mersin University, TURKEY

Hamza Gamgam, Gazi University, TURKEY

Hasan Bal, Gazi University, TURKEY

Hatice Hicret Özkoç, Muğla Sıtkı Koçman
University, TURKEY

Hülya Bayrak, Gazi University, TURKEY

Hüseyin Tatlıdil, Hacettepe University,
TURKEY

Hon Keung Tony Ng, Southern Methodist
University, USA

İlker Ercan, Uludag University, TURKEY

İnci Batmaz, Middle East Technical
University, TURKEY

İpek Deveci Karakoç, Dokuz Eylül University,
TURKEY

İsmail Erdem, Başkent University, TURKEY

İsmihan Bayramoğlu, İzmir Economy
University, TURKEY

Kamil Feridun Türkman, Lisbon University,
PORTUGAL

Lovorka Gotal Dmitrovic, University North,
University Centre Varazdin, CROATIA

M. Akif Bakır, Gazi University, TURKEY

Maria Antonia Amaral Türkman, Lisbon
University, PORTUGAL

Mehmet Aksaraylı, Dokuz Eylül University,
TURKEY

Mehmet Aydın Erar, Mimar Sinan University
of Fine Arts, TURKEY

Mehmet Mert, Akdeniz University, TURKEY

Meral Çetin, Hacettepe University, TURKEY

Minoo Aminnejad, Razi University, IRAN

Murat Alper Başaran, Alanya Alaaddin
Keykubat University, TURKEY

Müjgan Tez, Marmara University, TURKEY

Mohammad Qamarul Islam, Middle East
Technical University, TURKEY

Nalan Cinemre, Mimar Sinan University of
Fine Arts, TURKEY

Narayanaswamy Balakrishnan, McMaster
University, CANADA

Nesrin Alkan, Sinop University, TURKEY

Nimet Yapıcı Pehlivan, Selçuk University,
TURKEY

Nurgül Okur Bekar, Giresun University,
TURKEY

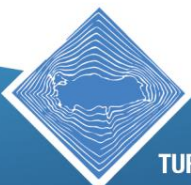
Olca Arslan, Ankara University, TURKEY

Ömer Gebizlioğlu, Kadir Has University,
TURKEY

Özlem Alpu, Eskişehir Osmangazi University,
TURKEY

Özlem Ege Oruç, Dokuz Eylül University,
TURKEY

Özlem Gürünlü Alma, Muğla Sıtkı Koçman
University, TURKEY



Philippe Castagliola, Université de Nantes,
FRANCE

Rama Shanker, Eritrea Institute of Technology
University, AFRICA

Reza Arabi Belaghi, Tabriz University, IRAN

Roland Fried, Technical University of
Dortmund, GERMANY

Sadullah Sakallıoğlu, Çukurova University,
TURKEY

Selahattin Güriş, Marmara University,
TURKEY

Selahattin Kaçıranlar, Çukurova University,
TURKEY

Selma Gürler, Dokuz Eylül University,
TURKEY

Serpil Aktaş Altunay, Hacettepe University,
TURKEY

Serpil Cula, Başkent University, TURKEY

Sevil Bacanlı, Hacettepe University, TURKEY

Sukhdev Singh, Indian Institute of Technology
Patna, INDIA

Süleyman Günay, Hacettepe University,
TURKEY

Sylwia Gwozdziwicz, Jacob of Paradies
University, POLAND

Şule Şahin, Liverpool University, UK

Şüheda Güray, Başkent University Ankara,
TURKEY

Turhan Menteş, Hacettepe University,
TURKEY

Uğur Eliyi, Dokuz Eylül University,
TURKEY

Vedide Rezan Uslu, Samsun 19 Mayıs
University, TURKEY

Veysel Yılmaz, Eskişehir Osmangazi
University, TURKEY

Vilda Purutçuoğlu, Middle East Technical
University, TURKEY

Wali Khan Mashwani, KOHAT University,
PAKISTAN

Yaprak Arzu Özdemir, Gazi University,
TURKEY

Yeliz Mert Kantar, Eskişehir Teknik
University, TURKEY

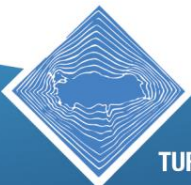
Yılmaz Akdi, Ankara University, TURKEY

Zafer Küçük, Karadeniz Technical University,
TURKEY

Zehra Muluk, Başkent University, TURKEY

Zeki Yıldız, Eskişehir Osmangazi University,
TURKEY

Zeynep Filiz, Eskişehir Osmangazi University,
TURKEY



ORGANIZING COMMITTEE

Ayşen Apaydın, Turkish Statistical Association

Atila Gökteş, Mugla Sıtkı Koçman University, TURKEY

Erol Eğrioğlu, Giresun University, Turkey

Özge Akkuş, Mugla Sıtkı Koçman University, TURKEY

Süzülay Hazar, Turkish Statistical Association

Yüksel Akay Ünvan, Ankara Yıldırım Beyazıt University, TURKEY

A. Sevtap Kestel, Turkish Statistical Association

Ali Zafer Dalar, Giresun University, TURKEY

Nebahat Bozkuş, Giresun University, TURKEY

Ilgım Yaman, Giresun University, TURKEY

Cem Koçak, Hitit University, TURKEY

Derya Erdoğan, Giresun University, TURKEY

Eren Baş, Giresun University, TURKEY

Fatma Zehra Doğru, Giresun University, TURKEY

Furkan Başer, Turkish Statistical Association

Gürol İlhan, Turkish Statistical Association

İsmet Temel, Turkish Statistical Association

Onur Toka, Turkish Statistical Association

Özge Çağcağ Yolcu, Giresun University, TURKEY

Özlem Türkşen, Turkish Statistical Association

Selman Mermi, Mugla Sıtkı Koçman University, TURKEY

Serpil Cula, Turkish Statistical Association

Ufuk Yolcu, Giresun University, TURKEY

CONFERENCE CO-CHAIRS

Atila Gökteş, Mugla Sıtkı Koçman University, Turkey

Erol Eğrioğlu, Giresun University, Turkey

Özge Akkuş, Mugla Sıtkı Koçman University, Turkey

Süzülay Hazar, Turkish Statistical Association

Yüksel Akay Ünvan, Ankara Yıldırım Beyazıt University, Turkey

CONFERENCE CHAIR

Ayşen Apaydın, Turkish Statistical Association

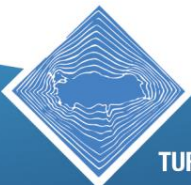
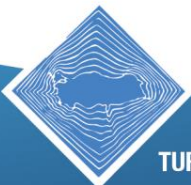


TABLE of CONTENTS

KEYNOTES.....	I
PROGRAMME.....	IX
KEYNOTES' TALKS.....	1
ABSTRACTS.....	11
CONTENTS.....	12
POSTERS.....	256
AUTHOR INDICES.....	259



INVITED SPEAKERS



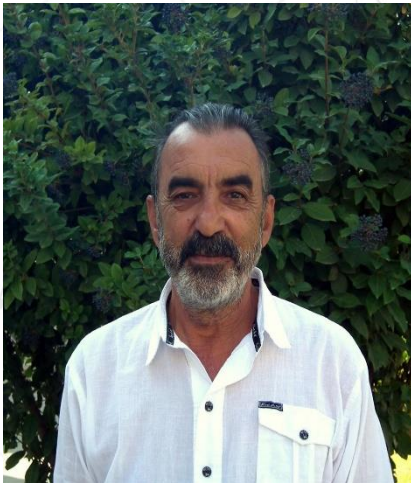
Andrzej Sokolowski

Department of Statistics, Cracow University of Economics,
Poland

Title of Speech: Composite Indicators – History and Present

Short CV

Prof. Andrzej Sokolowski is a Professor of Statistics at Cracow University of Economics, Poland. He received his PhD in Economics in Cracow University of Economics in Poland. His research interests are mathematical statistics, cluster analysis, multidimensional analysis, medical statistics, time series analysis, data analysis and data mining and statistical software. He is a member of following scientific organizations: Polish Biometric Society, Polish Statistical Society, Data Analysis and Classification – Section of the Polish Statistical Society (Vice-President), International Federation of Classification Societies (Council member), Polish Academy of Science, Cracow Branch, Statistics and Demography Commission, international conference on higher education (member of the Steering Committee). He is an editorial board member on Nowotwory- Journal of Oncology, Emergo – Journal of Transforming Economies and Societies, Zarządzanie Publiczne (Public Governance).



Güçkan Yapar

Department of Statistics, Dokuz Eylül University, Turkey

Title of Speech: M-Competitions: ATA Method and Others

Short CV

He is a Professor in the Department of Statistics at the Dokuz Eylül University. He completed his MS. at Carnegie Mellon University and received his Ph. D. from Middle East Technical University. In recent years he has focused on better techniques for modelling time-series data and obtaining better forecasts. His other research interests are risk analysis and actuarial models.



Hak-Keung Lam

Department of Informatics, King's College London, United Kingdom

Title of Speech: Interval Type – 2 Fuzzy System and Its Applications

Short CV

H. K. Lam received the B.Eng. (Hons.) and Ph.D. degrees from the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, in 1995 and 2000, respectively. During the period of 2000 and 2005, he worked with the Department of

Electronic and Information Engineering at The Hong Kong Polytechnic University as Post-Doctoral Fellow and Research Fellow respectively. He joined as a Lecturer at King's College London in 2005 and is currently a Reader.

His current research interests include intelligent control, computational intelligence and machine learning. He has authored/co-authored over 340 publications (3 monographs, 2 edited book, 9 editorials, 6 book chapters, 203 journal papers and 122 conference papers).

He has served as a program committee member, international advisory board member, invited session chair and publication chair for various international conferences and a reviewer for various books, international journals and international conferences. He was an associate editor for IEEE Transactions on Fuzzy Systems (2009-2018) and is an associate editor for IEEE Transactions on Circuits and Systems II: Express Briefs, IET Control Theory and Applications, International Journal of Fuzzy Systems, Neurocomputing and Nonlinear Dynamics; and guest editor for a number of international journals. He is on the editorial board of Journal of Intelligent Learning Systems and Applications, Journal of Applied Mathematics, Mathematical Problems in Engineering, Modelling and Simulation in Engineering, Annual Review of Chaos Theory, Bifurcations and Dynamical System, The Open Cybernetics and Systemics Journal, Cogent Engineering and International Journal of Sensors, Wireless Communications and Control. He is an IEEE senior member and a highly cited researcher.

He is a coeditor of two edited volumes: Control of Chaotic Nonlinear Circuits (World Scientific, 2009) and Computational Intelligence and Its Applications (World Scientific, 2012), and author/coauthor of three monographs: Stability Analysis of Fuzzy-Model-Based Control Systems (Springer, 2011), Polynomial Fuzzy Model Based Control Systems (Springer, 2016) and Analysis and Synthesis for Interval Type-2 Fuzzy-Model-Based Systems (Springer, 2016).



Hamza Erol

Department of Computer Engineering, Mersin University,
Turkey

Title of Speech: Mixture Distribution Models With Variable
Weights: Reliability and Risk Modeling In Complex Chaotic
Systems

Short CV

Bachelor's degree from Middle East Technical University, Department of
Mathematics in 1989.

Master's degree from Çukurova University, Department of Mathematics in 1991. Supervisor Prof.Dr. Fikri Akdeniz.

Ph.D degree from Çukurova University, Department of Mathematics in 1995. Supervisor Prof.Dr. Fikri Akdeniz.

In the years 2005 - 2011; He worked as a Professor in the Department of Statistical Information Systems at the
Department of Statistics, Çukurova University.

In the years 2011-2014; He worked as a professor at Abdullah Gül University, Software Engineering Department.

In 2017, He began to work as a professor at Department of Computer Engineering, Mersin University.

He is currently the Head of the Department of Computer Engineering at Mersin University.

His research interests include: Data Mining, Artificial Intelligence, Artificial Brain, Algorithms and Computational
Theory in the Basic Field of Computer Science and Engineering.



Mu-Yu Chen

Department of Information Management, National Taichung University of Science and Technology, Taiwan

Title of Speech: Generative Adversarial Networks: Future Has Coming Now

Short CV

Dr. **Mu-Yen Chen** is a Professor of Information Management at National Taichung University of Science and Technology, Taiwan. He received his PhD from Information Management from National Chiao-Tung University in Taiwan. His current research interests include artificial intelligent, soft computing, bio-inspired computing, data mining, deep learning, context-awareness, machine learning, and healthcare, with more than 100 publications in these areas. He has served as Associate Editor of *Journal of Information Processing Systems* and *International Journal of Social and Humanistic Computing* while he is an editorial board member on several SCI journals. In addition, he has also co-edited 12 special issues in International Journals (e.g. Computers in Human Behavior, Applied Soft Computing, Soft Computing, Information Fusion, Neurocomputing, Journal of Medical and Biological Engineering, The Electronic Library, Library High Tech).



Jie Chen

Biostatistics and Data Science, Augusta University, USA

Title of Speech: Statistical Approaches for Retrospective Detection of Multiple Change Points with Applications

Short CV

Dr. Jie Chen is Professor of Statistics, Program Director and Division Chief of Biostatistics & Data Science in Department of Population Health Sciences at Medical College of Georgia of Augusta University. Dr. Chen's research interests include, but not limited to, statistical change point analysis, applied statistics, statistical inference, model selection criteria, statistics in bioinformatics, biostatistics, statistical modeling of high throughput genomic data and biomedical data. She is the leading author of the book entitled "Parametric Statistical Change Point Analysis: With Application to Genetics, Medicine, and Finance" (Birkhäuser, 2nd edition 2012). She has been a member of American Statistical Association (ASA) Caucus of Statistics Academic Representatives. Dr. Chen was elected an ASA Fellow in 2014 and is currently Editor-in-Chief of the Journal of Applied Statistics (a Taylor & Francis journal).



Saleha Habibullah

Department of Statistics, Kinnaird College for Women,
Pakistan

Title of Speech: Establishment of an International Consortium of Statisticians for the Achievement of Sustainable Development Goals

Short CV

Prof. Dr. Saleha Naghmi Habibullah is one of the most outstanding female statisticians in Pakistan. Her entire professional career is that of Kinnaird College For Women, one of the most prestigious higher education institutions in the country. As far back as in 1985-86, Dr. Habibullah commenced a data-based project with her students that turned out to be the harbinger of a series of an ongoing series of projects and programs aimed at the uplift and enhancement of statistical education in Pakistan. These included small-scale sample surveys, statistical exhibitions, statistical competitions and training workshops. The Kinnaird College Statistics Forum for the Enlightenment of Students, Teachers, Innovators, Veterans, Amateurs and Learners (KC Statistics FESTIVAL) attracted a large number of teachers, students and practitioners of statistics from various cities and towns of Pakistan. More recently, in her capacity as Director Research at Kinnaird, Dr. Habibullah introduced and implemented a variety of initiatives for the development of the research culture at this esteemed institution. She has presented a large number of papers in numerous international conferences all over the world and is the recipient of national and international prizes and awards. In 2018, she has been selected as the educational ambassador of a highly prestigious American Statistical Association. As well, during the year 2019, she has served Vice President of the International Association for Statistical Education (IASE). As far as her research interests are concerned, her broad area is of research is distribution theory and her focus is on an interesting class of distributions that posses a remarkable property leading to better-fitting models.



Tim O'Brien

Department of Mathematics and Statistics, Loyola University
Chicago, USA

Title of Speech: Analysis of Nonlinear Longitudinal Data:
Statistical Modelling and Experimental Design

Short CV

Timothy O'Brien completed his PhD in environmental and biostatistics at NCSU in 1993 followed by post-docs in INRA (in France) and Universitaet Augsburg (in Germany). He joined Loyola University Chicago in US in 1998 and is now professor and graduate program director, where he has a joint appointment in Loyola's Department of Mathematics and Statistics and Institute of Environmental Sustainability (IES). In 2009, Dr. O'Brien won Loyola's Master Teacher Award, and, since 2013, he has led five MSc Applied Statistics students in Loyola Research Experience for Master's Programs Fellowships. His current research comprises:

- (1) providing scientists with practical (optimal) experimental design strategies (one result of this research was a 2019 joint publication with a recent MSc Applied Statistics graduate)
- (2) drug and similar compound synergy modelling (one result of this research a 2017 joint publication with IES faculty and student)
- (3) nonlinear and pharmacokinetic modelling, and
- (4) longitudinal data analysis and finite mixture models.



Wali Khan Mashwani

Department of Mathematics, Kohat University of Science & Technology, Pakistan

Title of Speech: Ensemble Strategy Based Evolutionary Algorithm for Global Optimization Problems

Short CV

Wali Khan Mashwani has received M.Sc. degree in Mathematics from University of Peshawar, Khyber Pakhtunkhwa, Pakistan in 1996 and Ph.D. in Mathematics from University of Essex, UK, in 2012. He is currently working as Associate Professor of Mathematics and Director of the Institute of Numerical Sciences, Kohat University of Science & Technology (KUST), Khyber Pakhtunkhwa, Pakistan. His main research interests are Evolutionary Computation, Hybrid Evolutionary Multiobjective Algorithms, and Decomposition Based Evolutionary Methods for Multi-objective Optimization, Mathematical Programming, Numerical Analysis and Statistical Models. Dr. Mashwani has more than seventy academic articles in peer-reviewed international journals and conference proceedings to his credit. He remained as a Chief organizer of Seven Conferences sponsored by Higher Education of Pakistan. He has been awarded NRP with ID: 5892 as PI and Start-Up grant as Co-PI by HEC, Pakistan. Dr. Mashwani is working as an active member of project, "Efficient Hybrid Meta-heuristic Algorithms for Solving Optimization Problems" under the research group "Hybrid Meta-heuristic Computing Research Group King Khalid University (KKU)" awarded by the Deanship of Scientific Research, KKU. 2017-2018. He has visited Hacettepe University under the Project ID 1543 awarded by Scientific Research Coordination Unit, 06800, Beytepe-Anakra, Turkey, in 2017 for about one month.

PROGRAMME

4th October 2019 -Friday

08:30-10:30

REGISTRATION

10:30-12:00

OPENING CEREMONY--ISTANBUL HALL

12:00 - 14:00

LUNCH TIME

14:00-14:40

INVITED SPEAKER—ISTANBUL HALL

Composite Indicators – History and Present

Andrzej Sokolowski

14:40-15:20

INVITED SPEAKER—ISTANBUL HALL

M4 Competition: Ata Method and Others

Güçkan Yapar

15:20 - 15:40

COFFEE BREAK

15:40-17:00

ISTANBUL HALL

Chair: Lovarka Gotal Dmitrovic (ENG)

Optimal Bandwidth Selection with Density Functional Estimation for Contaminated Data

Celal Aydın, Necla Gündüz

A new Robust Ridge Parameter Based on Search Method for Linear Regression Model

Atila Göktas, Özge Akkuş, Aykut Kuvat

What kind of patterns underlie the ratings students give their professors? A Modern Machine Learning Analysis

Necla Gündüz, Ernest Fokoue

The Returns to College Education and Subject of Degrees in Turkey: A Cross Sectional Analysis on 2008-2017 Period

Hakan Ulucan, Cemil Çiftçi

15:40-17:00

KAPADOKYA/CAPPADOCIA HALL

Chair: Agnieszka Kulawik (ENG)

Multivariate Intuitionistic Fuzzy Inference System for Time Series Prediction

Ozge Cagcag Yolcu, Erol Egrioglu, Eren Bas, Ufuk Yolcu

A Novel Time Series Forecasting Model Based on Single-Valued Neutrosophic Sets

Ufuk Yolcu, Eren Bas, Erol Egrioglu, Ozge Cagcag Yolcu

15:40-17:00

Deep Intuitionistic Fuzzy Time Series Forecasting Method Based on Long Short Term Memory

Cem Koçak, Erol Egrioglu, Eren Bas

Panel Data Analysis

Yüksel Akay Ünvan

EFES/EPHESUS HALL

Chair: Ayşen Dener Akkaya (ENG)

The impact of ICT penetration on financial development: An International Cross-Country Analysis

Mei-Se Chien, Chien-Chiang Lee, Chih-Yang Cheng, Meta Ayu Kurniawati

Determinants of Trade Flows in Turkey

Bilgin Bari

A Longitudinal Data Analysis to Examine the Environmental Kuznets Curve and the European Union Enlargement: The Case of Carbon Monoxide

İsmail Onur Baycan

Identification of Wealthy Households from the Residential Property Price Index Database for Sample Selection for Household Surveys

Evren Ceritoğlu, Ozlem Ardic Sevinc

15:40-17:00

NEMRUT HALL

Chair: Ebru Özgür Güler (TR)

A Specific Application of Comparison of OECD Countries with Credit Risk Swap Premiums with Quantile Regression and Multiple Regression Analysis Techniques

İrem Yanık, Hüseyin Tatlıdil

Bibliometric Analysis of Five Journals with The Highest Impact Factor In The Field of Statistics Using Citespace

Fatma Yardibi, Mehmet Ziya Fırat

A Comparison of Variety of Ridge Parameters by Means of Their Distributions and MSE Values

Selman Mermi, Atila Göktaş, Özge Akkuş

15:40 – 17:00

PAMUKKALE/HIERAPOLIS HALL

Chair: Hüseyin Güler (ENG)

Finite Mixtures of Skew Laplace Normal Distributions with Random Skewness

Fatma Zehra Doğru, Olcay Arslan

Convex Combination of Some Estimators: A Comparative Study

Buatikan Mirezi, Selahattin Kaçiranlar, Hüseyin Güler

15:40 – 17:00

Type-1 Penalized Regression Functions

Nihat Tak, Deniz İnan

Mismatch of Ridge Logistic Regression and Support Vector Machine:
Financial Distress Estimation

Ömer Utku Erzengin

GOBEKLI TEPE HALL

Chair: Özlem Türker Bayrak (ENG)

Does Gender Diversity, Institutional Ownership and Capital Structure Affect
Firm Performance? An Indonesia Real Estate and Property Firms Analysis
Rani Raharjanti

Synoptic climatology of Black Sea-effect snowfall events in Istanbul, Turkey
Hakki Baltacı, Bulent Oktay Akkoyunlu

Regularized Elimination for Variable Selection in Categorical Partial Least
Squares

Maryam Sadiq, Tahir Mehmood, Muhammad Aslam

Estimation of AR(1) Model Having Generalized Logistic Disturbances

Özlem Türker Bayrak, Ayşen Dener Akkaya

15:40 – 17:00

HEVSEL BAHÇELERİ/HEVSEL GARDENS HALL

Chair: Deniz Türsel Eliyi (TR)

Analysis of the Efficiency of Venture Capital Investment Trusts Traded in
Borsa İstanbul: A Proposal about Regulating Negative Data
M. Esra Atukalp

The Changes in Last 10 Years of the Marriage and Divorce Statistics Based
on Regions of Level 2 According Statistical Regional Units Classification
(NUTS) in Turkey and Age Groups: Examples 2008 and 2018
Fatma Sevinç Kurnaz

Classification with Two-Stage Correlation-Based Attribute Selection on the
Big Data platform

Muhammet Sinan Başarslan, Fatih Kayaalp

17:20 – 19:00

POSTER PRESENTATIONS – ISTANBUL HALL

The Effects of Time Interval on Test-Retest Reliability
İlker Ercan, Fatma Ezgi Can, Guven Ozkaya, Mehmet Onur Kaya

Penalized Power Properties of Computation Methods for the Generalized F-
test

Berna Yazıcı, Mustafa Çavuş

5th October 2019 -Saturday

09:00-09:40

INVITED SPEAKER—ISTANBUL HALL

Statistical Approaches for Retrospective Detection of Multiple Change Points with Applications

Jie Chen

09:40-10:20

INVITED SPEAKER—ISTANBUL HALL

Interval Type-2 Fuzzy System and its Applications

Hak-Keung Lam

9:00 – 10:20

POSTER PRESENTATIONS - HEVSEL BAHÇELERİ/ HEVSEL GARDENS HALL

The Effects of Time Interval on Test-Retest Reliability

İlker Ercan, Fatma Ezgi Can, Guven Ozkaya, Mehmet Onur Kaya

Penalized Power Properties of Computation Methods for the Generalized F-test

Berna Yazıcı, Mustafa Çavuş

10:20 - 10:40

COFFEE BREAK

10:40 – 12:20

ISTANBUL HALL

Chair: Mohammad Abiad (ENG)

Liu and Restricted Liu Estimators in Linear Measurement Error Models

Gülesen Üstündağ Şiray

Application of Statistical Methodologies for Customer Churn: A case study of Kuwait Telecommunication Sector

Mohammad Abiad, Sorin Ionescu

A Proposed Algorithm for Predictive Maintenance Using Statistics

Mohammad Abiad, Sorin Ionescu

A Study on Diagnostics for Two-level Multilevel Regression Models

Yasin Altinisik, Emel Cankaya

Parameter estimation of capital asset pricing model through Quantile regression: An application on Karachi Stock Exchange

Amna Riaz, Shahid Kamal, Nauman Riaz Chaudhry, Mohsin Riaz

10:40 – 12:20

HEVSEL BAHÇELERİ/ HEVSEL GARDENS HALL

Chair: Müjgan Tez (ENG)

A Mathematical Programming Procedure for Examining Biasing Parameters of Two Stage Two Parameter Estimator

Selma Toker, Nimet Özbay

Point Estimation for Chen Distribution based on Ranked Set Sampling
Muhammet Çavdar, Buğra Saraçoğlu

An improper form of Weibull distribution for competing risks analysis with Bayesian approach
Ahmad Reza Baghestani, Fatemeh Sadat Hosseini-Baharanchi

Developing a First Order Two Parameter Estimator for Generalized Linear Model
Selma Toker, Gülesen Üstündağ Şiray, Muhammad Qasim

Adaptive MCMC for Generalized Method of Moments with Many Moment Conditions
Masahiro Tanaka

10:40 - 12:20

EFES/EPHESUS HALL

Chair: Adnan Karaibrahimoğlu (ENG)

Optimal Bus-Garage Allocations and Garage Capacity Analysis Using a Map Interface
Uğur Eliyi, Selcen Coşkun

Re-evaluation of Subtypes of Positional OSAS by Clustering Algorithms
Adnan Karaibrahimoglu, Onder Ozturk

Detection of Related Species by Denoising Phylogenetic Trees
Nebahat Bozkus, Stuart Barber

Examination of Aggregating Clustering Validation Indexes on the Football Players Performance Data Set
Serhat Emre Akhanlı, Oya Kalaycıoğlu

Performance of Variance ratio tests in presence of heteroskedasticity: Application to stock returns data in Pakistan
Nuzhat Aftab, Muhammed Sohail

10:40 – 12:20

NEMRUT HALL

Chair: Mária Bohdalová (ENG)

The Effect of Uncertainty Shocks on Global Housing Markets
Mei-Se Chien, Nur Setyowati

Examination of Cereal Production, Cereal Yield and Land Under Cereal Production in Turkey
Melis Temiz, Nilsu Coskun, Semra Cakmak, Ozlem Senvar

The Dependent-Conditional Value-at-Risk Forecast for Aggregate Risk Models
Bony Josaphat, Khreshna Syuhada

Point Cloud Matching with a Vertex Weighted Graph Kernel

Pelin Kaya, Mehmet Ali Balci, Ömer Akgüller

Testing the Hypothesis of Environmental Kuznets Curve with Renyi and Tsallis Mutual Informations

Elif Tuna, Atif Evren

10:40 – 12:20

PAMUKKALE/HIERAPOLIS HALL

Chair: Atakan Gerger (TR)

Depth Based Nonparametric Multivariate Tolerance Regions

E. Burcu Mamak Ekinci, İhsan Karabulut

Predicting Academic Performance of Foreign Students via Machine Learning Algorithms

Özge Altıntaş, Furkan Başer, Ömer Kutlu, Ayşen Apaydın

Investigation of Possibility of Being Poor in Turkey with Probit Analysis

Tuba Yakıcı Ayan, Nurdan Değirmenci

Investigation of Item Bias of the Foreign Student Examination

Burcu Hasançebi, Yüksel Terzi, Zafer Küçük

Asymmetric Open Vehicle Routing Problem with Heterogeneous Fleet and Time Limitations

Nilay Yaren Çelikoklar, Uğur Eliyi

10:40 – 12:20

GOBEKLI TEPE HALL

Chair: Murtaza Abbas Rizvi (ENG+TR)

A Gravity Model: Maritime Transport Costs and Their Impact on Economy

Nurten Çalışkan, Burcu Özge Özaslan Çalışkan

Big Data Analytics: Future Perspective and Scope in IT Applications

M A Rizvi, Anu Mangal

Big Data Analytics for Green Supplier Selection Under Sustainable Conditions

Abdullah Yıldızbaşı, Yağmur Arıöz

Evaluate the Performances of Restaurants and Hotels Listed in BIST 100 Under Balanced Scorecard Perspective Via Plithogenic Decision Making Approach

Çağlar Karamaşa

12:20 - 14:00

LUNCH TIME

14:00 – 15:20

ISTANBUL HALL

Chair: Ayşen Dener Akkaya (ENG)

Iterative Robust Bayesian Estimation of Regression Modelling based on Ramsay-Novick Distribution

Mutlu Kaya, Emel Cankaya, Olcay Arslan

Parameter Estimation of Dirichlet Distribution Based on Entropy

Büşra Şahin, Atıf Evren

Multi-Criteria Decision Making in Site Selection for Nuclear Power Plants:

Multi-Attribute Utility Approach

Berna Tuncer, Ayşen Dener Akkaya

14:00 – 15:20

HEVSEL BAHÇELERİ/ HEVSEL GARDENS HALL

Chair: Aytaç Pekmezci (TR)

The Impact of Medical Tourism on the Turkey Economic Performance

Berhan Çoban

Automatic Determination of K Parameter Value of K-Means for Time Series Clustering

Fatma Özge Özkök, Mete Celik

Predicting Sepsis Via Unsupervised Machine Learning Using Bedside Device Data in Hospitals

Gokhan Silahtaroglu, Zehra Nur Canbolat, Kadir İdinNevin Yılmaztürk

Clustering of Vector Autoregressive Models: An Application on Air Pollution

Aytaç Pekmezci

14:00 – 15:20

EFES/EPHESUS HALL

Chair: Ayten Yiğiter (TR)

Time Series Forecasting via Dendritic Neuron Model Neural Network Trained by Particle Swarm Optimization

Ayşe Yılmaz, Ufuk Yolcu

Financial Time Series Prediction via Convolutional Neural Network

Melih Kirişçi, Özge Cagcag Yolcu

Multiplicative Neuron Model Based on Threshold Value for Forecasting Problem of Time Series

Asiye Nur Yıldırım, Eren Bas, Erol Egrioglu

A Comparison Study on Stock Exchange Forecasting

Onur Derya, Ali Zafer Dalar

14:00 – 15:20

NEMRUT HALL

Chair: Gökhan Silahtaroglu (ENG)

The Missing Link: Are Individuals with More Social Capital in Better Health? Evidence from India

Barış Alpaslan, Julide Yıldırım

Predicting Patients Future Laboratory Test Result Patterns Via A Novel Unsupervised Deep Learning Model

Gokhan Silahtaroglu, Zehra Nur Canbolat, Kadir İdin, Nevin Yilmaztürk

The Role of Social Capital in Environmental Protection Efforts: Evidence from Turkey

Julide Yıldırım, Barış Alpaslan, Erdener Emin Eker

Prediction of Failures in Plastic Extrusion Process with Deep Learning

Fatma Demircan Keskin, Ural Gökay Çiçekli, İsmail Doğukan İçli

14:00 – 15:20

PAMUKKALE/HIERAPOLIS HALL

Chair: Deniz İnan (TR)

Application of Statistical Process Control in Production Process: A Case Study in Cleaning Products Sector

Gülşen Aydın Keskin, Murat Çolak, Bengü Güngör

Evaluation and Estimation of Wind Speed with Data Mining Methods

Kadriye Ergün, Melike Sultan Karasu Aşnaz

Investigating the Effectiveness of Metaheuristic Approaches in Parameter Estimation via TOPSIS Method

Aynur Şahin, Nimet Yapıcı Pehlivan

Determination of The Number of In-Port Transportation Vehicles by Simulation Modelling

Bengü Güngör, Gülşen Aydın Keskin

14:00 – 15:20

GOBEKLI TEPE HALL

Chair: Ebru Dumlupınar (ENG)

The Effect of Different Antibiotics Use on Prognosis of Patients who Hospitalized with Urinary Tract Infection: A Retrospective Study

Sevilay Kiliçkaya, Osman Gürdal

Analyzing hospital high length of stay outliers

Pakize Yiğit

15:20 - 15:40

Receiver Operating Characteristic (ROC) Analysis in Clinical Diagnosis:
A Diabetes Case

Ebru Dumlupınar, Osman Gürdal, Yusuf Bilen

COFFEE BREAK

15:40 – 17:00

ISTANBUL HALL

Chair: Berna Yazıcı (ENG)

Considering Linear Constraints for Almon Two Parameter Ridge Estimation
Nimet Özbay, Selma Toker

Modelling Disease Transmission Scenarios Using Field Data and Heavy
Tailed Random Effects: The Zika Case

Zafer Bekiryazici, Tülay Kesemen, Mehmet Merdan, Tahir Khaniyev

Mixed Frequency Estimation by Smoothed Least Squares
Selma Toker, Nimet Özbay, Kristofer Månsson

Unit-Lindley Mixed-Effect Model for Ratio Data
Hatice Tul, Kubra Akdur

15:40 – 17:00

HEVSEL BAHÇELERİ/ HEVSEL GARDENS HALL

Chair: Mária Bohdalová (ENG)

Bitcoin Inefficiency Analysis
Mária Bohdalová, Michal Greguš

Catch-Up Performance of Turkish Manufacturing Firms and International
Trade Effect

Aslıhan Atabek Demirhan

Investigation of Effects of Some Cryptocurrencies on Bitcoin: Causality
Analysis

Yüksel Akay Ünvan

15:40 – 17:00

EFES/EPHESUS HALL

Chair: Amna Riaz (ENG)

Drought Monitoring of Marmara Region using Novel Machine Learning
Tools

Ö. Ozan Evkaya, Fatma Sevinç Kurnaz

Modeling of BACE-1 Inhibitory activity Using MLR, ANN and SVM
Techniques

Bashirulla Shaik

15:40 – 17:00

Household Expenditures Pattern on Durable and Nondurable Goods and Services: A Comparison of Multiple Regression and Neural Network Model
Khadija Fatima, Muhammad Suleman, Zahoor Ahmed, Abeeda Fatima, Amna Riaz

NEMRUT HALL

Chair: Laura Plazola (ENG)

Stochastic frontier analysis to measure the impact on the productive efficiency of lemon producers in Michoacán, through the incorporation of the economic costs of insecurity

Laura Plazola, Mariela Sandoval

Asymptotic Results for Stationary Characteristics of Random Walk with a General Interference of Chance

Ozlem Ardic Sevinc, Tahir Khaniyev

Mixed Lasso Estimator for Stochastic Restricted Regression Models

Huseyin Guler, Ebru Ozgur Guler

On the Limiting Behaviour of the Ergodic Distribution of Random Walk with a Generalized Reflecting Barrier

Başak Gever, Zülfiye Hanalioğlu

15:40 -17 :00

PAMUKKALE/HIERAPOLIS HALL

Chair: Esra Satıcı (TR)

Forecasting the Volatility of Stock Price Index

Eda Yalçın Kayacan

Using Data Mining Methods for Prediction of Women's Birth Time in Turkey

Özer Bakar, Duygu İçen

An Improved Search Engine for E-Commerce Websites Based on Word2vec Model

Bahadır Durak, Buse Dedeakayoğulları

Comparing Forecasting Accuracies of ATA and Exponential Smoothing

Beyza Cetin, İdil Yavuz

15:40 – 17:00

GOBEKLI TEPE HALL

Chair: Tahir Mehmood (ENG)

A Study on Industry 4.0 awareness of undergraduate students by using Structural Equational Modelling

Sevde Ceren Yıldız, Seniye Ümit Fırat



Good Agricultural Practices and Technical Efficiency in Chili Production in Thailand

Wirat Krasachat

The Use of Generalized Estimating Equations and Panel Causality Tests in the Evaluation of Countries' Economic

Harun Yonar, Neslihan İyit

17:00 - 17:20

COFFEE BREAK

17:20 – 18:40

ISTANBUL HALL

Chair: Aytaç Pekmezci (TR)

Bi-Responses Mars Modelling through Earth Package for Regression Type Problems

Ecevit Eyduran, Meleksen Akın, Demet Canga, Ali Eylem Celik, Sadiye Peral Eyduran

New Goodness of Fit Tests for Unit-Lindley Regression Model

Deniz Ozonur

Robust Regression Method for Parameters of Power Lindley Distribution

Berivan Çakmak, Fatma Zehra Doğru

Use of Bootstrap Aggregating (Bagging) MARS to Improve Predictive Accuracy for Regression Type Problems

Ecevit Eyduran, Demet Çanga, Harun Sevgenler, Ali Eylem Celik

17:20 – 18:40

HEVSEL BAHÇELERİ/ HEVSEL GARDENS HALL

Chair: İlker Ercan (TR)

The Relationship Between the Graph Energies and Vulnerabilities

Erva Kol, Mehmet Ali Balcı, Ömer Akgüller

On Solving the Generalized Assignment Problem via Hypergraphs

Kezban Afetsu Başdoğan, Mehmet Ali Balcı, Ömer Akgüller

The Cubic Rank Transmutation Method and Applications

Demet Sezer, Nuri Celik, Aydın Karakoca

Hierarchies in Communities of FTSE100 from the Perspective of Brexit

Serdar Can Güzel, Mehmet Ali Balcı, Ömer Akgüller

6th October 2019 -Friday

09:00 - 09:40

INVITED SPEAKER—ISTANBUL HALL

Mixture Distribution Models with Variable Weights: Reliability and Risk Modeling in Complex Chaotic Systems

Hamza Erol

09:00 – 09:40

INVITED SPEAKER— KAPADOKYA/ CAPPADOCIA HALL

Generative Adversarial Networks: Future Has Coming Now

Mu-Yen Chen

09:40 - 10:20

INVITED SPEAKER—ISTANBUL HALL

Analysis of Nonlinear Longitudinal Data: Statistical Modelling and Experimental Design

Timothy O'Brien

10:20 - 10:40

COFFEE BREAK

10:40 – 12:20

ISTANBUL HALL

Chair: Nebile Korucu Gümüsoğlu (TR)

A Comparison of Regression Analysis Methods for Beer Consumption
Duygu Dede Şener

Modified Two Parameter Estimator to Combat Multicollinearity: A Comparative Study
Selahattin Kaçıranlar, Nimet Özbay, Ecem Özkan

The Government Revenue - Expenditure Nexus: Symmetric and Asymmetric Causality Tests
Şule Nazlı Arslan, Havvanur Feyza Erdem

The Validity of Easterlin Hypothesis in Turkish Economy
Ayşenur Üçeş, Havvanur Feyza Erdem

A Modeling Study for Discrete and Continuous Variables by Using Parametric and Nonparametric Regression Analysis
Gözde Karakoç, Özlem Türkşen

10:40 – 12:20

KAPADOKYA/ CAPPADOCIA HALL

Chair: Şüheda Güray (TR)

Alignment Method: Concept and Application
Gizem Uyumaz, Gözde Sırgancı, Alperen Yandı

Different Methods for Testing of Measurement Invariance with Many Groups
Gizem Uyumaz, Gözde Sırgancı, Alperen Yandı

Institutional Quality and Income Inequality: Panel Cointegration Analysis in 12 Developing Countries
Seyit Ali Miçooğulları, Süleyman Değirmen

Examine of Mathematical Anxiety of Engineering Students by Factor Analysis
Şüheda Güray, Tülay Korkmaz Devrani

The Impact of Perceptions of Social Media Advertisements on Advertising Value and Brand Awareness: Research on Generation Y Instagram Users
İbrahim Halil Efendioğlu, Yakup Durmaz

10:40 – 12:20

EFES/EPHESUS HALL

Chair: Henryk Gacki (ENG)

Data Dimensionality Reduction based on Variables Clustering
Arnaldo Rabello de Aguiar Vallim-Filho, Pedro Okidoi, Leandro Augusto da Silva, Marco Vinicius Bhering de Aguiar Vallim, Leilton Santos da Silva

Neutrosophic Linear Regression Analysis of Construction Accidents in Turkey
Şüra Toptancı, Nihal Erginel, Ilgın Acar

An Application of Cluster Analysis Based on Competitive Intensity and Market Turbulence of Firm
Sinem Peker

The Huber Functions and Their Application to a Classification Problem
Henryk Gacki, Agnieszka Kulawik

Classification of Artworks with Deep Learning Methods: Ebru and Suminagashi
Aysin Tasdelen, Sebnem Ozdemir

10:40 – 12:20

HEVSEL BAHÇELERİ/ HEVSEL GARDENS HALL

Chair: Süleyman Günay (TR)

Determining the Most Risky Production Area By Using Aggregated Criteria Importance Weights
Gülin Feryal Can, Pelin Toktaş

Creating Risk Based Quality Management Procedures in A Company Producing Medical Devices
Selin Sezer, Pelin Toktaş

Vehicle Headway Modeling with Exponentiated Weibull Distribution using Ranked Set Sampling Method
Busra Sevinc, Selma Gürler, Melek Esemem

10:40 – 12:20

Control Charts for the Compound Weibull Exponential Distribution under Truncated Life Tests

Canan Hamurkaroğlu, Ayten Yiğiter

Statistical Analysis of Wheat Flour Quality Parameters

Barış Ergül, Arzu Altın Yavuz

PAMUKKALE/HIERAPOLIS HALL

Chair: Selin Özen (ENG)

Performance Evaluation of Non-Life Insurance Companies via Grey Relational Analyse and Analytic Hierarchy Process

Murat Kırkağaç, Övgücan Karadağ Erdemir

A Numerical Approximation of Inversion Method to Generating Random Numbers

Orhan Kesemen, Buğra Kaan Tiryaki, Ayşe Pak

10:40 – 12:20

GOBEKLI TEPE HALL

Chair: Qais Mustafa Abdulqader (ENG)

Sector-wise analysis of cardinality constraint portfolio optimization based on a hybrid approach: nonlinear neural network and genetic algorithm

Ilgın Yaman, Turkan Erbay Dalkilic

Statistical Analysis Techniques of Determination of a Key Comparison Reference Value in International Measurement Comparisons

Sevda Kaçmaz, Hilmi Ege Subat

Forecasting tourist arrivals with economic regressors: A grid search on combinations of different individual models

Emrah Gülay, Korkut Bekiroğlu, Okan Duru

ARIMA Model for Monthly Crude Oil Forecasting in Iraq

Qais Mustafa Abdulqader

12:20 - 14:00

LUNCH TIME

14:00 – 14:40

INVITED SPEAKER – ISTANBUL HALL

Establishment of an International Consortium of Statisticians for the Achievement of Sustainable Development Goals

Saleha Habibullah

14:40 – 15:20

INVITED SPEAKER – ISTANBUL HALL

Ensemble Strategy Based Evolutionary Algorithm for Global Optimization Problems

Wali Khan Mashwani



15:20 - 15:40

COFFEE BREAK

15:40 – 17:00

ISTANBUL HALL

Chair: Betül Kan Kılınç (ENG)

Debt and Economic Growth Nexus in the US: Evidence from Asymmetric Causality

Nebile Korucu

Investigation of the Effects of Research Supports on Academic Performance of Turkish Universities by Using Correlation, Regression and Path Analyses
Esra Satıcı, Bulut Tunca, Özge Akkuş

Lifelong Learner: Question-Answering Chatbot

Ezgi Özen, Adil Alpkoçak

Generalization and Compounding Statistical Distributions, Goals and Benefits: A Review

Ehsan Fayyazi, Serpil Kiliç Depren

15:40 – 17:00

KAPADOKYA/CAPPADOCIA HALL

Chair: Hakan Kızılkaya (TR)

Alpha Power Transformed Two-Parameters Rayleigh Distribution: Properties and Inference

Hakan Kızılkaya, Cenker Biçer

Statistical Errors in Specialization in Medicine Thesis

Ayşegül Yabacı, Fatma Ezgi Can, Fisun Kesin, Robab Ahmadian, Ilker Ercan

Log - Two - Parameter Xgamma Distribution: Properties and Inference

Esra Kızılkaya, Cenker Biçer

Statistical Evaluation of The Cultural Difference Problem Experienced by Refugees in Health Services

Arzu Altın Yavuz, Barış Ergül, Şenay Kızılkum

15:40 – 17:00

EFES/EPHESUS HALL

Chair: Pelin Toktaş (TR)

An Approach of Feature Selection Consensus for the Gas Sensors' Drift Data
Çağatay Berke Erdaş, Gülnur Begüm Ergün, Selda Güney

The Ranking of Cities for Household Consumption Expenditures in Turkey with RANKING-KEMIRA-M Approach

Pelin Toktaş

Evaluation of Feature Reduction and Classification Methods for Activity Recognition

Çağatay Berke Erdaş, Didem Ölçer

15:40 – 17:00

Some Properties Provided by Fourier Transform with Distribution Function in Statistics

Seçil Yalaz

HEVSEL BAĞÇELERİ/ HEVSEL GARDENS HALL

Chair: M. Akif Bakır (ENG)

Examining the response of olfactory stimulus in EEG signals

Begüm Kara Gülay, Neslihan Demirel, Alper Vahaplar, Çağdaş Güdücü

Research of Stochastic Properties of a Population Model with Fokker Planck Equation Under Allee Effect

Çağatay Eskin, Özgür Gültekin

A Satellite Images Based Sampling Approach in An African City

Bayram Samet Şahin, Hakan Demirbüken, M. Akif Bakır

Data Science and Big Data Analytics Education - Serbian perspective

Dragan Vukmirovic, Tijana Comic, Zoran Radojicic

15:40 – 17:00

PAMUKKALE/HIERAPOLIS HALL

Chair: Uğur Eliyi (ENG)

Preemptive Goal Programming for the Daily Shipment Planning of an Automotive Lubricant Company

Deniz Türsel Eliyi, Sel Özcan, Uğur Eliyi

The Use of AHP-TOPSIS Method in Technopark Preferences of Technology Companies: Istanbul Case

İsmail Durak, Hakan Murat Arslan, Yahya Özdemir

Two-stage solution framework for a workload balancing and workforce scheduling problem: A comparative study

Sel Özcan, Uğur Eliyi, Deniz Türsel Eliyi

Machine Learning Based Feature Extraction for Determination of Balance Disorders

Serhat İkizoglu, Ezgi Demir, Batuhan Atasoy

15:40 – 17:00

GOBEKLI TEPE HALL

Chair: M. Aydın Erar (ENG)

Supervised Learning Based Apparel Recommendation System for Online Shopping System

Selver Ezgi Küçükbay, Nafiye Kübra Turhan

Predicting Loss Aversion Behaviour with Machine Learning Methods

Omur Saltik, Ridvan Soyu, Suleyman Degirmen

17:00 – 17:20

17:20 – 18:40

17:20 – 18:40

17:20 – 19:00

Predicting Bubbles in the S&P 500 Index with Macroeconomic Indicators: A Machine Learning Approach

Fatma Başoğlu Kabran, Kamil Demirberk Ünlü

Credit Scoring with Using Alternative Model Techniques

Zeynep Bal, M. Aydın Erar

COFFEE BREAK

ISTANBUL HALL

Chair: Alessandro Barbiero (ENG)

Algebraic Methods for the Problem of Fitting Concentric Ellipses to Digitized Observations

Ali-Al Saradqah, Lorenzo Rulli

Two Way Mixed Design: Symmetry Assumption and Missing Data

Sevgi Demircioğlu

Approximation of Continuous Random Variables for Evaluating Reliability of Complex Stress-Strength Models

Alessandro Barbiero

Adaptation of the r-d class prediction method to linear mixed models

Özge Kuran

KAPADOKYA/CAPPADOCIA HALL

Chair: Ayşegül Tuş (TR)

Multi Criteria Group Decision Making with Neutrosophic Sets

Ayşegül Tuş, Gülin Zeynep Öztaş, Abdullah Özçil, Tayfun Öztaş, Esra Aytaç Adalı

The relationship between the Stock of Car Loans and the growth of the Automotive Sector from a times series analysis

Burcu Barlas

Neutrosophic ARAS Method for Selection Problems

Esra Aytaç Adalı, Tayfun Öztaş, Abdullah Özçil, Gülin Zeynep Öztaş, Ayşegül Tuş

EFES/EPHESUS HALL

Chair: Oğuz Akpolat (ENG)

A New Heuristic Algorithm for the Label Printing Problem

Emre Çankaya, Ali Ekici, Okan Örsan Özener

Curve Fitting for Environmental Measurements by Artificial Intelligence Techniques

Ayşe Nurcan Kapucu, Ceyhan Kapucu, Oğuz Akpolat

Estimating Optimal Value for the Shrinkage and Biasing Parameters in Liu-Type Logistic Regression based on Particle Swarm Optimization

Nuriye Sancar, Deniz Inan

A traffic accident prediction model based on ANN

Beyzanur Cayir Ervural

Improvement of official statistics by applying the concept of Big Data - Case study: Serbia

Tijana Comic, Dragan Vukmirovic, Aleksandar Djokovic

17:20 – 18:40

HEVSEL BAHÇELERİ/ HEVSEL GARDENS HALL

Chair: İlker Ercan (TR)

Circular Analyses of Dates on Patients with Gastric Carcinoma

Adnan Karaibrahimoğlu, Seren Ayhan

Importance of Fractal Structures in Modern Biology and Discussion of Forward Kolmogorov Equation with Fractional Derivative Approach

Esra Yazıcıoğlu, Özgür Gültekin

Examination of Incidences of HIV Infection and Tuberculosis in Some Selected European Countries with Statistical Process Control

Özlem Toluk, İlker Ercan, Halis Akalın

A Two-Population Model with Renewal Process for Measuring Longevity Basis Risk

Selin Özen, Şule Şahin

17:20 – 18:40

PAMUKKALE/HIERAPOLIS HALL

Chair: Ali Acaravcı (TR)

The Effects of Financial Innovation on Economic Growth in The Banking Sector: Case of Regions and Provinces

Aynur Karaçoban, Ömür Saltık, Süleyman Değirmen

Net External Position, Financial Development, and Banking Crisis

Aytül Ganioglu

The Gravity Model of International Immigration: Evidence from OECD Countries

Ali Acaravcı, Seyfettin Artan, Cemalettin Kalaycı, Sinan Erdoğan

A Promotion Planning Model with Product Substitution and Promotion Fatigue

Dilek Tüzün Aksu, Bahadır Durak, Buse Dedeakayoğulları

17:20 – 18:40

GOBEKLI TEPE HALL

Chair: Adem Doğaner (TR)

Human Identification with Deep LSTM by Using Biometric Walking Patterns

Çağatay Berke Erdaş, Didem Ölçer

Different Approaches to Reducing Bias in Classification of Medical Data by Ensemble Learning Methods

Adem Doğaner

Reduce Unexpected Airline Diverts: Modelling with Neural Network

Hazal Berve Doğan, Derya Gözen, Hakan Maral, Tahir Hanalioğlu

Financial Distress Prediction Using Support Vector Machine with Optimal Parameter and Feature Selection

Seyyide Doğan, Şenol Altan

7th October 2019 - Saturday

9:00 – 10:20

ISTANBUL HALL

Chair: Özlem Türkşen (ENG)

Distribution the DJSTOXXE50 Index

Yasemin Ulu

Spatial Statistical Analysis for Geodetic Points of a Simulated Fault Plane through Geographical Information Systems

Özlem Türkşen

Assessment of Longevity Risk on Pension Funds: Credibility Approach

Bükre Yıldırım Külekci, A. Sevtap Selçuk-Kestel

The Effects of a Promotion on Store Traffic, Conversion, and Customer Expenditures: Evaluation with a Prediction Approach

Ignacio E. Inostroza-Quezada, Leonardo D. Epstein

9:00 – 10:20

KAPADOKYA/CAPPADOCIA HALL

Chair: Melike Kaya Bahçecitapar (ENG)

Nonparametric Statistical Methods for the Comparison of Two Treatments in the Presence of Competing Risks

Jingjing Lyu, Jinbao Chen, Yawen Hou, Zheng Chen

Modification of S Test Statistics for Randomized Block Designs with Autocorrelated Longitudinal Data

Melike Kaya Bahçecitapar

Novel Approaches for Choosing Optimum Cut-Off Value

Naime Meric Konar

Prognostic model Animated Film Festival - VAFI

Lovorka Gotal Dmitrovic, Hrvoje Selec, Mario Lesina



9:00 – 10:20

EFES/EPHESUS HALL

Chair: Ebru Özgür Güler (TR)

Classification of OECD Countries with Socio-Economic and Education Indicators

Ebru Özgür Güler, Dilek Veysikarani

Theoretical and Practical Comparison of the Frequency Based Classification Methods: Education Sample

Cengiz Gazeloğlu, Sinan Saraçlı

ATAforecasting: An R Package for Automatic Seasonal or Non-seasonal Time Series Forecasting

Hanife Taylan Selamlar, Ali Sabri Taylan

Ranking of Turkish Universities Based on Academic Performance by Using Fuzzy c-means Clustering Algorithm

Ayşen Apaydın, Fatih Tank, Furkan Başer

9:00 – 10:20

NEMRUT HALL

Chair: Oğuz Akpolat (ENG)

An Experimental Design Application in Chemistry for Optimization of Assay Conditions for tRNA

Oğuz Akpolat

Smoothing Techniques in Solving Absolute Value Equations

Nurullah Yilmaz, Ahmet Sahiner

Approximated Convex Bezier Curve for the Continuation Global Optimization Approach

Ahmet Sahiner, Idris A. Masoud Abdulhamid, Nurullah Yilmaz

A New Smoothing Technique for Total Variation Minimization in Image Denoising

Ahmet Sahiner, Shehab A. Ibrahim

9:00 – 10:20

PAMUKKALE/HIERAPOLIS HALL

Chair: Atakan Gerger (ENG)

An Exact Approach for A Dynamic Workforce Scheduling Problem

Okan Örsan Özener

Use of Six Sigma Method as a Process Improvement Techniques

Atakan Gerger, Ali Rıza Firuzan

Data Mining in Finance: Comparing Decision Trees and Association Rules for Stock Price Prediction in BIST 100

G.Ataman, S.Kahraman

9:00 – 10:20

GOBEKLI TEPE HALL

Chair: İlhan Usta (TR)

Energy Consumption and Economic Growth: Panel Simultaneous Equation Models

Pınar Hayaloğlu, Seyfettin Artan, Selim Koray Demirel

Cumulant Generating Function of Sample Minimum of Order Statistics from Geometric Distribution

Aslıhan Demir, Ayşe Metin Karakaş, Sinan Çalık

Volatility Measurement Entropy Methods

Ayşe Metin Karakaş, Aslıhan Demir, Sinan Çalık

Entropy Approach for Wind Energy

Ayşe Metin Karakaş, Aslıhan Demir, Sinan Çalık

9:00 – 10:20

HEVSEL BAHÇELERİ/HEVSEL GARDENS HALL

Chair: M. Akif Bakır (ENG)

Combination of Forecasting Benchmarks on M4 Competition Data Sets

Tuğçe Ekiz Yılmaz

Financial Data Analysis by Exponential Smoothing and ATA Method

Selma Şalk, Güçkan Yapar

Exchange Rate Volatility and Stock Returns: A case of China and Turkey

Tao Wang, Özlem Sayılır

Examining the Financial Success of Stocks by Using Data Mining Techniques

Serpil Altınırmak, Ezgi Demir, Batuhan Atasoy, Çağlar Karamaşa

10:20 – 10:40

COFFEE BREAK

10:40 – 12:20

ISTANBUL HALL

Chair: Deniz İnan (ENG)

The p-Splines Model in Predicting PM10 Concentrations

Betül Kan-Kilinc, Huruy D. Asfha

Evaluating Negative Binomial Regression and Zero Truncated Negative Binomial Regression for Different Dispersion Parameters

Eda Çakmak, Ergun Karaağaoğlu

Binary Particle Swarm Optimization as a Detection Tool for Influential Subsets in Linear Regression

Deniz İnan, Gokce Deliorman

J-Inflated longitudinal regression models for analyzing responses with finite support

Mojtaba Ganjali, Taban Baghfalaki

10:40 – 12:20

KAPADOKYA/CAPPADOCIA HALL

Chair: İlhan Usta (TR)

Opposition Based Particle Swarm Optimization for Feature Selection
Merve Özdemir, Ümmügülsüm Can, Berna Kiraz

Artificial Bee Colony Algorithm for Estimating Parameters of a Flexible Weibull Distribution
Ilhan Usta, Volkan Yesildere

A Study on Loss Probability of a Blocked Queueing System
Erdoğan Yücesoy, Vedat Sağlam, Murat Sağır

Parameter Estimation of the Kies Distribution by Artificial Bee Colony Algorithm
Ilhan Usta, Volkan Yesildere

Data Mining of SILC Data: Turkey Case
Olgun Özdemir, İbrahim Demir

Wage Dynamics According to Income Group in Turkey: 2014-2017 Period
Şenay Üçdoğru Birecikli, Nilgün Acar Balaylar, Hamdi Emeç

Modelling the Mediator Effect of Demographic Features on Clients' Deposit via Structural Equation Modelling
Sinan Saraçlı, Cengiz Gazeloğlu

10:40 – 12:20

NEMRUT HALL

Chair: Brahma Deo (ENG)

The Impacts of Oil Price Fluctuations on the Turkish Stock Market: Correlations, Volatility Spillover and Hedging
Vasif Abioglu

Linear and Nonlinear Unit Root Tests Application: Turkish Electricity Consumption Per Person
Fatih Çemrek, Büşra Yıldız

Noise Reduction In Chaotic Time Series For Improved Feature Selection, Abnormal Process Detection, And Process Modeling
Puneet Choudhary, Shibu Meher, Brahma Deo, Manish Goyal

Detection of Mode-Mixing Problem in EMD of Time Series Signal Data
Atacan Erdiş, M. Akif Bakır, Muhammed I Jaiteh

Investigation of The Relationship Between Turkey and Economic Growth Industrial Production Index
Fatih Çemrek, Büşra Yıldız

10:40 – 12:20

PAMUKKALE/HIERAPOLIS HALL

Chair: Latifa Ghalayini (ENG)

Estimating the Output GAP for Lebanon: A Production Function methodology

Latifa Ghalayini

Sunk Costs and Exporting Behavior: A Sectoral Analysis

Kurmaş Akdoğan, Laura M. Werner

Complexly Interwoven Stochastic and Chaotic Tendencies in the Profit-Interest Ratios at a Recent Crisis Juncture in Turkey: Possibilities for Crisis-Management

Ahmet Kara

Practical Tests of Financial Bubbles: Better Critical Values for Detecting Exuberant Behavior

Ufuk Demiroğlu, Aytül Ganioglu

Inference of Economic Performance Using Global Trade Data

Orhun Sevinç

10:40 – 12:20

GOBEKLI TEPE HALL

Chair: Özge Gündoğdu (TR)

Change Point Detection in ARMA(p,q) Process and Application on Headway Data

Almila Hacıoğlu, Aylin Alin, Süheyla Pelin Çalışkanelli, Ufuk Beyaztaş

A New Robust Unit Root Test with a Simulation Study

Hakan Türkay, Özge Gündoğdu

A Meta Analysis Approach on Work – Family Balance and Job Satisfaction

Ebru Özgür Güler, Didem Keskin

KEYNOTES' TALKS

Composite Indicators – History and Present

Andrzej Sokołowski
Cracow University of Economics, Poland

Rankings are popular both in scientific research and in everyday life. Many institutions rank different objects from the best to the worst. We have rankings of countries, universities, cities, hospitals, jobs, provinces, books, songs, actors, footballers etc. If such a ranking is based on just one variable measured in strong scale, then the task is trivial. The only thing we should decide is the direction – the bigger the better or the smaller the better. The problem is more complicated with multidimensional case, even with just two diagnostic variables.

The aim of the paper is to present and discuss issues connected with the consecutive stages of the process of constructing composite indicators.

Linear ordering methods are sometimes divided into those with benchmark and without. It can be argued that every method has some benchmark – assumed or calculated from the data.

Short review of the literature will be given and some example provided and critically discussed – such as Human Development Index or Best Jobs in USA. Finally, some recommendations are given. They are concentrated on issues connected with ordering stages such as weighting systems, normalization of variables, distributions of variables, aggregation formulas, point systems, and some new proposals like step-wise and flexible procedures.

M-Competitions: ATA Method and Others

Güçkan Yapar

Dokuz Eylül University, Department of Statistics
guckan.yapar@deu.edu.tr

Abstract

Forecasting plays a critical role for almost all of science and the “M” competitions organized by Spyros Makridakis have had an enormous influence on the field of forecasting. They focused attention on what models produced good forecasts, rather than on the mathematical properties of those models. In this presentation, a new forecasting approach, ATA method, will be introduced and the forecasting accuracy of this new forecasting method that is alternative to two major forecasting approaches: exponential smoothing (ES) and ARIMA, will be evaluated using M4 competition data set. Finally, I will present ideas on how the forecasting performance of the method can be improved so that desirable results can be obtained for the most recent of the series M4- competition.

Key Words: Exponential smoothing, forecasting, initial value, forecasting competition, smoothing parameter

Mixture Distribution Models With Variable Weights: Reliability and Risk Modeling In Complex Chaotic Systems

Hamza Erol

Mersin University, Faculty of Engineering, Department of Computer Engineering, 33110 Yenişehir Mersin – Turkey.
e-mail: herol@meu.edu.tr

Abstract

Reliability and risk modeling in complex chaotic systems is one of important topics in the area of systems engineering. A complex chaotic system is a system with some of its components having increasing risk, some of its components having constant risk or some of its components having decreasing risk. Reliability and risk modeling in complex chaotic systems uses the reliability block diagrams of complex chaotic systems. Mixture distribution models with variable weights are proposed for modeling reliability and risk of reliability block diagrams for complex chaotic systems.

Complex chaotic system's non-linear reliability and hazard functions are transformed to mixture reliability and mixture hazard functions respectively in this study. Mixture reliability and mixture hazard functions are expressed as linear functions of component functions with variable non-linear mixture weight functions. Complex chaotic system's mixture reliability and mixture hazard functions are obtained as finite sum of products of component's variable non-linear mixture weight functions with component pure reliability and hazard functions respectively. It is shown that the number of terms in complex chaotic system's mixture reliability and hazard functions are equal to the number of components in the complex chaotic system. The effect of the structure of reliability block diagram of the complex chaotic system is reflected to the component's variable non-linear mixture weight functions.

Mixture distribution models with variable non-linear mixture weight functions are used both for prediction of reliability or life times and risk of complex chaotic system and its components concurrently. The working principle and computational steps of the proposed mixture model reliability and risk analysis of reliability block diagram of a complex chaotic system were explained on an application.

Interval Type-2 Fuzzy System and its Applications

Hak-Keung Lam

Department of Informatics, King's College London, United Kingdom

This talk will be delivered in two parts while the first part is a brief introduction of fuzzy logic systems from the control point of view while the second part is about the fuzzy-logic related applications. In the first part, the fuzzy logic system will be introduced and its fundamentals, working principle and rationale will be explained in detail with the support of block diagrams and examples. The role of type-1, interval type-2 and general type-2 fuzzy sets will be discussed, and their pros and cons will be highlighted. In the second part, the applications using fuzzy-model-free and fuzzy-model-based techniques will be covered which include control of mobile robot, decision making and obstacle avoidance for robot soccer, drug administration of anesthesia, classification of epilepsy phases, balancing of inverted pendulum, torque control of bolt-tightening applications and tracking control of continuum manipulator. The contributions of the presenter made to the field, which initiated two research sub-fields underpinning the imperfect premise matching concept, membership-function-dependent analysis and interval type-2 fuzzy-model-based control systems, will be summarized.

Statistical Approaches for Retrospective Detection of Multiple Change Points with Applications

Jie Chen

Biostatistics & Data Science, DPHS, MCG, Augusta University, USA

Two key features of retrospective, statistical change point analysis are: (i) to test if change points exist in the underlying distributions from which a sequence of observations are generated and (ii) to estimate the locations of any change points that do exist. Retrospective change point detection can be done through: a hypothesis testing process, a model selection perspective, a Bayesian approach, and penalized regression, among other methods. Change point analysis has a wide range of applications in research fields, such as statistical quality control, finance, economics, climate study, medicine, genetics and genomics.

In this talk, I will present some methods that we developed for retrospectively detecting multiple change points. First, a likelihood-procedure-based test statistic was developed to detect variance or covariance change points with applications to the detection of changes in the Dow Jones Weekly stock price indices. Then, we investigated the problems of detecting multiple change points for a single profile and multiple profiles using the framework of penalized regression. Simulation studies and application on the next-generation sequencing data in genomics show that the proposed approaches can effectively identify change points of both a single profile and multiple profiles.

Generative Adversarial Networks: Future Has Coming Now

Mu-Yen Chen

Department of Information Management, National Taichung University of Science and Technology, Taiwan

GANs through a small amount of real data, generated a large amount of training data, which is an unsupervised model. Compared with the previous CNNs/RNNs, it is a supervised model, which must rely on a large amount of labeled data. GANs is a major advancement in the Neural Network. GANs uses an unsupervised learning style training that can be widely used in unsupervised learning and semi-supervised learning. It can produce clearer and more realistic samples. Compared to Variational Auto - Encoder (VAE), GANs has no variation lower bound, if the discriminator is well trained, then the generator can perfectly learn the distribution of the training samples. In other words, the GANs are gradual, but the VAE is biased. In addition, GANs is applied to some scenes, such as image style transform, super-resolution, image completion, denoising, avoiding the difficulty of loss function design, as long as there is a benchmark, directly on the discriminator, the rest handed over to the adversarial network training. But GANs has some problems, such as GANs is not suitable for dealing with discrete forms of data, such as text, GANs has unstable training, gradient disappears, mode collapse, etc. In this presentation, we will introduce the GANs concepts and its applications in many successful domains.

Establishment of an International Consortium of Statisticians for the Achievement of Sustainable Development Goals

Saleha Naghmi Habibullah

Professor of Statistics, Kinnaird College For Women, Lahore, Pakistan

saleha.habibullah@kinnaird.edu.pk, salehahabibullah@gmail.com

Abstract

Towards the end of the year 2015, the world adopted seventeen Sustainable Development Goals, the target being to build a better world for the people and the planet by the year 2030. Four years down the road, not much progress seems to have been made in this regard in developing countries. Needless to say, the solution to multifarious problems faced by the common man in any part of the world lies in evidence-based decision-making which is the basis for appropriate policy-formulation. As such, the importance of deriving **extensive** information from large-scale data-sets being produced by national and international statistical organizations leading to a variety of socio-economic indicators cannot be over-emphasized. This target can be realized effectively through the establishment of an **International Consortium of Statisticians** comprising, among others, professional statisticians, academia and MS/PhD students of Statistics belonging to under-developed, middle-income and technologically advanced countries. Higher education institutions will have an important role to play in the Consortium as no segment of the statistical community other than young, budding students enrolled in masters/doctoral programs can be expected to possess the **dynamism, enthusiasm** and **passion** that is required for spreading awareness regarding the socio-economic situation of the people. Wide-spread publicity by members of an integrated group regarding the problems and deprivations experienced by the masses will push governments toward policy formulation aimed at improving the quality of life of the people.

Key Words: Sustainable development goals, evidence-based decision-making, consortium

Analysis of Nonlinear Longitudinal Data: Statistical Modelling and Experimental Design

Timothy O'Brien

Loyola University Chicago (USA)

Longitudinal data are ubiquitous in biomedical research, economics, environmental research, psychometrics as well as many other domains, and analysis of these data present unique and farreaching challenges in applied statistical research. These data often also contain latent (hidden) cohorts/groups, which – with the aid of the EM algorithm and associated methods – can be discerned in order to help researchers in better understanding their data and underlying phenomena. Although the fields of Finite Mixture Models and Trajectory Analysis in the context of longitudinal data analysis is relatively new, controversy exists as to how best to discern these patterns and data.

This talk focuses on the larger field of estimation and design of longitudinal data, with an eye to trajectory analysis and finite mixture models in modelling nonlinear phenomena. We make connections to the linear and generalized linear cases - as well as highlighting important differences and relevant software packages.

Global Optimization Based on Hybrid Swarm Intelligence Techniques

Wali Khan Mashwani and Fouzia

Institute of Numerical Sciences,

KUST, Khyber KPK, Pakistan

E-mail: mashwanigr8@gmail.com; walikhan@kust.edu.pk

Abstract

Global optimization is the mathematical procedure to find the global solution for optimization and search problems. Optimization problems have wide application in advanced engineering design, biotechnology, data analysis, environmental management, financial planning, process control, risk management, scientific modeling, and others. Their solution often requires a global search approach. Swarm intelligence based approaches have shown good performances for dealing with various real-world problems and complicated benchmark functions (Wazir, et al. 2016). In the last two decades, several different types of swarm algorithms (SAs) have been suggested including the Bat algorithm (BA) (Yang 2014), (Kennedy and Eberhart 1995) (PSO), Firefly algorithm (FA) (Yang 2014) Ant Colony optimization (ACO) and many others (Eberhart, Shi and Kennedy 2001). All these types of swarm intelligence based algorithms are mainly inspired by the social and combined behaviors of swarm such as animal herds, bird flocking and schooling of fish (Blum and Merkle 2008).

This research thesis proposes hybrid swarm intelligence techniques for global optimization. The suggested algorithm employs Bat Algorithm (BA) (Karamanoglu, et al. 2013) as search operator in combination with practical swarm optimization (PSO) to cope with recently designed benchmark functions for the special session of the 2017 IEEE congress of evolutionary computation (CEC'17) (Awad, et al. 2017). The solutions obtained by suggested algorithm are either optimal or very close to the known best solutions for most of CEC'17 benchmark functions (Awad, et al. 2017).

Key Words: Global Optimization, Optimization Problems, Evolutionary Algorithm, Evolutionary algorithms, Nature Inspired Algorithms, Swarm Intelligence Approaches and Hybrid Evolutionary Algorithms.

Bibliography

Awad, N. H., M. Z. Ali, J. J. Liang, B. Y. Qu, and P. N. Suganthan. *Problem Definitions and Evaluation Criteria for the CEC 2017 Special Session and Competition on Single Objective Bound Constrained Real-Parameter Numerical Optimization*. Singapore: Nanyang Technological University, 2017.

Blum, Christian, and Daniel Merkle. *Swarm Intelligence: Introduction and Applications*. Springer-Verlag Berlin Heidelberg, 2008.

Eberhart, Russell C., Yuhui Shi, and James Kennedy. *Swarm Intelligence*. The Morgan Kaufmann Series in Evolutionary Computation, 2001.

Karamanoglu, Mehmet, Renbin Xiao, Zhihua Cui, and Xin-She Yang. *Swarm Intelligence and Bio-Inspired Computation: Theory and Applications*. Science Direct, 2013.

Kennedy, James, and Russell C. Eberhart. "Particle swarm optimization." *IEEE International Conference on Neural Networks. IV*. IEEE, 1995. 942–948.

Storn, Rainer, and Kenneth Price. "Differential Evolution-A Simple and Efficient Heuristic for global Optimization over Continuous Spaces." *Journal of Global Optimization* 11, no. 4 (December 1997): 341–359.

Wazir, Hamza, Muhammad Asif Jan, Wali Khan Mashwani, and Tayyaba Tabasum Shah. "A Penalty Function Based Differential Evolution Algorithm for Constrained Optimization." *The Nucleus Journal* 53, no. 1 (2016): 155-161.

Yang, Xin-She. *Introduction to Nature-Inspired Optimization Algorithms*. Elsevier, 2014.

ABSTRACTS

CONTENTS

OPTIMAL BANDWIDTH SELECTION WITH DENSITY FUNCTIONAL ESTIMATION FOR CONTAMINATED DATA ..	22
A NEW ROBUST RIDGE PARAMETER BASED ON SEARCH METHOD FOR LINEAR REGRESSION MODEL	23
WHAT KIND OF PATTERNS UNDERLIE THE RATINGS STUDENTS GIVE THEIR PROFESSORS? A MODERN MACHINE LEARNING ANALYSIS!	24
THE RETURNS TO COLLEGE EDUCATION AND SUBJECT OF DEGREES IN TURKEY: A CROSS SECTIONAL ANALYSIS ON 2008-2017 PERIOD	25
MULTIVARIATE INTUITIONISTIC FUZZY INFERENCE SYSTEM FOR TIME SERIES PREDICTION	26
A NOVEL TIME SERIES FORECASTING MODEL BASED ON SINGLE-VALUED NEUTROSOPHIC SETS.....	27
DEEP INTUITIONISTIC FUZZY TIME SERIES FORECASTING METHOD BASED ON LONG SHORT TERM MEMORY	28
MARKET BASKET ANALYSIS WITH ASSOCIATION RULES	28
THE IMPACT OF ICT PENETRATION ON FINANCIAL DEVELOPMENT: AN INTERNATIONAL CROSS-COUNTRY ANALYSIS	30
DETERMINANTS OF TRADE FLOWS IN TURKEY	31
A LONGITUDINAL DATA ANALYSIS TO EXAMINE THE ENVIRONMENTAL KUZNETS CURVE AND THE EUROPEAN UNION ENLARGEMENT: THE CASE OF CARBON MONOXIDE	32
IDENTIFICATION OF WEALTHY HOUSEHOLDS FROM THE RESIDENTIAL PROPERTY PRICE INDEX DATABASE FOR SAMPLE SELECTION FOR HOUSEHOLD SURVEYS.....	33
A SPECIFIC APPLICATION OF COMPARISON OF OECD COUNTRIES WITH CREDIT RISK SWAP PREMIUMS WITH QUANTILE REGRESSION AND MULTIPLE REGRESSION ANALYSIS TECHNIQUES	34
BIBLIOMETRIC ANALYSIS OF FIVE JOURNALS WITH THE HIGHEST IMPACT FACTOR IN THE FIELD OF STATISTICS USING CITESPACE	35
A COMPARISON OF VARIETY OF RIDGE PARAMETERS BY MEANS OF THEIR DISTRIBUTIONS AND MSE VALUES	36
FINITE MIXTURES OF SKEW LAPLACE NORMAL DISTRIBUTIONS WITH RANDOM SKEWNESS.....	37
CONVEX COMBINATION OF SOME ESTIMATORS: A COMPARATIVE STUDY	38
TYPE-1 PENALIZED REGRESSION FUNCTIONS.....	39
MISMATCH OF RIDGE LOGISTIC REGRESSION AND SUPPORT VECTOR MACHINE: FINANCIAL DISTRESS ESTIMATION.....	40
DOES GENDER DIVERSITY, INSTITUTIONAL OWNERSHIP AND CAPITAL STRUCTURE AFFECT FIRM PERFORMANCE? AN INDONESIA REAL ESTATE AND PROPERTY FIRMS ANALYSIS.....	41
SYNOPTIC CLIMATOLOGY OF BLACK SEA-EFFECT SNOWFALL EVENTS IN ISTANBUL, TURKEY	42
REGULARIZED ELIMINATION FOR VARIABLE SELECTION IN CATEGORICAL PARTIAL LEAST SQUARES	43



ESTIMATION OF AR(1) MODEL HAVING GENERALIZED LOGISTIC DISTURBANCES	44
ANALYSIS OF THE EFFICIENCY OF VENTURE CAPITAL INVESTMENT TRUSTS TRADED IN BORSA İSTANBUL: A PROPOSAL ABOUT REGULATING NEGATIVE DATA.....	45
THE CHANGES IN LAST 10 YEARS OF THE MARRIAGE AND DIVORCE STATISTICS BASED ON REGIONS OF LEVEL 2 ACCORDING STATISTICAL REGIONAL UNITS CLASSIFICATION (NUTS) IN TURKEY AND AGE GROUPS: EXAMPLES 2008 AND 2018.....	46
CLASSIFICATION WITH TWO-STAGE CORRELATION-BASED ATTRIBUTE SELECTION ON THE BIG DATA PLATFORM	47
LIU AND RESTRICTED LIU ESTIMATORS IN LINEAR MEASUREMENT ERROR MODELS	48
APPLICATION OF STATISTICAL METHODOLOGIES FOR CUSTOMER CHURN: A CASE STUDY OF KUWAIT TELECOMMUNICATION SECTOR	49
A PROPOSED ALGORITHM FOR PREDICTIVE MAINTENANCE USING STATISTICS	50
A STUDY ON DIAGNOSTICS FOR TWO-LEVEL MULTILEVEL REGRESSION MODELS	51
PARAMETER ESTIMATION OF CAPITAL ASSET PRICING MODEL THROUGH QUANTILE REGRESSION: AN APPLICATION ON KARACHI STOCK EXCHANGE	52
A MATHEMATICAL PROGRAMMING PROCEDURE FOR EXAMINING BIASING PARAMETERS OF TWO STAGE TWO PARAMETER ESTIMATOR.....	53
POINT ESTIMATION FOR CHEN DISTRIBUTION BASED ON RANKED SET SAMPLING	54
AN IMPROPER FORM OF WEIBULL DISTRIBUTION FOR COMPETING RISKS ANALYSIS WITH BAYESIAN APPROACH	55
DEVELOPING A FIRST ORDER TWO PARAMETER ESTIMATOR FOR GENERALIZED LINEAR MODEL	56
ADAPTIVE MCMC FOR GENERALIZED METHOD OF MOMENTS WITH MANY MOMENT CONDITIONS	57
OPTIMAL BUS-GARAGE ALLOCATIONS AND GARAGE CAPACITY ANALYSIS USING A MAP INTERFACE	58
RE-EVALUATION OF SUBTYPES OF POSITIONAL OSAS BY CLUSTERING ALGORITHMS.....	59
DETECTION OF RELATED SPECIES BY DENOISING PHYLOGENETIC TREES.....	60
EXAMINATION OF AGGREGATING CLUSTERING VALIDATION INDEXES ON THE FOOTBALL PLAYERS PERFORMANCE DATA SET	61
PERFORMANCE OF VARIANCE RATIO TESTS IN PRESENCE OF HETEROSKEDASTICITY: APPLICATION TO STOCK RETURNS DATA IN PAKISTAN	62
THE EFFECT OF UNCERTAINTY SHOCKS ON GLOBAL HOUSING MARKETS	63
EXAMINATION OF CEREAL PRODUCTION, CEREAL YIELD AND LAND UNDER CEREAL PRODUCTION IN TURKEY	64
THE DEPENDENT-CONDITIONAL VALUE-AT-RISK FORECAST FOR AGGREGATE RISK MODELS	65
POINT CLOUD MATCHING WITH A VERTEX WEIGHTED GRAPH KERNEL.....	66



TESTING THE HYPOTHESIS OF ENVIRONMENTAL KUZNETS CURVE WITH RENYI AND TSALLIS MUTUAL INFORMATIONS	67
DEPTH BASED NONPARAMETRIC MULTIVARIATE TOLERANCE REGIONS	68
PREDICTING ACADEMIC PERFORMANCE OF FOREIGN STUDENTS VIA MACHINE LEARNING ALGORITHMS.....	69
INVESTIGATION OF POSSIBILITY OF BEING POOR IN TURKEY WITH PROBIT ANALYSIS	70
INVESTIGATION OF ITEM BIAS OF THE FOREIGN STUDENT EXAMINATION.....	71
ASYMMETRIC OPEN VEHICLE ROUTING PROBLEM WITH HETEROGENEOUS FLEET AND TIME LIMITATIONS ..	72
A GRAVITY MODEL: MARITIME TRANSPORT COSTS AND THEIR IMPACT ON ECONOMY	73
BIG DATA ANALYTICS: FUTURE PERSPECTIVE AND SCOPE IN IT APPLICATIONS	74
BIG DATA ANALYTICS FOR GREEN SUPPLIER SELECTION UNDER SUSTAINABLE CONDITIONS.....	75
EVALUATE THE PERFORMANCES OF RESTAURANTS AND HOTELS LISTED IN BIST 100 UNDER BALANCED SCORECARD PERSPECTIVE VIA PLITHOGENIC DECISION MAKING APPROACH.....	76
ITERATIVE ROBUST BAYESIAN ESTIMATION OF REGRESSION MODELLING BASED ON RAMSAY-NOVICK DISTRIBUTION	77
PARAMETER ESTIMATION OF DIRICHLET DISTRIBUTION BASED ON ENTROPY	78
MULTI-CRITERIA DECISION MAKING IN SITE SELECTION FOR NUCLEAR POWER PLANTS: MULTI-ATTRIBUTE UTILITY APPROACH	79
THE IMPACT OF MEDICAL TOURISM ON THE TURKEY ECONOMIC PERFORMANCE	80
AUTOMATIC DETERMINATION OF K PARAMETER VALUE OF K-MEANS FOR TIME SERIES CLUSTERING	81
PREDICTING SEPSIS VIA UNSUPERVISED MACHINE LEARNING USING BEDSIDE DEVICE DATA IN HOSPITALS..	82
CLUSTERING OF VECTOR AUTOREGRESSIVE MODELS: AN APPLICATION ON AIR POLLUTION	83
TIME SERIES FORECASTING VIA DENDRITIC NEURON MODEL NEURAL NETWORK TRAINED BY PARTICLE SWARM OPTIMIZATION	84
FINANCIAL TIME SERIES PREDICTION VIA CONVOLUTIONAL NEURAL NETWORK	85
MULTIPLICATIVE NEURON MODEL BASED ON THRESHOLD VALUE FOR FORECASTING PROBLEM OF TIME SERIES	86
A COMPARISON STUDY ON STOCK EXCHANGE FORECASTING.....	87
THE MISSING LINK: ARE INDIVIDUALS WITH MORE SOCIAL CAPITAL IN BETTER HEALTH? EVIDENCE FROM INDIA.....	88
PREDICTING PATIENTS FUTURE LABORATORY TEST RESULT PATTERNS VIA A NOVEL UNSUPERVISED DEEP LEARNING MODEL.....	89
THE ROLE OF SOCIAL CAPITAL IN ENVIRONMENTAL PROTECTION EFFORTS: EVIDENCE FROM TURKEY	90
PREDICTION OF FAILURES IN PLASTIC EXTRUSION PROCESS WITH DEEP LEARNING	91



APPLICATION OF STATISTICAL PROCESS CONTROL IN PRODUCTION PROCESS: A CASE STUDY IN CLEANING PRODUCTS SECTOR	92
EVALUATION AND ESTIMATION OF WIND SPEED WITH DATA MINING METHODS	93
INVESTIGATING THE EFFECTIVENESS OF METAHEURISTIC APPROACHES IN PARAMETER ESTIMATION VIA TOPSIS METHOD	94
DETERMINATION OF THE NUMBER OF IN-PORT TRANSPORTATION VEHICLES BY SIMULATION MODELLING .	95
THE EFFECT OF DIFFERENT ANTIBIOTICS USE ON PROGNOSIS OF PATIENTS WHO HOSPITALIZED WITH URINARY TRACT INFECTION: A RETROSPECTIVE STUDY.....	96
ANALYZING HOSPITAL HIGH LENGTH OF STAY OUTLIERS	97
RECEIVER OPERATING CHARACTERISTIC (ROC) ANALYSIS IN CLINICAL DIAGNOSIS: A DIABETES CASE.....	98
CONSIDERING LINEAR CONSTRAINTS FOR ALMON TWO PARAMETER RIDGE ESTIMATION	99
MODELLING DISEASE TRANSMISSION SCENARIOS USING FIELD DATA AND HEAVY TAILED RANDOM EFFECTS: THE ZIKA CASE.....	100
MIXED FREQUENCY ESTIMATION BY SMOOTHED LEAST SQUARES.....	101
UNIT-LINDLEY MIXED-EFFECT MODEL FOR RATIO DATA.....	102
BITCOIN INEFFICIENCY ANALYSIS.....	103
CATCH-UP PERFORMANCE OF TURKISH MANUFACTURING FIRMS AND INTERNATIONAL TRADE EFFECT	104
INVESTIGATION OF EFFECTS OF SOME CRYPTOCURRENCIES ON BITCOIN: CAUSALITY ANALYSIS	105
DROUGHT MONITORING OF MARMARA REGION USING NOVEL MACHINE LEARNING TOOLS	106
MODELING OF BACE-1 INHIBITORY ACTIVITY USING MLR, ANN AND SVM TECHNIQUES.....	107
HOUSEHOLD EXPENDITURES PATTERN ON DURABLE AND NONDURABLE GOODS AND SERVICES: A COMPARISON OF MULTIPLE REGRESSION AND NEURAL NETWORK MODEL	108
STOCHASTIC FRONTIER ANALYSIS TO MEASURE THE IMPACT ON THE PRODUCTIVE EFFICIENCY OF LEMON PRODUCERS IN MICHOACÁN, THROUGH THE INCORPORATION OF THE ECONOMIC COSTS OF INSECURITY.	109
ASYMPTOTIC RESULTS FOR STATIONARY CHARACTERISTICS OF RANDOM WALK WITH A GENERAL INTERFERENCE OF CHANCE	110
MIXED LASSO ESTIMATOR FOR STOCHASTIC RESTRICTED REGRESSION MODELS	111
ON THE LIMITING BEHAVIOUR OF THE ERGODIC DISTRIBUTION OF RANDOM WALK WITH A GENERALIZED REFLECTING BARRIER	112
FORECASTING THE VOLATILITY OF STOCK PRICE INDEX.....	113
USING DATA MINING METHODS FOR PREDICTION OF WOMEN'S BIRTH TIME IN TURKEY.....	114
AN IMPROVED SEARCH ENGINE FOR E-COMMERCE WEBSITES BASED ON WORD2VEC MODEL	115
COMPARING FORECASTING ACCURACIES OF ATA AND EXPONENTIAL SMOOTHING	116



A STUDY ON INDUSTRY 4.0 AWARENESS OF UNDERGRADUATE STUDENTS BY USING STRUCTURAL EQUATIONAL MODELLING.....	117
GOOD AGRICULTURAL PRACTICES AND TECHNICAL EFFICIENCY IN CHILI PRODUCTION IN THAILAND.....	118
THE USE OF GENERALIZED ESTIMATING EQUATIONS AND PANEL CAUSALITY TESTS IN THE EVALUATION OF COUNTRIES' ECONOMIC PROFILES	119
BI-RESPONSES MARS MODELLING THROUGH EARTH PACKAGE FOR REGRESSION TYPE PROBLEMS	120
NEW GOODNESS OF FIT TESTS FOR UNIT-LINDLEY REGRESSION MODEL	121
ROBUST REGRESSION METHOD FOR PARAMETERS OF POWER LINDLEY DISTRIBUTION.....	122
USE OF BOOTSTRAP AGGREGATING (BAGGING) MARS TO IMPROVE PREDICTIVE ACCURACY FOR REGRESSION TYPE PROBLEMS.....	123
THE RELATIONSHIP BETWEEN THE GRAPH ENERGIES AND VULNERABILITIES	124
ON SOLVING THE GENERALIZED ASSIGNMENT PROBLEM VIA HYPERGRAPHS	125
THE CUBIC RANK TRANSMUTATION METHOD AND APPLICATIONS.....	126
HIERARCHIES IN COMMUNITIES OF FTSE100 FROM THE PERSPECTIVE OF BREXIT	127
A COMPARISON OF REGRESSION ANALYSIS METHODS FOR BEER CONSUMPTION	128
MODIFIED TWO PARAMETER ESTIMATOR TO COMBAT MULTICOLLINEARITY: A COMPARATIVE STUDY.....	129
THE GOVERNMENT REVENUE - EXPENDITURE NEXUS: SYMMETRIC AND ASYMMETRIC CAUSALITY TESTS....	130
THE VALIDITY OF EASTERLIN HYPOTHESIS IN TURKISH ECONOMY	131
A MODELING STUDY FOR DISCRETE AND CONTINUOUS VARIABLES BY USING PARAMETRIC AND NONPARAMETRIC REGRESSION ANALYSIS	132
ALIGNMENT METHOD: CONCEPT AND APPLICATION	133
DIFFERENT METHODS FOR TESTING OF MEASUREMENT INVARIANCE WITH MANY GROUPS.....	134
INSTITUTIONAL QUALITY AND INCOME INEQUALITY: PANEL COINTEGRATION ANALYSIS IN 12 DEVELOPING COUNTRIES.....	135
EXAMINE OF MATHEMATICAL ANXIETY OF ENGINEERING STUDENTS BY FACTOR ANALYSIS	136
THE IMPACT OF PERCEPTIONS OF SOCIAL MEDIA ADVERTISEMENTS ON ADVERTISING VALUE AND BRAND AWARENESS: RESEARCH ON GENERATION Y INSTAGRAM USERS	137
DATA DIMENSIONALITY REDUCTION BASED ON VARIABLES CLUSTERING	138
NEUTROSOPHIC LINEAR REGRESSION ANALYSIS OF CONSTRUCTION ACCIDENTS IN TURKEY	139
AN APPLICATION OF CLUSTER ANALYSIS BASED ON COMPETITIVE INTENSITY AND MARKET TURBULENCE OF FIRM.....	140
THE HUBER FUNCTIONS AND THEIR APPLICATION TO A CLASSIFICATION PROBLEM	141



DETERMINING THE MOST RISKY PRODUCTION AREA BY USING AGGREGATED CRITERIA IMPORTANCE WEIGHTS	142
CREATING RISK BASED QUALITY MANAGEMENT PROCEDURES IN A COMPANY PRODUCING MEDICAL DEVICES	143
VEHICLE HEADWAY MODELING WITH EXPONENTIATED WEIBULL DISTRIBUTION USING RANKED SET SAMPLING METHOD.....	144
CONTROL CHARTS FOR THE COMPOUND WEIBULL EXPONENTIAL DISTRIBUTION UNDER TRUNCATED LIFE TESTS.....	145
STATISTICAL ANALYSIS OF WHEAT FLOUR QUALITY PARAMETERS	146
PERFORMANCE EVALUATION OF NON-LIFE INSURANCE COMPANIES VIA GREY RELATIONAL ANALYSE AND ANALYTIC HIERARCHY PROCESS	147
A NUMERICAL APPROXIMATION OF INVERSION METHOD TO GENERATING RANDOM NUMBERS	149
SECTOR-WISE ANALYSIS OF CARDINALITY CONSTRAINT PORTFOLIO OPTIMIZATION BASED ON A HYBRID APPROACH: NONLINEAR NEURAL NETWORK AND GENETIC ALGORITHM	150
STATISTICAL ANALYSIS TECHNIQUES OF DETERMINATION OF A KEY COMPARISON REFERENCE VALUE IN INTERNATIONAL MEASUREMENT COMPARISONS	151
FORECASTING TOURIST ARRIVALS WITH ECONOMIC REGRESSORS: A GRID SEARCH ON COMBINATIONS OF DIFFERENT INDIVIDUAL MODELS.....	152
ARIMA MODEL FOR MONTHLY CRUDE OIL FORECASTING IN IRAQ.....	153
DEBT AND ECONOMIC GROWTH NEXUS IN THE US: EVIDENCE FROM ASYMMETRIC CAUSALITY	154
INVESTIGATION OF THE EFFECTS OF RESEARCH SUPPORTS ON ACADEMIC PERFORMANCE OF TURKISH UNIVERSITIES BY USING CORRELATION, REGRESSION AND PATH ANALYSES.....	155
LIFELONG LEARNER: QUESTION-ANSWERING CHATBOT.....	156
GENERALIZATION AND COMPOUNDING STATISTICAL DISTRIBUTIONS, GOALS AND BENEFITS: A REVIEW ...	157
ALPHA POWER TRANSFORMED TWO-PARAMETERS RAYLEIGH DISTRIBUTION: PROPERTIES AND INFERENCE	158
STATISTICAL ERRORS IN SPECIALIZATION IN MEDICINE THESIS	159
LOG - TWO - PARAMETER XGAMMA DISTRIBUTION: PROPERTIES AND INFERENCE	160
STATISTICAL EVALUATION OF THE CULTURAL DIFFERENCE PROBLEM EXPERIENCED BY REFUGEES IN HEALTH SERVICES	161
AN APPROACH OF FEATURE SELECTION CONSENSUS FOR THE GAS SENSORS' DRIFT DATA.....	162
THE RANKING OF CITIES FOR HOUSEHOLD CONSUMPTION EXPENDITURES IN TURKEY WITH RANKING-KEMIRA-M APPROACH	163
EVALUATION OF FEATURE REDUCTION AND CLASSIFICATION METHODS FOR ACTIVITY RECOGNITION	164



SOME PROPERTIES PROVIDED BY FOURIER TRANSFORM WITH DISTRIBUTION FUNCTION IN STATISTICS....	165
EXAMINING THE RESPONSE OF OLFACTORY STIMULUS IN EEG SIGNALS	166
RESEARCH OF STOCHASTIC PROPERTIES OF A POPULATION MODEL WITH FOKKER PLANCK EQUATION UNDER ALLEE EFFECT.....	167
A SATELLITE IMAGES BASED SAMPLING APPROACH IN AN AFRICAN CITY	168
DATA SCIENCE AND BIG DATA ANALYTICS EDUCATION - SERBIAN PERSPECTIVE.....	169
PREEMPTIVE GOAL PROGRAMMING FOR THE DAILY SHIPMENT PLANNING OF AN AUTOMOTIVE LUBRICANT COMPANY	170
THE USE OF AHP-TOPSIS METHOD IN TECHNOPARK PREFERENCES OF TECHNOLOGY COMPANIES: ISTANBUL CASE.....	171
TWO-STAGE SOLUTION FRAMEWORK FOR A WORKLOAD BALANCING AND WORKFORCE SCHEDULING PROBLEM: A COMPARATIVE STUDY	172
MACHINE LEARNING BASED FEATURE EXTRACTION FOR DETERMINATION OF BALANCE DISORDERS	173
SUPERVISED LEARNING BASED APPAREL RECOMMENDATION SYSTEM FOR ONLINE SHOPPING SYSTEM	174
PREDICTING LOSS AVERSION BEHAVIOUR WITH MACHINE LEARNING METHODS.....	175
PREDICTING BUBBLES IN THE S&P 500 INDEX WITH MACROECONOMIC INDICATORS: A MACHINE LEARNING APPROACH	176
CREDIT SCORING WITH USING ALTERNATIVE MODEL TECHNIQUES	177
ALGEBRAIC METHODS FOR THE PROBLEM OF FITTING CONCENTRIC ELLIPSES TO DIGITIZED OBSERVATIONS	178
TWO WAY MIXED DESIGN: SYMMETRY ASSUMPTION AND MISSING DATA	179
APPROXIMATION OF CONTINUOUS RANDOM VARIABLES FOR EVALUATING RELIABILITY OF COMPLEX STRESS-STRENGTH MODELS	180
ADAPTATION OF THE R-D CLASS PREDICTION METHOD TO LINEAR MIXED MODELS	181
MULTI CRITERIA GROUP DECISION MAKING WITH NEUTROSOPHIC SETS.....	182
THE RELATIONSHIP BETWEEN THE STOCK OF CAR LOANS AND THE GROWTH OF THE AUTOMOTIVE SECTOR FROM A TIMES SERIES ANALYSIS.....	183
NEUTROSOPHIC ARAS METHOD FOR SELECTION PROBLEMS.....	184
A NEW HEURISTIC ALGORITHM FOR THE LABEL PRINTING PROBLEM.....	185
CURVE FITTING FOR ENVIRONMENTAL MEASUREMENTS BY ARTIFICIAL INTELLIGENCE TECHNIQUES.....	186
ESTIMATING OPTIMAL VALUE FOR THE SHRINKAGE AND BIASING PARAMETERS IN LIU-TYPE LOGISTIC REGRESSION BASED ON PARTICLE SWARM OPTIMIZATION	187
A TRAFFIC ACCIDENT PREDICTION MODEL BASED ON ANN	188



IMPROVEMENT OF OFFICIAL STATISTICS BY APPLYING THE CONCEPT OF BIG DATA - CASE STUDY: SERBIA .	189
CIRCULAR ANALYSES OF DATES ON PATIENTS WITH GASTRIC CARCINOMA	190
IMPORTANCE OF FRACTAL STRUCTURES IN MODERN BIOLOGY AND DISCUSSION OF FORWARD KOLMOGOROV EQUATION WITH FRACTIONAL DERIVATIVE APPROACH	191
EXAMINATION OF INCIDENCES OF HIV INFECTION AND TUBERCULOSIS IN SOME SELECTED EUROPEAN COUNTRIES WITH STATISTICAL PROCESS CONTROL.....	192
A TWO-POPULATION MODEL WITH RENEWAL PROCESS FOR MEASURING LONGEVITY BASIS RISK.....	193
THE EFFECTS OF FINANCIAL INNOVATION ON ECONOMIC GROWTH IN THE BANKING SECTOR: CASE OF REGIONS AND PROVINCES	194
NET EXTERNAL POSITION, FINANCIAL DEVELOPMENT, AND BANKING CRISIS	195
THE GRAVITY MODEL OF INTERNATIONAL IMMIGRATION: EVIDENCE FROM OECD COUNTRIES	196
A PROMOTION PLANNING MODEL WITH PRODUCT SUBSTITUTION AND PROMOTION FATIGUE	197
HUMAN IDENTIFICATION WITH DEEP LSTM BY USING BIOMETRIC WALKING PATTERNS	198
DIFFERENT APPROACHES TO REDUCING BIAS IN CLASSIFICATION OF MEDICAL DATA BY ENSEMBLE LEARNING METHODS.....	199
REDUCE UNEXPECTED AIRLINE DIVERTS: MODELLING WITH NEURAL NETWORK	200
FINANCIAL DISTRESS PREDICTION USING SUPPORT VECTOR MACHINE WITH OPTIMAL PARAMETER AND FEATURE SELECTION.....	201
DISTRIBUTION THE DJSTOXXE50 INDEX.....	202
SPATIAL STATISTICAL ANALYSIS FOR GEODETIC POINTS OF A SIMULATED FAULT PLANE THROUGH GEOGRAPHICAL INFORMATION SYSTEMS	203
ASSESSMENT OF LONGEVITY RISK ON PENSION FUNDS: CREDIBILITY APPROACH	204
THE EFFECTS OF A PROMOTION ON STORE TRAFFIC, CONVERSION, AND CUSTOMER EXPENDITURES: EVALUATION WITH A PREDICTION APPROACH.....	205
NONPARAMETRIC STATISTICAL METHODS FOR THE COMPARISON OF TWO TREATMENTS IN THE PRESENCE OF COMPETING RISKS	206
MODIFICATION OF S TEST STATISTICS FOR RANDOMIZED BLOCK DESIGNS WITH AUTOCORRELATED LONGITUDINAL DATA.....	207
NOVEL APPROACHES FOR CHOOSING OPTIMUM CUT-OFF VALUE	208
PROGNOSTIC MODEL ANIMATED FILM FESTIVAL - VAFI.....	209
CLASSIFICATION OF OECD COUNTRIES WITH SOCIO-ECONOMIC AND EDUCATION INDICATORS	210
THEORETICAL AND PRACTICAL COMPARISON OF THE FREQUENCY BASED CLASSIFICATION METHODS: EDUCATION SAMPLE	211



ATAFORECASTING: AN R PACKAGE FOR AUTOMATIC SEASONAL OR NON-SEASONAL TIME SERIES FORECASTING.....	212
RANKING OF TURKISH UNIVERSITIES BASED ON ACADEMIC PERFORMANCE BY USING FUZZY C-MEANS CLUSTERING ALGORITHM.....	213
AN EXPERIMENTAL DESIGN APPLICATION IN CHEMISTRY FOR OPTIMIZATION OF ASSAY CONDITIONS FOR TRNA	214
SMOOTHING TECHNIQUES IN SOLVING ABSOLUTE VALUE EQUATIONS	215
APPROXIMATED CONVEX BEZIER CURVE FOR THE CONTINUATION GLOBAL OPTIMIZATION APPROACH.....	216
A NEW SMOOTHING TECHNIQUE FOR TOTAL VARIATION MINIMIZATION IN IMAGE DENOISING	217
AN EXACT APPROACH FOR A DYNAMIC WORKFORCE SCHEDULING PROBLEM.....	218
USE OF SIX SIGMA METHOD AS A PROCESS IMPROVEMENT TECHNIQUES	219
DATA MINING IN FINANCE: COMPARING DECISION TREES AND ASSOCIATION RULES FOR STOCK PRICE PREDICTION IN BIST 100.....	220
ENERGY CONSUMPTION AND ECONOMIC GROWTH: PANEL SIMULTANEOUS EQUATION MODELS	221
CUMULANT GENERATING FUNCTION OF SAMPLE MINIMUM OF ORDER STATISTICS FROM GEOMETRIC DISTRIBUTION	222
VOLATILITY MEASUREMENT ENTROPY METHODS.....	223
ENTROPY APPROACH FOR WIND ENERGY	224
COMBINATION OF FORECASTING BENCHMARKS ON M4 COMPETITION DATA SETS	225
FINANCIAL DATA ANALYSIS BY EXPONENTIAL SMOOTHING AND ATA METHOD	226
EXCHANGE RATE VOLATILITY AND STOCK RETURNS: A CASE OF CHINA AND TURKEY*	227
EXAMINING THE FINANCIAL SUCCESS OF STOCKS BY USING DATA MINING TECHNIQUES	228
THE P-SPLINES MODEL IN PREDICTING PM10 CONCENTRATIONS.....	229
EVALUATING NEGATIVE BINOMIAL REGRESSION AND ZERO TRUNCATED NEGATIVE BINOMIAL REGRESSION FOR DIFFERENT DISPERSION PARAMETERS	230
BINARY PARTICLE SWARM OPTIMIZATION AS A DETECTION TOOL FOR INFLUENTIAL SUBSETS IN LINEAR REGRESSION.....	231
J-INFLATED LONGITUDINAL REGRESSION MODELS FOR ANALYZING RESPONSES WITH FINITE SUPPORT	232
OPPOSITION BASED PARTICLE SWARM OPTIMIZATION FOR FEATURE SELECTION	233
ARTIFICIAL BEE COLONY ALGORITHM FOR ESTIMATING PARAMETERS OF A FLEXIBLE WEIBULL DISTRIBUTION	234
A STUDY ON LOSS PROBABILITY OF A BLOCKED QUEUEING SYSTEM.....	235
PARAMETER ESTIMATION OF THE KIES DISTRIBUTION BY ARTIFICIAL BEE COLONY ALGORITHM.....	236



DATA MINING OF SILC DATA: TURKEY CASE	237
COMPARISON OF PREDICTIVE PERFORMANCES OF CART AND LOGISTIC REGRESSION MODELS: APPLICATION OF LIFE SATISFACTION.....	238
SOME PROPERTIES OF COVARIANCE MATRICES OF PREDICTORS UNDER SEEMINGLY UNRELATED REGRESSION MODELS	239
BAYESIAN INFERENCE FOR TOBIT MODEL AND AN APPLICATION USING WITH WINBUGS PROGRAM	240
WAGE DYNAMICS ACCORDING TO INCOME GROUP IN TURKEY: 2014-2017 PERIOD	241
MODELLING THE MEDIATOR EFFECT OF DEMOGRAPHIC FEATURES ON CLIENTS' DEPOSIT VIA STRUCTURAL EQUATION MODELLING	242
THE IMPACTS OF OIL PRICE FLUCTUATIONS ON THE TURKISH STOCK MARKET: CORRELATIONS, VOLATILITY SPOILOVER AND HEDGING	243
LINEAR AND NONLINEAR UNIT ROOT TESTS APPLICATION: TURKISH ELECTRICITY CONSUMPTION PER PERSON	244
NOISE REDUCTION IN CHAOTIC TIME SERIES FOR IMPROVED FEATURE SELECTION, ABNORMAL PROCESS DETECTION, AND PROCESS MODELING	245
DETECTION OF MODE-MIXING PROBLEM IN EMD OF TIME SERIES SIGNAL DATA	246
INVESTIGATION OF THE RELATIONSHIP BETWEEN TURKEY AND ECONOMIC GROWTH INDUSTRIAL PRODUCTION INDEX.....	247
ESTIMATING THE OUTPUT GAP FOR LEBANON: A PRODUCTION FUNCTION METHODOLOGY	248
SUNK COSTS AND EXPORTING BEHAVIOR: A SECTORAL ANALYSIS	249
COMPLEXLY INTERWOVEN STOCHASTIC AND CHAOTIC TENDENCIES IN THE PROFIT-INTEREST RATIOS AT A RECENT CRISIS JUNCTURE IN TURKEY: POSSIBILITIES FOR CRISIS-MANAGEMENT	250
PRACTICAL TESTS OF FINANCIAL BUBBLES: BETTER CRITICAL VALUES FOR DETECTING EXUBERANT BEHAVIOR	251
INFERENCE OF ECONOMIC PERFORMANCE USING GLOBAL TRADE DATA.....	252
CHANGE POINT DETECTION IN ARMA(P,Q) PROCESS AND APPLICATION ON HEADWAY DATA	253
A NEW ROBUST UNIT ROOT TEST WITH A SIMULATION STUDY.....	254
A META ANALYSIS APPROACH ON WORK – FAMILY BALANCE AND JOB SATISFACTION	255
POSTERS.....	256
THE EFFECTS OF TIME INTERVAL ON TEST-RETEST RELIABILITY	257
PENALIZED POWER PROPERTIES OF COMPUTATION METHODS FOR THE GENERALIZED F-TEST	258
AUTHOR INDICES.....	259

Optimal Bandwidth Selection with Density Functional Estimation for Contaminated Data

Celal Aydın¹, Necla Gündüz²

¹Department of Statistics, Gazi University, Ankara, Turkey

²Department of Statistics, Gazi University, Ankara, Turkey

Abstract

The bandwidth selection is the most important part of the kernel density estimation. Various bandwidth selection methods have been proposed in past decades. In the nonparametric cases density functionals $\int \eta(u) f^2(u) du$ are very common in use for estimating location and scale parameter. The purpose of this study is mainly directed to evaluate the performance of optimal bandwidth selection based on the mean square error (MSE) of density functionals $\int \eta(u) f^2(u) du$ for contaminated data with outliers.

A simulation study is constructed for the random samples from Normal and Cauchy distributions with different parameters and with different contamination levels. In addition, the performance of bandwidth selections is compared in terms of MSE of density functionals $\int \eta(u) f^2(u) du$.

Key Words: Bandwidth, density functionals, Kernel smoothing, contamination data.

¹ aydinc@gazi.edu.tr

² Corresponding author: ngunduz@gazi.edu.tr, gunduznecla@yahoo.com

A New Robust Ridge Parameter Based on Search Method for Linear Regression Model

Atila Göktaş¹, Özge Akkuş², Aykut Kuvat³

Department of Statistics, Muğla Sıtkı Koçman University, Muğla, Turkey

Abstract

A large and wide variety of ridge parameters proposed for linear regression models exist in the literature. However, there is no ridge parameter that can best serve for any sample size or any degree of collinearity among regressors. To compare the efficiency of the ridge parameters proposed in the literature, Göktaş and Sevinç performed a simulation study in 2016 for a broad range of different sample sizes and different degrees of collinearity. Using the obtained four best ridge parameters from the study of Göktaş and Sevinç (2016), a new ridge parameter has been developed in the current study in a way of search method to provide the smallest mean square error values of regression parameters as a collection of those four. The proposed robust ridge parameter guarantees that it is free of sample size and degree of collinearity. After that another simulation study is conducted to show that the proposed parameter is robust. In conclusion, it is found that this ridge parameter is promising in any case. Besides, a popular data set is used as an example to illustrate the performance of the parameters in practice.

Key Words: Ridge Regression, multicollinearity, Ridge parameters, robust Ridge parameter.

What Kind of Patterns Underlie the Ratings Students Give their Professors? A Modern Machine Learning Analysis!

Necla Gündüz¹, Ernest Fokoue²

¹Department of Statistics, Gazi University, Ankara, Turkey

²Rochester Institute of Technology, School of Mathematical Sciences, Rochester, USA

Abstract

The evaluation of instructors by their students has been practiced at most universities for many decades, and there has always been a great interest in a variety of aspects of the evaluations. Are students matured and knowledgeable enough to provide useful and dependable feedback for the improvement of their instructors' teaching skills/abilities? Does the level of difficulty of the course have a strong relationship with the rating the student give an instructor? In this paper, we attempt to answer questions such as these using some state of the art statistical data mining techniques such support vector machines, classification and regression trees, boosting, random forest, factor analysis, *k*-Means clustering, hierarchical clustering. We explore various aspects of the data from both the supervised and unsupervised learning perspective. The data set analysed in this paper was collected from a university in Turkey. The application of our techniques to this data reveals some very interesting patterns in the evaluations, like the strong association between the student's seriousness and dedication (measured by attendance) and the kind of scores they tend to assign to their instructors.

Key Words: Questionnaire, likert, student evaluation, professor, instructor, random forest, clustering, *k*-means, pattern recognition, discovery, zero-variation, classification, factor analysis.

¹ Corresponding author: ngunduz@gazi.edu.tr.

² epfeqa@rit.edu

The Returns to College Education and Subject of Degrees in Turkey: A Cross Sectional Analysis on 2008-2017 Period

Hakan Ulucan¹, Cemil Çiftçi²

¹ Department of Economics, Pamukkale University, Denizli, Turkey

² Department of Economics, Pamukkale University, Denizli, Turkey

Abstract

A rapid expansion in both the numbers of new universities in the regions and the quotas of the existing universities have characterized the higher education of Turkey in the last decade. As a result, coverage of the college education in Turkey has increased enormously. The main aim of this study is to find out the effects of the university education expansion in Turkey on the college premium in the wages in two dimensions. First, the general impact of the higher education expansion on the wages of college graduates is presented. Second, the college premiums of the college majors is estimated. These estimations are conducted by using Quantile and Ordinary Least Squares Regression methods. The data from Household Labor Force Surveys of Turkish Statistical Institute covering the period 2010-2017 is used in this study to derive the Mincerian wage equations for Turkey. The findings show that the average returns to education for college graduates have been decreasing since the beginning of 2010s. A closer analysis investigating the returns to different majors shows that the wage premiums to science-technology based subject of degrees have been on the rise while the earnings of social science based majors have been falling.

Key Words: College wage premium, cross sectional analysis, quantile regression.

¹ Assistant Professor, mail: hulucan@pau.edu.tr

² Associate Professor, mail: cciftci@pau.edu.tr

Multivariate Intuitionistic Fuzzy Inference System for Time Series Prediction

Ozge Cagcag Yolcu ¹, Erol Egrioglu ², Eren Bas ², Ufuk Yolcu ³

¹Department of Industrial Engineering, Giresun University, Giresun, Turkey

²Department of Statistics, Giresun University, Giresun, Turkey

³Department of Econometrics, Giresun University, Giresun, Turkey

Abstract

Time series prediction is an important issue for decision-maker in a variety of fields such as finance, environment, informatics, agriculture etc... As it known that almost all the time series prediction tools have the univariate structure. However, in the time series prediction problem, more than one time series can interact which each other. In such a case, the analysis can be performed by taking one of them as the main time series which is desired to predict, and the others as auxiliary time series. Different fuzzy inference systems have been used successful in time series prediction literature. Especially, in some recent studies, intuitionistic fuzzy inference systems have improved prediction accuracy by using additional information obtained from intuitionistic fuzzy sets. In this study, a multivariate intuitionistic fuzzy time series definition is given and to predict such a time series, a multivariate intuitionistic fuzzy inference system is introduced. By using this kind of inference system (M-IFIS), it is taken advantage of additional info from both non-memberships and auxiliary time series. The proposed M-IFIS uses pi-sigma artificial neural network to specified the relationship between inputs and outputs. Inputs of the system consist of memberships, non-memberships and crisp observations of time series, and the crisp observations of main time series constitute the outputs of the system. The training of the pi-sigma neural network is realized by particle swarm optimization. Moreover, the memberships and non-memberships are achieved via intuitionistic fuzzy C-means clustering method. To bring into light the prediction performance of the proposed system, some real-world time series are analysed and the obtained results are compared with the results of some state-of-the-art approaches.

Key Words: Intuitionistic fuzzy time series, multivariate time series, pi-sigma artificial neural network, particle swarm optimization, prediction.

¹ ozgecagcag@yahoo.com (Corresponding Author)

² erol.egrioglu@ giresun.edu.tr

² eren.bas@giresun.edu.tr

³ varyansx@hotmail.com

A Novel Time Series Forecasting Model based on Single-Valued Neutrosophic Sets

Ufuk Yolcu ¹, Eren Bas ², Erol Egrioglu², Ozge Cagcag Yolcu³

¹ Department of Econometrics, Giresun University, Giresun, Turkey

² Department of Statistics, Giresun University, Giresun, Turkey

³ Department of Industrial Engineering, Giresun University, Giresun, Turkey

Abstract

Neutrosophic sets (NSs) generate 3 kinds of memberships. Single-valued neutrosophic sets (SV-NSs) are a special case of NSs and it also based three membership functions as truth, indeterminacy and falsity which take the values between zero and one. Although, some fuzzy-based and intuitionistic fuzzy-based approaches have been proposed with the aim of time series forecasting, there is no forecasting tool based on NSs. A NS-based time series forecasting model, thanks to allowing to use whole information contained by all three membership functions in the analysis process, may be expected to be a more effective forecasting tool. From this point of view, in this study, a novel time series forecasting model based on single-valued neutrosophic sets (TSFMbSVNS) is presented. TSFMbSVNS uses three membership values like truth, indeterminacy and falsity as separately inputs for three separate sigma-pi neural networks (SPNNs) to determine the neutrosophic relationships between inputs and outputs. The outputs of each SPNN, which are composed of crisp observations, are combined by using neutrosophic weights which are specified by particle swarm optimization (PSO) for each membership function and so final forecasts are obtained. PSO is also used to the training of SPANNs in a simultaneous process with specifying neutrosophic weights. The membership values are achieved by single valued neutrosophic clustering. It is aimed to be demonstrated the forecasting performance of TSFMbSVNS via some real-world time series implementations. The obtained results are evaluated with the results of some other fuzzy-based approaches, together.

Key Words: Forecasting, single valued neutrosophic sets, time series, sigma pi neural networks, single valued neutrosophic clustering.

¹ varyansx@hotmail.com (Corresponding Author)

² eren.bas@giresun.edu.tr

² erol.egrioglu@giresun.edu.tr

³ ozgecagcag@yahoo.com

Deep Intuitionistic Fuzzy Time Series Forecasting Method Based on Long Short Term Memory

Cem Koçak¹, Erol Egrioglu², Eren Bas³

¹Faculty of Health Sciences, Hitit University, Çorum, Turkey

²Department of Statistics, Giresun University, Giresun, Turkey

³Department of Statistics, Giresun University, Giresun, Turkey

Abstract

Recent years, deep artificial neural networks produce good forecasting results than many of other artificial neural networks. Long short time memory has not been used to determine fuzzy relations in an intuitionistic fuzzy time series forecasting methods. In this paper, a new intuitionistic fuzzy time series forecasting method is proposed. The proposed method is called as deep intuitionistic fuzzy time series forecasting method because determining of fuzzy relations is made by using long short term memory artificial neural network. The proposed method employed intuitionistic fuzzy c-means for obtaining membership and non-membership values. The inputs of the long short term memory are merged membership and non-membership values by minimum operator. Moreover, lagged crisp values are inputs of the long short term memory. The architecture of the long short term memory contains multiple input and single output. The proposed method is applied to well-known time series in the literature. It is shown that the proposed method outperforms some methods in the literature.

Key Words: Intuitionistic fuzzy time series, long short term memory, intuitionistic fuzzy c-means, deep learning.

References

- [1] Tran N., Nguyen T., Nguyen B.M., Nguyen G. (2018), A Multivariate Fuzzy Time Series Resource Forecast Model for Clouds using LSTM and Data Correlation Analysis., *Procedia Computer Science*, 126, 636–645.
- [2] Sadaei H.J., Silva P.C.L., Guimaraes F.G., Lee M.H. (2019), Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series, *Energy* 175, 365-377.

Panel Data Analysis

Yüksel Akay Ünvan¹

¹Assoc. Prof., Ankara Yıldırım Beyazıt University, Faculty of Management, Banking and Finance Department, Turkey

Abstract

In this study, among the developing countries; Argentina, Azerbaijan, Egypt, Indonesia, Russia, and Turkey were selected and Dollar (USD) and Euro exchange rates, GDP, GNI and export values of these countries were taken as a variable. The annual average of these variables between 2016-2018 was obtained and the data set was formed. The main purpose of this study is to see the effect of other variables on USD. For this purpose, Panel Data Analysis was used. Firstly, descriptive statistics and correlation coefficients of variables were given. Then, the Hadri Unit Root Test was performed to test the stationarity of the variables. After determining that all variables were stationary and did not need to be differentiated, a regression model was established with panel data. However, it was not known whether it is fixed or random. Therefore, the Hausman Test was performed on the established model. As a result of this test, it was decided that a fixed effect model would be established and so a fixed effect panel regression model was formed. According to the results obtained from this model; between 2016-2018, it was determined that only EURO and constant value were significant when USD was the dependent variable. On the basis of countries, Indonesia has a positive coefficient, which is different from other countries, and this was due to the fact that the USD and EURO are extremely high in this country compared to other countries and the difference between them was therefore high.

Key Words: Panel data analysis, hadri unit root test, hausman test, exchange rates.

¹ aunvan@ybu.edu.tr

The Impact of ICT Penetration on Financial Development: An International Cross-Country Analysis

Mei-Se Chien ¹, Chien-Chiang Lee², Chih-Yang Cheng³, Meta Ayu Kurniawati⁴

¹ Department of Finance and Information, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

² Department of Finance, National Sun Yat-sen University, Kaohsiung, Taiwan

³ Department of Finance, National Sun Yat-sen University, Kaohsiung, Taiwan

⁴ Department of International Relations, Ritsumeikan university, Kyoto, Japan

Abstract

The purpose of this study is to explore the effects of information and communication technology (ICT) diffusion on financial development. We apply the dynamic generalized moment method (GMM) to estimate these effects for the data sample of 81 countries with different income levels and geographic classifications from 1990 to 2015. Furthermore, the panel smooth transition regression (PSTR) model is applied to investigate the nonlinear relationship between ICT diffusion and financial development. The empirical results are as follows: First, the ICT diffusion generally is conducive to financial development for most of cases, but mobile cellular has a negative impact on financial development in the high-income countries, which is the same as the countries with higher mobile penetration rates. Secondly, strong evidence shows that the PSTR models capture the smooth non-linear effects of ICT diffusion and financial development. The effects of all three ICT variables on financial development are significantly positive in the lower regime of ICT diffusion, but these effects change in the higher regime. From the lower to higher regime of ICT diffusion, the effect becomes smaller for the case of fixed telephone, but the effect is from positive to negative for the cases of mobile cellular or the internet. Finally, regardless of the geographical area, upgrade the internet and fixed telephone will promote financial development, while mobile cellular will only benefit financial development in Africa.

Key Words: ICT diffusion, financial development, Dynamic panel GMM, PSTR.

Determinants of Trade Flows in Turkey

Bilgin Bari¹

Department of Economics, Asst. Prof.Dr., Eskisehir, Turkey

Abstract

This study analyzes the trade flows for the long and short term in the Turkish economy. For this purpose, the main determinants of exports, imports, and foreign trade deficit are examined by different time series approaches. The dependence on imported inputs and intermediate goods for production in the Turkish economy is an essential structural problem, and the effects of this dependency on imports and trade deficit are examined. The effects of the exchange rate on exports are investigated using the nominal exchange rate, real exchange rate, and real effective exchange rate. Uncertainties about pricing caused by excessive volatility in exchange rates are also included in the analysis as an essential factor. While the ARDL (Autoregressive Distributed Lag) model is estimated for the analysis of long-term relationships, the Error Correction model is used for the analysis of short-term effects. Besides, the VEC (Vector Error Correction) model is also used to reveal the interaction between economic variables. The primary purpose of the research is to reveal the structural problems related to production and foreign trade by detailing the help of descriptive and econometric methods. In the study, the effects of foreign income on the export of Turkey are examined by using an index of a weighted average of the gross domestic product of Turkey's main trade partners. The findings emphasize the dependence of production and exports on imported inputs and intermediate goods. In addition, changes in exchange rates and an increase in volatility have significant effects on trade flows.

Key Words: Trade deficit, export, import, autoregressive distributed lag model, error correction model, vector error correction model.

¹ e-mail: bbari@anadolu.edu.tr

A Longitudinal Data Analysis to Examine the Environmental Kuznets Curve and the European Union Enlargement: The Case of Carbon Monoxide

İsmail Onur Baycan¹

¹Department of Economics, Faculty of Economics and Administrative Sciences, Anadolu University, Eskisehir, Turkey.

Abstract

This study investigates the Environmental Kuznets Curve hypothesis between the levels of Carbon Monoxide concentration in the air and per capita income growth, considering the European Union's largest enlargement process. Carbon monoxide, which is one of the most common and widely distributed key air pollutants, is utilized for the group of three different countries: the core group of initial fifteen EU countries before the largest enlargement, the group of overall twenty-five EU countries after the enlargement, and the group of countries that became the members of the EU right after this enlargement process. The study verifies that there is no evidence of a significant original or reverse Environmental Kuznets curve relationship between carbon monoxide concentration and per capita income growth for the group of core fifteen EU member countries before the enlargement and the group of twenty-five countries combined after the enlargement. As a result, empirical findings suggest that the largest enlargement has not changed the fact that an Environmental Kuznets curve in any type does not exist for the EU. The findings reveal that the largest enlargement policy of the EU does not have a significant effect on the relationship between the levels of CO concentration and per capita income growth in the union.

Key Words: Environmental Kuznets curve, economic growth, European Union enlargement, air pollution.

Identification of Wealthy Households from the Residential Property Price Index Database for Sample Selection for Household Surveys

Evren Ceritoğlu¹, Ozlem Ardic Sevinc^{1,2}

¹Structural Economic Research Department, Central Bank of the Republic of Turkey, Ankara, 06050 TURKEY

²Department of Industrial Engineering, TOBB University of Economics and Technology, Ankara, 06560 TURKEY

Abstract

This paper aims to identify wealthy households in Turkey for sample selection for household surveys. In the absence of income and wealth tax data, we analyze house prices from the Residential Property Price Index (RPPI), which is constructed by the Central Bank of the Republic of Turkey (CBRT) from dwelling appraisal reports to monitor price movements. RPPI is announced monthly by the CBRT for Turkey and 26 geographical regions at NUTS2 level since January 2010. The RPPI database comprises more appraisal observations from İstanbul and western provinces, where house prices are significantly higher than national average. However, the number of appraisal observations is low for Eastern provinces, since house sales are limited in poor and small provinces.

We examine unit house prices from the CBRT – RPPI database from 2016 to 2018 at both province and neighborhood level. Unit house prices are calculated by dividing the value (TL) to the gross usage area (m2) at current prices. Only neighborhoods with 30 or more observations are examined in the analysis. We regress the natural logarithm of the number of home appraisals on the natural logarithm of unit house prices using mean values. We perform fixed effects regressions using our unbalanced panel data set. We control for year effects by introducing time dummy variables into the regressions. We find that there is a positive and statistically significant relationship between unit house prices and the number of home appraisals. Moreover, we perform the same regressions for neighborhoods that have more than 100 observations as a robustness check. We observe that the size and the sign of the regression coefficients do not change when we restrict our data set.

The direction of the relationship might be from the number of home appraisals to unit house prices or both it could be both ways. Therefore, as another robustness check, we regress the natural logarithm of unit house prices on the natural logarithm of the number of home appraisals. We observe that there is a statistically significant relationship between the number of home appraisals and unit house prices. However, the size of the regression coefficients is considerably lower in this case. As a result, our empirical analysis indicates that the number of observations is higher in administrative units, where house prices are higher. Hence, we argue that identification of wealthy households according to their neighborhoods using the RPPI database is a reliable and consistent method.

Key Words: Unit house prices, wealthy households, panel data.

A Specific Application of Comparison of OECD Countries with Credit Risk Swap Premiums with Quantile Regression and Multiple Regression Analysis Techniques

İrem Yanık¹, Hüseyin Tatlıdıl²

¹Hacettepe University Department of Statistics, Ankara, Turkey

²Hacettepe University Department of Statistics, Ankara, Turkey

Abstract

Regression analysis is one of the most important statistical analysis methods in statistics and econometrics. Regression is the association of a dependent variable with one or more independent variables, the correlation between them and the analysis of this correlation. Quantile regression was developed to provide a more comprehensive regression image as an alternative to the least squares method used in simple linear and multiple linear regression analysis. Quantile regression should be preferred when the model assumptions on which the least squares estimator is based are not met and where extreme values are important.

OECD (Organization for Economic Cooperation and Development) countries are composed of industrialized and developing countries. The OECD aims to assist governments in achieving prosperity through cooperation in the fields of economic growth, financial stability, trade and investment, technology, innovation, entrepreneurship and development and combating poverty and also Turkey is member of OECD.

Credit risk swap premiums mean insurance of a debt as a result of agreement with third parties against the possibility of default. In other words, credit risk draft premiums are the premiums paid by a creditor to a third party or institution to insure the debt papers in its portfolio. Credit Default Swap is abbreviated as CDS using initials. The credit risk swap premium is translated into Turkish at the rate of credit default swap.

In this study, simple linear regression model, multiple linear regression model, quantile regression, theoretical historical development and application areas of quantile regression are included. It is aimed to investigate the effect of OECD countries' Current Deficit, Consumer Confidence Index and Gross Domestic Product indicators on Credit Risk Clearing Premiums. For this purpose, the effect of these indicators on Credit Risk Clearing Premiums was first examined by Quantile Regression followed by Multiple Regression Analysis. In the last section, the results are interpreted and it is tried to be shown that Quantile Regression is a more robust regression technique.

Key Words: Quantile regression, OECD, credit default Swap.

Bibliometric Analysis of Five Journals with the Highest Impact Factor in the Field of Statistics Using Citespace

Fatma Yardibi¹, Mehmet Ziya Fırat²

¹ Akdeniz University, Department of Animal Science, 07050 Antalya, Turkey

² Akdeniz University, Department of Animal Science, 07050 Antalya, Turkey

Abstract

This study aims to identify trends in statistics literature in bibliometric terms by analyzing scientific studies carried out in the last 20 years. The research data consisted of 8700 studies published between 2000-2019 in the top-five journals of Statistics field, according to the Journal Citation Reports. The journals were analyzed in terms of number and types of publications, author, institution, country productivity, citation analysis, and citation burst. In the study, emerging trends and statistics intellectual structure are visualized with social network analysis. Research results show that USA is the leading country having the highest citations (6007) and highest centrality values (0.39) and England is the following leading country with the 810 citations and 0.19 centrality values. As a result of the keyword analysis, which is one of the most effective ways to show emerging trends, 'variable selection' was found to be the most active cluster. Result of cluster analysis, 'statistical significance' and 'longitudinal data' defined as active clusters, and this may become a popular topic in the future. Moreover, as a result of the word analysis conducted on the works made in the field, it was found that most repeated words are model, regression, and inference. Citation burst, one of the most effective methods to determine research trends in a discipline, occurred in cluster #2 (Bayesian model selection), according to the author. It is thought that this study will set an example for future studies in terms of the performance of the analysis.

Key Words: Bibliometric analysis, CiteSpaces, statistics, social network analysis.

A Comparison of Variety of Ridge Parameters by Means of Their Distributions and MSE Values

Selman Mermi¹, Atila Göktaş², Özge Akkuş³

¹Statistics Department, Muğla Sıtkı Koçman University, Muğla, Turkey

²Statistics Department, Muğla Sıtkı Koçman University, Muğla, Turkey

³Statistics Department, Muğla Sıtkı Koçman University, Muğla, Turkey

Abstract

In multiple regression, multicollinearity is a problem that always occurs when two or more explanatory variables are highly correlated with each other. When multicollinearity exists in data, the method of Ordinary Least Square (OLS) generally produces poor parameter estimates. Besides, OLS estimates of the regression coefficients might be inflated and OLS estimates will be too large and unstable. Ridge regression is one of the most popular methods used in the presence of multicollinearity and was introduced by Hoerl and Kennard (1970) as an alternative to the OLS estimator. In this method, ridge parameter k plays a vital role in parameter estimation. In this study, the comparison of all encountered ridge regression estimators in the literature have been performed in terms of their distributions and MSE values. For this purpose, 51 different estimators have been studied and comparisons have been carried out using Monte Carlo simulations. A large number of different models are constructed including different number of variables, sample size, variance of the random error and correlations among the explanatory variables. For each model 10000 replications have been performed. The distribution has been obtained for each ridge parameter and surprisingly found that most of them are skewed and only few of them behaves like the gauss distribution. Moreover, few of the ridge parameters interestingly are distributed within the range zero and ten.

Key Words: Multiple linear regression, Ridge regression, multicollinearity, Monte Carlo simulation.

Finite Mixtures of Skew Laplace Normal Distributions with Random Skewness

Fatma Zehra Doğru¹, Olcay Arslan²

¹Department of Statistics, Giresun University, Giresun, Turkey

²Department of Statistics, Ankara University, Ankara, Turkey

Abstract

The shape mixtures of the skew Laplace normal (SMSLN) distribution is introduced as a flexible extension of the skew Laplace normal distribution which is also a heavy-tailed distribution. The SMSLN distribution includes an extra shape parameter which controls skewness and kurtosis. Some distributional properties of this distribution are derived. In addition, we propose finite mixtures of SMSLN distributions to model both skewness and heavy-tailedness in heterogeneous data sets. The maximum likelihood (ML) estimators for parameters of interests are found via the expectation-maximization (EM) algorithm [Dempster, A.P., Laird, N.M. and Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 39, 1-38]. We also provide a simulation study and a real data example to illustrate the performance of proposed estimators.

Key Words: EM algorithm, finite mixture model, maximum likelihood, shape mixtures of the skew Laplace normal.

¹ fatma.dogru@giresun.edu.tr

² oarslan@ankara.edu.tr

Convex Combination of Some Estimators: A Comparative Study

Buatikan Mirezi¹, Selahattin Kaçiranlar², Hüseyin Güler³

¹Department of Statistics, Çukurova University, Adana, Turkey

²Department of Statistics, Çukurova University, Adana, Turkey

³Department of Econometrics, Çukurova University, Adana, Turkey

Abstract

Farebrother (1984) introduced an estimator combining the ordinary least squares (OLS) estimator and the restricted least squares (RLS) estimator. This new estimator is a general estimator defined by a convex combination of OLS and RLS and it includes OLS, RLS and ridge estimators. This estimator has a number of interesting properties since a convex combination of two estimators can be useful when both estimators appear to be appropriate in a specific situation. In addition, there is a possibility for the observed loss of the convex combination to be even smaller than both individual observed losses.

Farebrother (1984) theoretically compared the proposed estimator with OLS and RLS with respect to the mean square error (mse) criterion. However, to the best of the authors' knowledge, Farebrother's proposed estimator is not compared with the ridge estimator and there is no numerical or simulation study in the literature that compares aforementioned estimators. Therefore, in this study, we first theoretically compare Farebrother's proposed estimator with the ridge estimator according to the mse criterion. Then we present a Monte Carlo simulation experiment to compare Farebrother's estimator with OLS, RLS, and ridge estimators. Finally, we give a numerical example to demonstrate some of the theoretical results.

Key Words: Mean square error, multicollinearity, ordinary least squares, ridge regression, restricted least squares.

Type-1 Penalized Regression Functions

Nihat Tak¹, Deniz İnan²

¹Econometrics, Kırklareli University, Kırklareli, Turkey

²Statistics, Marmara University, Istanbul, Turkey

Abstract

Type-1 fuzzy functions (T1FFs) were introduced for classification and regression problems at first. However, they are also used for forecasting purposes. T1FFs starts with clustering the input matrix, constructed as the lagged values of the time series, by using fuzzy c-means clustering algorithm. The degree of membership values of the objects are included in the input matrices, next. Because different functions of the degree of membership values increase the forecasting ability of T1FFs, they are also added in the input matrices. However, adding very related functions of the degree of membership values in the input matrix increase the multicollinearity problem in T1FFs. In this sense, penalized regression methods are employed in T1FFs. Australian beer consumption and Istanbul stock exchange datasets are used to evaluate the performance of the proposed method. The results indicate very promising outcomes in terms of both root mean squared errors and mean absolute percentage errors.

Key Words: Forecasting, big data, p -value, fuzzy inference system, artificial intelligence.

Mismatch of Ridge Logistic Regression and Support Vector Machine: Financial Distress Estimation

Ömer Utku Erzengin¹

¹Department of Statistics, Süleyman Demirel University, Assistant Prof, Isparta, Turkey

Abstract

Enterprises can fall into a state of financial distress (FD). FD can be found by examining the cash and non-cash items in the income statements and balance sheets (IT&BS) of the enterprises. In this study, the IT&BS of 302 companies traded on Borsa Istanbul in 2018 were used. Ridge logistic regression (RLR) and support vector machine (SVM) was used to estimate FD. The discrepancy between the estimations of the two methods will be discussed.

The items which may be affected by another item of IT&BS are not examined alone. When the items on the IT&BS are compared to one another, the financial ratios (a kind of standardization) must be calculated. A financial ratio is a relative magnitude of two or more items taken from an enterprise's IT&BS. Financial ratios are Since the IT&BS the whole financial aspect of an entity, the financial ratios obtained are related to each other. The relationship between financial ratios also raises the problem of multicollinearity.

Edward I. Altman et al., published the formula for FD in 1983 based on financial ratios. Altman published an FD Z score in 1983 between 1.23 and 2.99 regions is defined as blurred. Altman Z score was less than 1.23 considered unsuccessful and over 2.99 considered successful.

FD risk estimates can be analyzed by logistic regression. Multicollinearity problem can be solved by RLR. Croux and Haesbroeck in 2003 offered the robust RLR method may be preferred instead of the MLE estimator in case of the connection problem between the independent variables of the data with multicollinearity problem derived from the financial ratios.

One of the classification methods used in data mining is machine learning. The SVM, a machine learning developed by Vapnik et al., is used to classify the variables of the two groups in the most appropriate way. The SVM has its own linear and nonlinear methods for solving problems such as multiple connection problems. If the data is in a complex state due to its multi-dimensional structure, nonlinear methods are used instead of linear methods.

FD failure was accepted as 0 and success was 1. Since successful unsuccessful areas are more precise, the model for the RLR will be found and the fuzzy region will be estimated according to the model. The same process will be done in the SVM. The conflict between RLR and the predicted performances of the financial failure of the SVM will be examined.

Key Words: Mismatch, Forecasting, Ridge Logistic Regression, Support Vector Machine, Financial Distress.

Does Gender Diversity, Institutional Ownership and Capital Structure Affect Firm Performance? An Indonesia Real Estate and Property Firms Analysis

Rani Raharjanti¹

¹Accounting Department, Politeknik Negeri Semarang, Semarang, Indonesia

Abstract

During emancipation era, women on board has been of major interest to corporate governance in recent years. The feminism already changed the mindset of millennium people. Feminism believes that there are equality at the top management between man and woman. This study analysis the effect of gender diversity, institutional ownership and capital structure on firm performance. Institutional ownership is a part of corporate governance's perspective. We added capital structure to see from finance management's perspective. By making use of cross sectional data for a sample of 48 real estate and property firms in Indonesia during period 2009-2018. To identify a causal effect of gender diversity, institutional ownership and capital structure on firm performance, we used multiple regression analysis. Our empirical evidence shows that gender diversity (woman boardroom) has no effect to firm performance measured by Return on Equity (ROE). It is indicating that the executive has no effect to oversight the monitoring effect. It is reinforced by Bloomberg Businessweek's research, woman on board in Indonesia in 2017 under 20 %. Based on Southeast Asia women perspective and culture, majority, Asian women did not have passion to become a leader than men. In addition, we found that findings of positive relationship between institutional ownership and firm performance. It means that institutional ownership did fully trusted on firm performance. The institutional ownership was oversight firm where they had been invested. There was a positive relationship between capital structure and firm performance. It can be concluded that the use of high debt is considered a positive signal for investors.

Key Words: Board diversity, institutional ownership, capital structure, corporate governance, real estate and property firms.

¹ Corresponding author. Email: rani.raharjanti@polines.ac.id

Synoptic Climatology of Black Sea-Effect Snowfall Events in Istanbul, Turkey

Hakki Baltaci^{1,*}, Bulent Oktay Akkoyunlu²

¹Turkish State Meteorological Service, Regional Weather Forecast and Early Warning Center, Istanbul, Turkey

²Marmara University, Department of Physics, Istanbul, Turkey

Abstract

A climatological analysis and overlying synoptic conditions of Black Sea-effect snowfall events were investigated for Istanbul, Turkey during the winter (DJF) period of 1971-2006. Concerning with the synoptic climatological approach, Lamb Weather Type (LWT) method was applied to NCEP/NCAR daily mean sea level pressure data. Basically, northwesterly (NW), northerly (N), and northeasterly (NE) circulation types (CTs), which flows blowing from Black Sea (BS), were thought to be relevant to occur sea-effect snowfall events. Wind speeds and flows at 850-hPa, directional shear, and temperature difference between sea surface and 850-hPa level thresholds were applied to these three CTs in order to find the suitable snowfall cases originated from Black-Sea. Based on our results, totally 4, 14, and 111 snowfall episodes occurred during NW, N, and NE types over Istanbul with the 2.8, 4.1, and 3.5 cm daily mean snow cover depths (DMSCD), respectively. In particular, it is found that interaction between surface high located over continental Europe and a low pressure located over central Black Sea, and relatively warm sea surface temperature (SST) and cold temperature anomaly at the low level of the atmosphere ($SST-T_{850} > 16$ °C) are favorable environment for the development of intense Black Sea-effect snowstorms (DMSCD > 10 cm) sourced by NE cases. When compared with NE types, although rare intense snowstorm events are shown, their contribution to DMSCD is greater during N types (18 cm SCD on daily average, a deep cyclone located over northern BS).

Key Words: Sea-effect snowfall, synoptic condition, Lamb Weather Type, composite analysis, Istanbul.

* Corresponding author. e-mail: baltacihakki@gmail.com, Tel: +902164573400 Fax: +902164573403

Regularized Elimination for Variable Selection in Categorical Partial Least Squares

Maryam Sadiq¹, Tahir Mehmood^{2*}, Muhammad Aslam³

^{1,3}Department of Mathematics and Statistics, Riphah International University, Islamabad, Pakistan.

² School of Natural Sciences (SNS), National University of Sciences and Technology (NUST), Islamabad, Pakistan.

Abstract

Like other scientific fields, public health also generating high-dimensional data, where obviously some variables are noise and some are influential for defining the variation in response variable. The high dimensional data usually faces with curse of dimensionality i.e. multicollinearity and dimensionality. Both of these problems can be handled with variable selection. For high dimensional data sets partial least squares (PLS) is an accepted algorithm to model the data. PLS in its standard form has no mechanism for variable selection, moreover PLS is designed for continues variables. In a recent study PLS algorithm is improved by introducing the categorical correlation measures for PLS loading weights (Cat-PLS). The categorical correlation measures includes Cramer's V, Phi coefficient, Tschuprow's T coefficient, Contingency Coefficient, Yule's Q and Yule's Y. For variable selection we have implemented the regularized elimination in Cat-PLS (re-Cat-PLS). The proposed method is implemented over the simulated data and is implemented to identify the significant variables (factors) associated with Cesarean section (CS). The maternal data was obtained from Pakistan Demographic and Health Surveys (PDHS). The proposed algorithm shows the significant improvement over the standard PLS algorithm on both simulated data and CS data.

Key Words: High dimensional, partial least square, variable selection. public health.

Estimation of AR(1) Model Having Generalized Logistic Disturbances

Özlem Türker Bayrak¹, Ayşen Dener Akkaya²

¹Department of Inter-Curricular Courses, Cankaya University, Ankara, Turkey

²Department of Statistics, Middle East Technical University, Ankara, Turkey

Abstract

Non-normality is becoming a common feature in real life applications. Using non-normal disturbances in autoregressive models induces non-linearity in the likelihood equations so that maximum likelihood estimators cannot be derived analytically. Thus, modified maximum likelihood estimation (MMLE) technique is introduced in literature to overcome this difficulty. However, this method assumes the shape parameter to be known which is not realistic in real life. Recently, for unknown shape parameter case, adaptive modified maximum likelihood estimation (AMMLE) method that combines MMLE with Huber estimation method is suggested in literature. In this study, we adopt AMMLE method to AR(1) model where the disturbances are Generalized Logistic distributed. Although Huber M-estimation is not applicable to skew distributions, the AMMLE method extends Huber type work to skew distributions. We derive the estimators and evaluate their performance in terms of efficiency and robustness via simulation. Besides, we compare them with MMLEs and Least Squares estimators.

Key Words: Adaptive modified maximum likelihood estimation, maximum likelihood estimation, least squares, efficiency, robustness.

¹ Corresponding author. E-mail: ozlemt@cankaya.edu.tr

² E-mail: akkay@metu.edu.tr

Analysis of the Efficiency of Venture Capital Investment Trusts Traded in Borsa İstanbul: A Proposal about Regulating Negative Data

M. Esra Atukalp¹

¹Department of International Trade, Giresun University, Giresun, Turkey

Abstract

Venture capital can be defined as the long term investment made by those who have surplus funds for the formation and operation of small and medium sized enterprises with high development potential. These venture enterprises should aim at the creation or development of tools, equipment, materials, services or new products, methods, systems and production techniques with industrial, agricultural application and commercial market potential, or be able to achieve these objectives through management, technical or capital support. Venture capital investment trust is one of the instruments used in the realization of these venture capital investments.

In this study, the efficiency of venture capital investment trusts traded in Borsa İstanbul was examined. Efficiency of the venture capital investment trusts was examined via input oriented data envelopment analysis method. 2014-2018 periods was selected as an analysis period. In the analysis, the ratio of venture capital investments to assets and equity to assets was determined as input and the return on assets variable was determined as output. The data required for the analysis were obtained from the Public Disclosure Platform. As a result of the analysis, it was determined that Gözde Venture Capital Investment Trust had the highest efficiency during the period. According to the results, 2018 was the year with the highest efficiency.

Since negative data cannot be used in data envelopment analysis, negative real data is made positive by researchers by adding a minimum value. In this study, because the variables are fractional numbers, the minimum value added to provide positivity is formulated by me as the ratio that makes the smallest negative variable positive.

Efficiency is an indicator of the extent to which it has reached its goals and objectives with limited resources and the relevance of efforts to expected and planned results. Therefore, in data envelopment analysis method especially in output oriented approaches based on output maximization, attention should be paid to the appropriateness of accepting these decision making units effectively with these negative outputs, even if the variables of decision making units with negative outputs are made positive for performing the analysis.

Key Words: Efficiency analysis, venture capital investment trusts, negative data, data envelopment analysis.

¹ Assistant Professor.

The Changes in Last 10 Years of the Marriage and Divorce Statistics Based on Regions of Level 2 According Statistical Regional Units Classification (NUTS) in Turkey and Age Groups: Examples 2008 and 2018

Fatma Sevinç Kurnaz¹

¹İstatistik Bölümü, Yıldız Teknik Üniversitesi, İstanbul, Türkiye

Abstract

Marriage and divorce are concepts that give important clues about societies, changes in societies and all social processes. In this study, the marriage and divorce statistics in Turkey for 2008 and 2018 are examined considering the changes in the age groups and regions of level 2 according to Statistical Regional Units Classification (NUTS). The findings are interpreted in detail.

Key Words: Marriage, divorce, statistics.

Classification with Two-Stage Correlation-Based Attribute Selection on Big Data Platform

Muhammet Sinan Bařarslan¹, Fatih Kayaalp²,

¹Department of Computer Programming, Doęus University, Istanbul, Turkey

²Department of Engineering Faculty, Düzce University, Düzce, Turkey

Abstract

The amount of created data has increased a lot in recent years. Because of this, data mining is used by various disciplines for extracting information through these datasets. Attribute selection which refers to identifying the attributes of dataset that have more contribution on the result is an important stage on data mining processes. The platform to be used for data mining processes also has effect on the performance of the task. Evaluating the performance of the new attribute selection method called Two Stage Correlation Based Attribute Selection (TSCBAS) which has been proposed by our previous works is aimed in this study. For this aim, SVM and Random Forest classification algorithms are applied on bank marketing data set from UCI machine learning data warehouse on two different data mining platforms such as Spark and R. The dataset was separated as training and testing data by 5-fold cross-validation method. According to the results, SVM has shown better classification performance than random Forest both on raw dataset and the dataset created with TSCBAS. In addition, Spark has performed better runtime result than R. The results have also confirmed the importance of attribute selection process.

Key Words: Attribute selection, TSCBAS, Apache Spark, R.

Liu and Restricted Liu Estimators in Linear Measurement Error Models

Gülesen Üstündağ Şiray¹

¹Department of Statistics, Çukurova University, Adana, Turkey

Abstract

Measurement error models are employed when the basic assumption of linear regression model that is all observations are correctly observed is violated. The situations that the observations of the variables are exposed to measurement errors are encountered frequently in data analysis. When measurement errors present in the data, usual statistical methods tend to lose their validity and ordinary least squares estimator of regression parameters become biased and inconsistent. So, some additional information is required, such as known covariance matrix of measurement errors and known matrix of reliability ratios in order to find the consistent estimators of the parameters. However, in the measurement error models alternative methods to the ordinary least squares for parameter estimation should be examined in the existence of multicollinearity. Recently, researchers have been investigating the estimation of the parameters of measurement error models in the presence of multicollinearity. In this study, Liu estimation approach, which is recommended for the linear regression, is applied to the measurement error models when the multicollinearity exists. In order to eliminate the multicollinearity in measurement error models, three estimators one of which is unrestricted and the others are restricted are proposed. The performances of the proposed estimators are inquired by theoretical and numerical evaluations.

Key Words: Liu estimator, restricted Liu estimator, measurement error, multicollinearity.

References

- Cheng, C.L. and Van Ness, J.W. (1999). Statistical Regression with Measurement Error, Arnold, London.
- Fuller, W.A. (1987). Measurement Error Models. Wiley, New York.
- Gleser, L.J. (1992). The importance of assessing measurement reliability in multivariate regression. Journal of the American Statistical Association, 87 (419):696-707.
- Ghapani, F., Rasekh, A.R., Babadi, B. (2016). The weighted ridge estimator in stochastic restricted linear measurement error models. Statistical Papers doi:10.1007/s00362-016-0786-3.
- Liu, K. (1993). A new class of biased estimate in linear regression, Communications in Statistics. Theory and Methods. 22: 393--402.
- Saleh, A.K.M.D. and Shalabh, (2014). A ridge regression estimation approach to the measurement error model, Journal of Multivariate Analysis. 123: 68-84.

¹ Corresponding author e-mail: gustundag@cu.edu.tr

Application of Statistical Methodologies for Customer Churn: A case study of Kuwait Telecommunication Sector

Mohammad Abiad^{1,2}, Sorin Ionescu¹

¹Faculty of Entrepreneurship, Business Engineering and Management, University Politehnica of Bucharest, Bucharest, Romania

²College of Business Administration, American University of the Middle East, Kuwait.

Abstract

Nowadays, the main objective of all businesses is to achieve a sustainable profitability. Companies are seeking to determine the factors, that has direct effect on their profit and try to analyze them in order to gain advantages on their competitors. With this stiff competitive market, retaining customers becomes the main target of mobile service providers instead of attracting new customers, since the cost of retaining a customer is considered to be very low compared to the cost of attracting a new customer as mentioned earlier by many scholars. Hence, Customer Churn Analytics is one of the important factors that companies should focus on to support their objective of achieving a sustainable profitability. In Telecommunication sector, the business exists to serve customers, and analyzing the customer churn will give the managers an idea about the likelihood of a customer to leave the company and what factors significantly leads to increase the customer churn. The main objective of this paper is to provide an important data analysis to enhance the business in service sectors by listing the different factors that can affect the decision of a customer to leave the company, that is studying the customer Churn analytics, and to give the reader an idea of how such data could be statistically analyzed.

At first, a list of factors that play a vital role in customer churn analytics is determined through the review of different researches done in this field, and then the appropriate statistical methodology is followed to analyze the customer churn based on these factors. Different Statistical techniques can be implemented for such purpose, starting with defining the appropriate sample size for a certain level of test power, followed by data collection and data cleaning, then displaying the data graphically and finally doing the appropriate regression analysis. This paper lists the suitable statistical techniques to be applied when analyzing Customer Churn and shows the statistical results of such analysis using statistical software such as SPSS and MINITAB. In addition, the interpretation of the output and how it should be read properly will be shown in order to give the reader an idea of how such data can be analyzed.

Key Words: Telecommunication, customer churn, statistical techniques, binary logistic regression.

A Proposed Algorithm for Predictive Maintenance Using Statistics

Mohammad Abiad^{1,2}, Sorin Ionescu¹

¹Faculty of Entrepreneurship, Business Engineering and Management, University Politehnica of Bucharest, Bucharest, Romania.

²College of Business Administration, American University of the Middle East, Kuwait.

Abstract

Predictive maintenance makes it possible for equipment users and manufacturing companies to evaluate the operating condition of equipment, helps in identifying failures, or predicts when the next possible error of an equipment is going to happen. Whenever you can diagnose or expect equipment error, you can schedule maintenance in ahead of time, effectively manage inventory, minimize downtime, and enhance the operational potency. A workflow to build a predictive maintenance algorithm is to create a detection model (for condition tracking) or a forecast model (for prediction) starting with data that defines your scheme in a range of good and defective circumstances. The first step in evolving predictive maintenance algorithm is data collection whether using sensors or other source or using data simulation in some cases, especially for limited failure data, due to low occurrence of errors because of regular maintenance.

This paper shows the role of statistics in each stage of the predictive maintenance algorithm; these stages are defined as data collection, data cleaning, Defining Condition Indices and improving the detection or prediction models before the implementation. In data collection, data could be recorded using tracking sensors and usually located in database or specialized software such as Hadoop, or data could be simulated when there is limited number of failure by using the Simulink model. In data cleaning stage, cleaning outliers and missing values, as well as applying data transformation in some cases are applied in order to prepare for the next stage, which is identifying condition indices. Statistical techniques is used to determine condition indices that result from simple analysis by finding some numerical measures such as the average over time, or complex analysis that explains or describes the changes in field over time. These condition indices could be generated by applying different techniques such as time-series analysis. Once the condition indices are defined, different statistical techniques are applied to determine the present status of the system or to predict its future possible failure and hence to test the models before the algorithm implementation stage.

Key Words: Predictive maintenance, error tracking, time to failure, data acquisition, statistical techniques.

A Study on Diagnostics for Two-level Multilevel Regression Models

Yasin Altinisik¹, Emel Cankaya²

¹ Department of Statistics, Sinop University, Sinop, Turkey

² Department of Statistics, Sinop University, Sinop, Turkey

Abstract

This study aims to construct a two-level multilevel regression model in investigating the reading achievement test scores of students, with a comprehensive diagnostic test evaluation.

Multilevel regression modelling is a well-established procedure for analysing data sets containing continuous outcomes and multilevel data structures. It is well known that this procedure is robust to the violation of the assumption of the classical regression analysis for the residuals to be independent. However, diagnostic checks for the assumptions of the normality and homoscedasticity of the residuals, also the linear form of the relationship between the outcome and predictors are still crucial for a two-level regression model, since it incorporates uncertainty into analysis at both levels which are the student and the school levels respectively in the present study. We employ diagnostic tests and display additional graphs to evaluate whether the model assumptions are met for the fitted model. We take the nested-structure of the data, i.e. students (N=10320) are nested within schools (J=160), into account by focusing on the effects of aggregate and student level predictors on the outcome that is the reading achievement test scores of students.

Findings indicate that, at the student level, vocabulary achievement test score, age, and gender (and their interactions) predict the reading achievement test scores of students, when a both intercept and slope varying model was utilized to estimate the parameters for the model under consideration.

We conclude that age negatively and vocabulary test score positively influence the outcome, the reading achievement test scores of students, for both males and females. Moreover, the interactive effect of vocabulary test scores with age and gender have a positive influence, while the interaction between age and gender presents a negative influence on the outcome.

Key Words: Reading achievement, nested-data structure, multilevel regression, model assumptions, diagnostic tests and graphs.

Parameter Estimation of Capital Asset Pricing Model Through Quantile Regression: An Application on Karachi Stock Exchange

Amna Riaz¹, Shahid Kamal², Nauman Riaz Chaudhry³, Mohsin Riaz⁴

¹Department of Statistics, University of Gujrat, Gujrat, Pakistan

²CSAS, University of Punjab, Lahore, Pakistan

³Department of Computer Science, University of Gujrat, Gujrat, Pakistan

⁴Department of Information Technology, University of Gujrat, Gujrat, Pakistan

Abstract

Important features of two import sectors (Chemicals, Oil and gas) of KSE are explored. Daily closing prices of ten years from July 1, 2004 to February 28, 2014 with exclusion of missing working days due to the internal financial crises of Pakistan for the ten firms considering five of each sector are taken. For the KSE-100 and ten firms of considered sectors, daily returns are obtained from daily basis closing prices. KSE-100 index is expected return of market. One year (12-month) treasury- bills (T-bills) return rate is considered benchmark of risk free rate (RFR). Excess returns are calculated for ten firms using RFR. For the evaluation of the CAPM validity, significance of simple regression intercept and fitting of polynomial was tested and the results were consistent with CAPM also values of R^2 showed only factor which is priced by market is beta i.e. most of the risk is systematic risk. Also the estimated expected returns of CAPM are calculated for all ten firms. CAPM parameter is estimated at quantiles ($q = 0.05, 0.1, 0.25, 0.40, 0.5, 0.6, 0.75, 0.9, 0.95$).

Key Words: Quantile regression, Karachi stock exchange, capital asset pricing model, returns, risk free rate.

A Mathematical Programming Procedure for Examining Biasing Parameters of Two Stage Two Parameter Estimator

Selma Toker¹, Nimet Özbay²

¹Department of Statistics, Çukurova University, Adana, Turkey

²Department of Statistics, Çukurova University, Adana, Turkey

Abstract

Two stage least squares has some adverse effects on estimation of simultaneous equations model when the problem of multicollinearity exists and thus it is no longer preferable. The multicollinearity is usually the reason for large values of both mean square error and length of an estimator. Özbay and Toker (2018) suggested a more efficient two stage two parameter estimator as an alternative to the two stage least squares estimator. Also, optimal estimators of the biasing parameters are derived theoretically in the paper of Özbay and Toker (2018). However, numerical techniques can be more attractive for selecting the optimal values of the biasing parameters (Ebaid et al., 2017). Within this framework, we develop a mathematical programming approach to specify two biasing parameters of the two stage two parameter estimator, simultaneously. By this approach, the biasing parameters are determined by minimizing the mean square error under the constraint that the length of the two stage two parameter estimator is less than the length of the two stage least squares estimator. Then, we test the performance of our mathematical programming method by means of a numerical example. In addition, we compute the optimal estimates of the biasing parameters and estimated values of mean square error and length of the foregoing estimators.

Key Words: Mathematical programming, multicollinearity, simultaneous equations model, two parameter estimation.

Acknowledgements

This paper is supported by Çukurova University Scientific Research Projects Unit with Project Number: FBA-2019-11884.

References

- Vinod, H.D., Ullah, A. (1981). Recent Advances in Regression Methods, Marcel Dekker, New York, Inc.
- Toker, S. (2018). "Investigating the Two Parameter Analysis of Lipovetsky for Simultaneous Systems", Statistical Papers, DOI: 10.1007/s00362-018-1021-1.
- Ebaid, R., Farghali, R., Abo-El-Hadid, S. (2017). "A mathematical programming approach for Liu-type estimator, Advances and Applications in Statistics, vol. 50, no. 4, pp. 293-314.
- Klein, L.R. (1950). Economic Fluctuation in the United States, 1921-1941, John Wiley and Sons, Inc., New York.
- Özbay, N., Toker, S. (2018). "Multicollinearity in simultaneous equations system: evaluation of estimation performance of two-parameter estimator", Computational and Applied Mathematics, vol. 37, no.4, pp. 5334-5357.

¹Corresponding author e-mail: stoker@cu.edu.tr.

Point Estimation for Chen Distribution based on Ranked Set Sampling

Muhammet Cavdar¹, Buğra Saraçoğlu²

¹ PhD Candidate, Selcuk University, Graduate School of Natural and Applied Science, Department of Statistics, Konya, Turkey

² Professor, Selcuk University, Science Faculty, Department of Statistics, Konya, Turkey

Abstract

In this study, we examine the problem of point estimation for unknown parameters of Chen distribution under Simple Random Sampling (SRS) and Ranked Set Sampling (RSS). First, maximum likelihood (ML) estimators for unknown parameters of Chen distribution under SRS and RSS are acquired. Then, Bayes estimators under squared error loss function for SRS and RSS are obtained by using Tierney-Kadane approximation. A Monte-Carlo simulation study is performed to compare ML and Bayes estimators under SRS and RSS. In the simulation study, ML and Bayes estimators based on SRS and RSS are compared in terms of mean squared error (MSE) and bias. Finally, relative efficiencies (REs) based on SRS and RSS are also compared.

Key Words: Bayesian estimation, Chen distribution, Monte-Carlo simulation, ranked set sampling, Tierney-Kadane approximation.

¹ muhammetcavdar@tuik.gov.tr

² bugrasarak@selcuk.edu.tr

An Improper Form of Weibull Distribution for Competing Risks Analysis with Bayesian Approach

Ahmad Reza Baghestani¹, Fatemeh Sadat Hosseini-Baharanchi²

¹Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

²Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran

Abstract

In survival analysis, individuals may fail due to multiple causes of failure called competing risks setting. Parametric models such as Weibull model are not improper that ignore the assumption of multiple failure times. In this study, a novel extension of Weibull distribution is proposed which is improper and then can incorporate to the competing risks framework. This model includes the original Weibull model before a pre-specified time point and an exponential form for the tail of the time axis. A Bayesian approach is used for parameter estimation. A simulation study is performed to evaluate the proposed model. The conducted simulation study showed identifiability and appropriate convergence of the proposed model. The proposed model and the 3-parameter Gompertz model, another improper parametric distribution, are fitted to the acute lymphoblastic leukemia dataset.

Key Words: Competing risks, parametric survival modeling, Weibull distribution, improper distribution, Bayesian analysis.

¹Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

²Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran

Developing a First Order Two Parameter Estimator for Generalized Linear Model

Selma Toker¹, Gülesen Üstündağ Şiray², Muhammad Qasim³

^{1,2} Department of Statistics, Çukurova University, Adana, Turkey

³ Department of Statistics & Computer Science, University of Veterinary & Animal Sciences, Lahore, Pakistan

Abstract

The generalized linear models were defined by Nelder and Wedderburn (1972) and these models allow us to fit regression models for univariate response data which follow a very common exponential family of distribution. The unknown regression coefficients of the generalized linear models are estimated by the maximum likelihood estimator. However, in the existence of multicollinearity, the variance of the maximum likelihood estimator becomes inflated and the statistical inferences based on the maximum likelihood method may not be reliable. In this study, we develop a first order two parameter estimator which combines the advantages of ridge and contraction estimators in the generalized linear models by extending the work of Özkale and Kaçiranlar (2007) in the linear model. The superiority of the first order two parameter estimator to the maximum likelihood, ridge and Liu estimators is investigated with regard to the mean square error criterion. We also examine some optimal estimators of biasing parameters. In addition to the theoretical comparisons, the performance of the estimators is judged by numerical evaluations where the mean square error is considered as a performance criterion.

Key Words: Generalized linear model, two parameter estimator, multicollinearity, first order approximation.

References

- Özkale, M.R., Kaçiranlar, S. (2007). "The restricted and unrestricted two parameter estimators", *Communications in Statistics-Theory and Methods*, 36(15): 2707–2725.
- Nelder, J. A., Wedderburn, R. W. (1972). "Generalized linear models", *Journal of the Royal Statistical Society: Series A (General)*, 135(3): 370-384.
- Hoerl, A.E., Kennard R.W. (1970). "Ridge regression: Biased estimation for nonorthogonal problems", *Technometrics*, 12: 55–67.
- Segerstedt, B. (1992). "On ordinary ridge regression in generalized linear models", *Communications in Statistics-Theory and Methods*, 21(8): 2227-224.
- Kurtoğlu F, Özkale M.R. (2016). "Liu estimation in generalized linear models: application on gamma distributed response variable", *Statistical Papers* 57(4): 911–928.
- Liu, K. (1993). "A new class of biased estimate in linear regression", *Communications in Statistics-Theory and Methods*, 22: 393-402.
- Mayer, L.S., Willke, T.A. (1973). "On biased estimation in linear models", *Technometrics*, 15(3): 497–508.

¹ Corresponding author e-mail: stoker@cu.edu.tr

Adaptive MCMC for Generalized Method of Moments with Many Moment Conditions

Masahiro Tanaka¹,

¹Graduate School of Economics, Waseda University, Tokyo, Japan; Director-General for Policy Planning (Statistical Standard), Ministry of Internal Affairs and Communications, Tokyo, Japan

Abstract

A generalized method of moments (GMM) estimator is unreliable when the number of moment conditions is large, that is, it is comparable or larger than the sample size. While a number of provisions for this problem is proposed in classical GMM literature, the literature on its Bayesian counterpart (i.e., Bayesian inference using a GMM criterion as a quasi-likelihood) has paid scant attention to this problem. This study fills this gap by proposing an adaptive Markov Chain Monte Carlo (MCMC) approach to a GMM inference with many moment conditions. Particularly, this study focuses on the adaptive tuning of a weighting matrix on the fly. Our proposal consists of two elements. The first is the use of the nonparametric eigenvalue-regularized precision matrix estimator, which contributes to numerical stability. The second is the random update of a weighting matrix, which substantially reduces computational cost, while maintaining the accuracy of the estimation. A simulation study and real data application are then presented to illustrate the performance of the proposed approach in comparison with existing approaches.

Key Words: Bayesian analysis, generalized method of moments, many instruments problem, adaptive Markov chain Monte Carlo, nonparametric eigenvalue-regularization precision estimator.

¹Address: 1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050 Japan, Email: gspddlnt45@toki.waseda.jp.

Optimal Bus-Garage Allocations and Garage Capacity Analysis Using a Map Interface

Uğur Eliiyi¹, Selcen Coşkun²

¹ Department of Computer Science, Dokuz Eylül University, Izmir, Turkey

² Department of Mathematics, Dokuz Eylül University, Izmir, Turkey

Abstract

Effective management of public transportation services, especially in metropolitan areas such as İzmir, involves several performance indicators. One of these is the incurred energy cost for operating the urban transportation system. Accordingly, our focus in this study is the minimization of dead mileage fuel consumption cost in public bus transportation.

Dead mileage or deadhead trip is the total idle distance covered by the buses with no passengers, between their parking garages and their route termini. Namely, it covers both the pull-out trips in the mornings that are from garages to the beginning terminals of their routes, and the pull-in trips in the nights that are from the ending terminals of their last services to the garages. Travel demand variations within the day are also considered, so that some of the buses must make extra pull-in and pull-out trips to and from possibly different garages before parking at their night garages.

The bus transportation authority is responsible for scheduling all trips in the network, which consists of 326 routes, 1,557 buses, 32 parking garages, and is also the larger one of the two operators serving the network. The other operator owns 328 buses of the overall fleet and one of the garages. The fleet comprises three different bus types, which are solo, articulated and midibus, mainly with respect to passenger capacities. These also differ significantly in terms of fuel consumption rates.

Considering possible scenarios including garage capacity expansions, the cancellation of ones with lower utilization rates, the fuel consumption rates of different types of buses, and the probable centralization of the network to be operated by only one firm, five mathematical models are developed for finding the optimal bus-garage allocations and parking capacities, which adapt similar ones in recent literature. Current bus schedule provides the models with the required number of buses, corresponding to the peak demand in the mornings or to the idle resources within the day due to the off-peak hours' service frequency levels. Thus, there is also the dynamic allocation logic using bi-directional inventory flow approach, besides the capacity constraints.

The optimal solutions obtained using IBM CPLEX Studio 12.8.0 are presented, and possible improvements are summarized in comparison with the current dead mileage cost. The bus-garage matchings based on actual locations, and garage utilization rates are displayed using Google Maps platform, providing the decision makers a visual tool for planning their next step for reducing fuel costs.

Key Words: Public bus transportation, fuel consumption, parking garage allocation, tactical transport planning, deadhead trip, garage capacity planning.

¹ ugur.eliiyi@deu.edu.tr, +90 232 3019515

² selcencskn@gmail.com

Re-evaluation of Subtypes of Positional OSAS by Clustering Algorithms

Karaibrahimoglu Adnan¹, Ozturk Onder²

¹Asst. Prof., Dept. of Biostatistics and Medical Informatics, Faculty of Medicine, Suleyman Demirel University, TURKEY.

²Prof., Dept. of Chest Diseases, Faculty of Medicine, Suleyman Demirel University, TURKEY.

Abstract

Obstructive sleep apnea syndrome (OSAS) is a common sleep disorder that occurs in approximately 5–10% of the general population, and characterized by excessive daytime sleepiness, disruptive snoring, recurrent episodes of apnea (no airflow) or hypopnea (partially obstructed airflow) and nocturnal hypoxemia. The subtypes of positional OSAS (PPs) are defined by conventional classification determined by apnea-hypopnea index (AHI). However, there were not enough studies about classification and characterization of PPs in the literature, and PPs subtypes show different clinical characteristics according to this classification. The aim of this study is to determine new subtypes of PPs by data mining algorithms.

After approval of the Ethics Committee, the study was admitted by 514 patients with OSAS. The training set with 24 attributes containing demographic and clinical features were analyzed by K-means clustering, C&RT and CHAID decision tree algorithms with gain plots. The algorithms were run by RStudio programming. Chi-square test was used for cross-validation and Kappa statistics was used to compare the re-evaluated values with classical values. In all methods, two clusters for PPs were obtained and CHAID algorithm gave us the most accurate results. The value for AHI nodes in CHAID was considered as a cut-off value, and cross-validated with the cut-off value obtained by AUC-ROC analysis with high accuracy (92%). Data mining techniques are very useful in health sciences for helping diagnosis, treatment and prognosis although the training set was not so big. As well as other methods, CHAID algorithms described the exact clusters of PPs. Therefore, specific treatments should be developed for new subtypes of PPs considering the centers of 14 significant attributes.

Key Words: Positional sleep apnea, clustering, decision tree, data mining.

¹ adnankaraibrahim@gmail.com

² dronderturk@gmail.com

Detection of Related Species by Denoising Phylogenetic Trees

Nebahat Bozkus¹, Stuart Barber²

¹Department of Statistics, University of Giresun, Giresun, Turkey

²Department of Statistics, University of Leeds, Leeds, UK

Abstract

One of the popular methods to summarize a molecular data set is clustering; in this way, related species can be detected. There are many classification indices proposed in the literature, but the performance of these indices decreases if different species share similar features. In addition, these methods may need user defined thresholds. We propose a new algorithm which automatically detects related species at phylogenetic trees by denoising lifting transformation of trees.

Our algorithm estimates its own threshold, so it removes the manual intervention on the choice of a threshold. In addition, the nature of our algorithm can deal with cases where there is no clear separation between different species. We compare the performance of our algorithm with some other available indices in the literature using some simulated and real data sets.

Key Words: Phylogenetic trees, hierarchical clustering, cluster validity indices, wavelets, lifting.

¹ nebahat.bozkus@giresun.edu.tr.

² S.barber@leeds.ac.uk.

Examination of Aggregating Clustering Validation Indexes on the Football Players Performance Data Set

Serhat Emre Akhanlı¹, Oya Kalaycıoğlu²

¹Univeristy College London Department of Statistical Science, PhD, London, UK

²Bolu Abant İzzet Baysal University Faculty of Medicine Department of Biostatistics and Medical Informatics, PhD, Bolu, Turkey

Abstract

In cluster analysis, an important question is how to choose an appropriate clustering method and how to determine the best number of clusters. For this aim, researchers should consider what data analytic characteristics the clusters they are aiming at are supposed to have. Following that, different clustering validation indexes can be evaluated, which is crucially dependent on the aim of clustering.

Several validation criteria that refer to different desirable characteristics of a clustering were introduced in the literature and the user can be interested in some of these criteria rather than just one of them. In this respect, a new concept that generates clusters by using new designed random clustering algorithms was generated in order to standardise the different characteristics so that users can aggregate them in a suitable way specifying weights for the various criteria that are relevant in the clustering application at hand.

In this study, the idea of random clustering is discussed on the football player performance data, in which a dissimilarity matrix was constructed between players ($n = 1500$) with the combination of several dissimilarity matrix in order to map football players information by using multidimensional scaling and to cluster football players. In order to make the optimal number of clusters necessarily to be informative and practicable for football squads and managers, it was anticipated to have a large number of clusters; in other words, a small group of players in each cluster. However, the existing clustering criteria as well as aggregation of clustering validation indexes did not provide strong differentiation between clusterings with large number of clusters. Therefore, validating the selection of different clustering solutions were conducted was considering some external information. The decisions were informed by interviews with some football experts, who are especially experienced with the analysis of football player performance, and by questioning people with different opinions who are familiar with this specific context. The survey questions were asked to 13 football professionals including the head coach, the assistant coaches, the football analysts and the scouts from Istanbul Basaksehir FC and some Turkish journalists. In the final phase, the best clustering solution was chosen based on the survey results from the football experts point of view.

Key Words: Cluster analysis, clustering validation indexes, football data, random clustering, aggregation.

Performance of Variance Ratio Tests in Presence of Heteroskedasticity: Application to Stock Returns Data in Pakistan

Nuzhat Aftab¹, Muhammed Sohail²,

¹CSAS, University of the Punjab Lahore, Pakistan

²CSAS, University of the Punjab Lahore, Pakistan

Abstract

The purpose of this article is to test the efficiency of the firms registered in the food sector of Pakistan stock market using different variance ratio tests. We also tried to explore the performance of variance ratio tests under heteroskedasticity and autocorrelation. The variance ratio tests are applied to the daily and weekly observed and corrected returns from Pakistan stock market. It is found that series are more efficient when corrected. In addition, it is found that weighted variance ratio tests are more consistent under both heteroskedasticity and homoskedasticity.

Key Words: Stock returns, heteroskedasticity, autocorrelation, bootstrap, variance ratio tests.

The Effect of Uncertainty Shocks on Global Housing Markets

Mei-Se Chien ¹, Nur Setyowati^{2,3}

¹Department of Finance and Information, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

²Department of International Business, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

³Department of Accounting, Politeknik Negeri Semarang, Semarang, Indonesia

Abstract

This paper investigates how different uncertainty shocks affect international housing prices. We set up a model of housing prices with 4 indices of uncertainty shocks, including macroeconomic instability, financial instability, economic policy uncertainty (EPU), and geopolitical risk (GPR), and panel GMM method, Granger test, and impulse response analysis are used to estimate these effects with the sample data of 56 countries in Asia, America, and Europe. Based on the full sample and different regional group, the main results are shown as follows.

First, based on GMM estimating results of the full sample, macroeconomic instability and economic policy uncertainty can cause positive effects on housing prices instability, but the effects of financial instability and geopolitical risk are converse. However, all the effects of four uncertainty indices are significant.

Second, the effects of four uncertainty indices present regional differences by using GMM estimation. In Europe, the effects of all four uncertainty indices are significantly positive; In America, economic policy uncertainty and geopolitical risk lead to significantly positive impacts, and macroeconomic instability is insignificant, while financial instability will bring about significantly negative effect; In Asia, there are significantly positive effects from two indices, macroeconomic instability and financial instability, and insignificant effect from economic policy uncertainty, but negative effect from geopolitical risk.

Third, based on the results of Granger test and impulse response analysis, macroeconomic instability can cause significant impact on housing prices instability in most of three regions, but the impacts of other three indices display inconsistent results in different regions, implying macroeconomic instability is the most important factor to affect housing prices instability in most of countries.

Key Words: Housing prices, macroeconomic instability, financial instability, economic policy uncertainty, and geopolitical risk.

¹ First author, email: cms@nkust.edu.tw

² Corresponding author, email: setyowati.first@gmail.com

Examination of Cereal Production, Cereal Yield and Land Under Cereal Production in Turkey

Melis Temiz¹, Nilsu Coskun², Semra Cakmak³, Ozlem Senvar⁴

¹Marmara University, Department of Industrial Engineering, Istanbul, Turkey

²Marmara University, Department of Industrial Engineering, Istanbul, Turkey

³Marmara University, Department of Industrial Engineering, Istanbul, Turkey

⁴Marmara University, Department of Industrial Engineering, Istanbul, Turkey

Abstract

Cereal production is an important source of income and a significant part of the nutritional resources. Due to the impact on the economic policies and having a great place in terms of employment and agriculture, investigating the variations in cereal production and productivity is necessary for Turkey. In this study, cereal productivity, cereal production, and land under cereal production are considered and examined together in order to have results for decision makers to develop strategies from wide-ranging perspectives. The aim of this study is to examine the cereal yield (kg per hectare), cereal production (metric tons) and land under cereal production (hectares) data in Turkey between years 1961 and 2014. Trend-based forecasting has been utilized to have predictions up to the year 2022. Descriptive statistics, inferential statistics, and graphical methods have also been included by using Minitab 17. According to the results of this study, it was observed that there is a strong positive linear relationship between cereal yield and cereal production while land under cereal production and cereal yield have a negative weak linear relationship. Concerning the demand growth, production is predicted to continue growing with fluctuations, and reach 34 million by 2022. Cereal yield is also expected to increase and reach 3121,61 (kg per hectare). It is observed that there is a decrease in land under cereal production from the year 2005 to 2014 implying that more technological improvements, cultivation techniques, qualified seeds should be considered for further directions. We believe that the results of this study can act as a guideline for policymakers.

Key Words: Statistical data analyses, regression analysis, trend based forecasting, correlation analysis.

¹melistemizz@gmail.com

²nilsucoskun@hotmail.com

³semracakmak023@gmail.com

⁴ozlem.senvar@marmara.edu.tr, ozlemsenvar1@gmail.com

The Dependent-Conditional Value-at-Risk Forecast for Aggregate Risk Models

Bony Josaphat¹, Khreshna Syuhada²

^{1,2}Statistics Research Division, Institut Teknologi Bandung, Bandung, Indonesia

Abstract

Value-at-Risk (VaR) forecast and its extension have been the system for risk performance indicator in the finance and insurance industries. Many efforts have been devoted by author to seek and develop methods, models, and other risk measures associated to VaR. In this paper, we propose another risk measure, namely Dependent-Conditional VaR or D-CoVaR, for the case of aggregate risk models. This alternative risk measure is basically obtained due to the fact that VaR is not coherent whereas the Co-VaR may have value quite far from VaR. Our proposed D-CoVaR is calculated by mean of the target risk, conditional on the associated risk, in which we need Copula to derive joint distribution function. A numerical analysis is carried out to illustrate our D-CoVaR where the associate risk is either a dependent random variable or a parameter of the target risk.

Key Words: Comonotonicity, Copula, conditional VaR, FGM family, Pareto distribution.

¹ Presenter, author.

² Co-author.

Point Cloud Matching with a Vertex Weighted Graph Kernel

Pelin Kaya¹, Mehmet Ali Balcı², Ömer Akgüller³

¹ Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

² Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

³ Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

Abstract

One of the significant methods in the engineering areas such as reverse engineering, deformation monitoring, and building information model is the surface reconstruction based on point cloud data. As the modern 3D measurement technologies are growing remarkable scale, a dense point cloud can be obtained rapidly by scanning the surface of a physical object or landscapes. Hence, point cloud matching is one of the key technologies of several scientific measurement. Most of the point cloud matching without landmark used the iterative closest point algorithm. In this study, we present a machine learning based graph kernel for matching point clouds. Our method is based on a vertex weighted nearest neighbour graph representation of given point clouds. The vertex weights are chosen to be geometric features.

Key Words: Point cloud matching, manifold learning, Laplacian matrix, graph kernels.

¹ pelinkaya@posta.mu.edu.tr

² mehmetalibalci@mu.edu.tr

³ oakguller@mu.edu.tr

Testing the Hypothesis of Environmental Kuznets Curve with Renyi and Tsallis Mutual Informations

Elif Tuna¹, Atıf Evren²

¹ Yildiz Technical University, Faculty of Sciences and Literature, Department of Statistics, Istanbul-Turkey

² Yildiz Technical University, Faculty of Sciences and Literature, Department of Statistics, Istanbul-Turkey

Abstract

Global warming and climate change have been one of the most important environmental problems in the last two decades. The Environmental Kuznets Curve (EKC) hypothesis postulates the existence of an inverse U-shape relationship between per capita gross domestic product (GDP) and Carbon dioxide emissions (CO₂). This article investigates the validity of the Environmental Kuznets Curve (EKC) hypothesis in a cross-section of 131 developed and developing countries with data from 2014. Nonlinearity means the evidence of EKC hypothesis in regression analysis. So testing linearity goes to the conclusion or not of EKC hypothesis. An F test or a comparison between linear determination coefficient and correlation ratio may be helpful. For detecting nonlinearity with independent of strict assumptions we proposed a method based on mutual information between partial residuals as a measure of nonlinearity. This method is based on comparing the mutual information between residuals obtained by removing linear dependence from the original ones and residuals obtained by regressing curvilinear effect variable X₂ on linear effect variable X. The last method is based on mutual information measure suggested by Tanaka, N., et al.,. In addition to their study we also showed that nonlinearity can be detected better with Rényi and Tsallis mutual informations because of their flexibilities due to α parameter. In this paper we studied these tests and then compared the results obtained by each test. By determining the nonlinearity between the variables we have shown that the EKC hypothesis is supported. The result is important because it may have important policy implications.

Key Words: Nonlinearity, entropy, Rényi mutual information, Tsallis mutual information, EKC hypothesis.

¹ Yildiz Technical University, Faculty of Sciences and Literature, Department of Statistics, Istanbul-Turkey

² Yildiz Technical University, Faculty of Sciences and Literature, Department of Statistics, Istanbul-Turkey

Depth Based Nonparametric Multivariate Tolerance Regions

E. Burcu Mamak Ekinci¹, İhsan Karabulut²

¹Computer Engineering, Baskent University, Ankara, Turkey

²Statistics Department, Ankara University, Ankara, Turkey

Abstract

Tolerance intervals cover a fixed p proportion of the population for a given confidence level $(1 - \alpha)$. Although not widely referenced among practitioners of statistics tolerance intervals and regions are important tools often utilized in areas such as engineering, manufacturing, and quality control. There are various types of tolerance intervals which have any distributional assumptions or not. If a sample is from a continuous distribution that does not fit a parametric model or fits a parametric model for which tolerance intervals are difficult to obtain, then nonparametric tolerance intervals provide an important convenience. The depth notion allows to build a nonparametric tolerance region in a natural way as the order statistics do for nonparametric tolerance intervals.

In this presentation, we introduce the approach of Li and Liu (2008, *The Ann. Statist.* Vol.36, No.3, 1299-1323) with an application.

Key Words: Tolerance intervals, multivariate tolerance regions, data depth.

Predicting Academic Performance of Foreign Students via Machine Learning Algorithms

Özge Altıntaş¹, Furkan Başer², Ömer Kutlu¹, Ayşen Apaydın²

¹Department of Educational Sciences, Ankara University, Ankara, Turkey

²Department of Actuarial Sciences, Ankara University, Ankara, Turkey

Abstract

In recent years, there has been an increase in the number of foreign students seeking admission into universities. Universities that accept students from abroad also expect international students to meet the necessary requirements. As a condition of admission to higher education, universities generally require high school diploma grades with international recognition and scores obtained from internationally recognized tests. The scores from the tests developed for the purpose of selecting and placing students should have the power to predict the students' future academic performance.

Modern universities' primary challenge in this phase is to thoroughly evaluate their performance and to build a strategy for further development and future actions. Universities should concentrate more on the profile of admitted students, becoming conscious of the characteristics of the students based on the data received. This paper describes a machine learning framework to develop a prediction model for academic performance of foreign students at Ankara University based on their personal and pre-university characteristics. The main conclusion is that students' achievement and success could be enhanced more efficiently via educational data mining techniques.

Key Words: Education, academic performance, predictive modeling, machine learning.

Investigation of Possibility of Being Poor in Turkey with Probit Analysis

Tuba Yakıcı Ayan¹, Nurdan Değirmenci²

¹ Karadeniz Teknik Üniversitesi, İİBF, Ekonometri Bölümü, Trabzon, Türkiye.

² Recep Tayyip Erdoğan Üniversitesi, FUBYO, Uluslararası Ticaret ve Lojistik Bölümü, Rize, Türkiye

Abstract

In this study, the demographic variables affecting poverty in Turkey, degree of impact and direction are specified with the probit analysis. The analysis data were compiled from the 2017 household income and living conditions survey by Turkish statistical institute (TÜİK).

The dependent variable was defined as two categories as “poor” and “non-poor” according to 2017 poverty line. The explanatory variables in the analysis were determined based on the literature and the possibilities of obtaining data.

Initial univariate analyses revealed that the formal marriage and marital status variables were not related to poverty. The probit model obtained with the remaining variables is as follows. The chi-square value of the model showed that the model had a good fit to the data.

$$PROBIT_i = 0,762 + 0,448HP - 0,351A - 0,816G + 0,618AS - 0,900EL + 0,119HS$$

where

HP: Household population,

A: Age of the household head (0-44 young, 45-59 middle age, 60- aged according to World Health Organization classification before 2018),

G: Gender of the household head (1: male, 2: female),

AS: Main activity status of the household head (1: employed, 2: looking for a job 3: retired-quit, 4: other),

EL: Educational level of the household head (1: not literate, did not finish school, primary school; 2: Primary, secondary, high school, vocational technical high school; 3: School, faculty, master, doctorate),

HS: General health status of the household head (1: Very good, good; 2: moderate; 3: bad, very bad).

For the interpretation of model coefficients, marginal effects approach with mean values of explanatory variables were used. The findings are as follow:

When the number of households increases by 1 person, the probability of poverty increases by 32,21%.

The probability of poverty decreases by 25.24% when the age of the household head increases from young to middle-aged or from middle-aged to aged.

The probability of poverty in families with female household heads is 58.67% less likely to be than with male head households.

The probability of poverty in families with head households who search for a job is 44.43% more likely to be poor than those who are currently employed.

As the education level of the household head increases, the probability of being poor decreases by 64.71%.

As the general health status of the household head deteriorates, the probability of the family being poor increases by 8.56%.

Key Words: Poverty, probit.

Investigation of Item Bias of the Foreign Student Examination

Burcu Hasacebi¹, Yksel Terzi², Zafer Kk³

¹Statistic and Computer Science, Karadeniz Technical University, Trabzon, Turkey

²Statistic, Ondokuz Mayıs University, Samsun, Turkey

³Statistic and Computer Science, Karadeniz Technical University, Trabzon, Turkey

Abstract

The primary task of a developed measurement tool is to measure the latent trait intended to be measured without mixing it with other traits. This will increase the validity of the measurement tool. Otherwise, adding variables other than the trait of interest to the items and presence of components in test items that will result for and against the respondent will lead to item bias (Eren Suna, 2012). Such items reduce the validity of the measurement tool and thus compromise the accuracy, sensitivity and objectivity of the results obtained. Bias, with the most common definition in the literature, is a systematic error that causes a subgroup to be advantageous or disadvantageous to an item (Shepard, 1981). According to the psychology dictionary (2018), bias is expressed as the tendency or predisposition to be in favor of or against something. Therefore, in this study, item bias was investigated by using the data obtained from Karadeniz Technical University 2019 Foreign Student Examination (FSE). In order to determine item bias, the Differential Item Function (DIF) analysis was performed. Analyses were made by using item reactions of Turkish and Iranian respondents. Based on these reactions, comments were made on the bias of the items in terms of gender and sociocultural aspects. In the study, Breslow-Day statistics and Logistic Regression method, which is one of the CTT based techniques, are used and one of the IRT-based techniques, Lord's χ^2 statistics and Raju's Area Measurements methods were used. According to the results, items containing DMF were determined and expert opinion was consulted for these items. In addition, according to the results obtained, the comparison of the methods based on the CTT and IRT were made.

Key Words: Item, item bias, differential item unction, classical test theory, item response theory.

References

- [1] Eren, H. S. (2012). Madde yanlılıđı: Test maddeleri, yanıtlayıcılara performanslarını gstermeleri iin eđit řans sađlıyor mu? *PiVOLKA*, 22(7), 15-16.
- [2] Shepard, L.A. (1981). Identifying bias in test items. In B.F. Green (Ed.), *Issues in testing: Coaching, disclosure and ethnic bias*. San Francisco, CA: Jossey Bass.
- [3] Karakař, S. (2018) Psikoloji szlđ. <http://www.psikolojisozlugu.com/>

Asymmetric Open Vehicle Routing Problem with Heterogeneous Fleet and Time Limitations

Nilay Yaren Çelikoklar¹, Uğur Eliiyi²

¹Department of Mathematics, Dokuz Eylül University, Izmir, Turkey

²Department of Computer Science, Dokuz Eylül University, Izmir, Turkey

Abstract

In this study, a personnel bus routing problem is considered. The objective is the minimization of total vehicle cost subject to capacity and time restrictions. In addition to the case-specific routing problem, other variants are examined with different objectives, constraints and input parameter values.

The problem owner is a faculty and the customers of the service shuttles correspond to bus stops, where several academic and administrative staff are to be picked up from in a timely manner, thus reaching their workplace before the beginning of their daily shifts. The personnel size considered is nearly 200, residing in different neighbourhoods of the city. For reducing the problem size, the dispersed home addresses are clustered into bus stops of the public transit network. The fleet is to be designed from alternative vehicles having three different capacities, namely with 16, 27 and 45 passenger seats. The outsourced vehicles have different rental costs regardless of having the same capacities. The distances between the stops are taken to be the traveling times, and they are asymmetric due to the road network directions and traffic conditions in the morning hours. The origin of the route of each vehicle is to be determined such that the total travel time till reaching the faculty building, which is the last stop for all routes, should fall within the 1-hour limit. There is no time window constraint, but rather the pick-up times will be determined. Moreover, each stop must be visited by only one vehicle, corresponding to no-split delivery. Lastly, as the return to depots is not considered, our problem becomes an open vehicle routing problem. The proposed integer programming formulation is solved optimally using IBM CPLEX Studio 12.8.0. The time-distances between bus stops and the routes are calculated, generated and displayed using Google Maps platform.

As different variants of the problem, split delivery, minimizing the maximum travel time and total weighted traveling time are separately considered with different problem sizes. Since all versions are NP-Hard, multi-stage neighbourhood search algorithms are proposed, combining random and adaptive approaches. A computational study is carried out with realistic location data, the results are reported for each solution approach and compared with optimal solutions when possible.

Key Words: Vehicle routing problem, time constraints, cost optimization, split delivery, mathematical modeling, heuristics.

¹ yarenly.ce@gmail.com

² ugur.eliiyi@deu.edu.tr, +90 232 3019515

A Gravity Model: Maritime Transport Costs and Their Impact on Economy

Nurten Çalışkan¹, Burcu Özge Özaslan Çalışkan²

¹International Trade Department, Istanbul Kultur University, Istanbul, Turkey

²Transportation and Logistics Department, Istanbul University, Istanbul, Turkey

Abstract

Due to transportation expenses, trade costs still protects its importance because no matter how companies try, they cannot zeroize their transportation cost. Maritime transportation is the most effective mode to move large quantities of cargo over long distances.¹ That is why maritime most preferred transportation mode in international trade. The Gravity Model is a popular model in international trade theory. It named after Newton's "law of gravity." This model says that, trade between countries can be explained by their GDP sizes and transportation costs which are varied by the geographical distance between them. Turkey carries out approximately 70% of its foreign trade using maritime transport. In 2017, Turkey major trading partner countries for exports were Germany, United Kingdom, United Arab Emirates, Iraq and United States and for imports, they were China, Germany, Russian Federation, United States and Italy.² Making use of these countries maritime transport costs, their GDP's and their distances between each other, this study investigate the role of maritime freight costs and its effects on these countries economy by using gravity model.

Key Words: Maritime, transportation cost, gravity model, GDP, Turkey.

¹ https://transportgeography.org/?page_id=1731

² <https://wits.worldbank.org/CountryProfile/en/Country/TUR/Year/2017/TradeFlow/EXPIMP/Partner/by-country>

Big Data Analytics: Future Perspective and Scope in IT Applications

M. A. Rizvi¹, Anu Mangal²

¹Department of Computer Engineering and Applications, National Institute of Technical Teachers' Training and Research, Bhopal – 462002, M.P. India.

²Department of Electronics & Communication Engineering, UIT-RGPV, Bhopal – 462033, M.P. India.

Abstract

The present work is focused on Big data and its analytics which is high volume and magnitude of data (terabytes and petabytes) used and analyzed by different firms for visualizing customer's interest, meet the demands and increase the quality services and efficiency of service provided so that organizations can have effective decision making related to their product and services. Big data is also a high velocity data. Velocity refers to the rate at which big data is generated and analyzed. Data obtained from various sources such as text, audio or video is usually in "raw" or unstructured form. So, for processing such data firstly data management must be done followed by its analytics/analysis leading to its interpretation. Recording and acquisition of data is done from several sources and then effectively represented for analysis of data for the specific purpose. Finally modeling and interpretations of "structured" data is done which provides organizations the areas for their improvement for the betterment of their products and services. Data must be mined carefully because it is complex since it is derived from myriad of sources. Thus, big data analytics can be viewed as a sub- process of insight extraction from big data. Big data is noisy, abundant, mostly unstructured and unreliable. So, development of "data science" to analyze big data for the conversion of "raw data" into "trustworthy" data is the real task. This paper deals with big data, importance of big data analysis and its necessity and also deals with the application of big data analytics in different sectors.

Key Words: Big data analytics, data science, data visualization, data mining.

¹ E-mail: marizvi@nittrbpl.ac.in

² E-mail: anu.mangal001@gmail.com

Big Data Analythics for Green Supplier Selection Under Sustainable Conditions

Abdullah Yıldızbaşı¹, Yağmur Arıöz²

¹Yıldırım Beyazıt University, Industrial Engineering, Ankara, Turkey

²Yıldırım Beyazıt University, Industrial Engineering, Ankara, Turkey

Abstract

With globalization, the rapid depletion of vital resources, the rapid development of customer and environmental awareness, and the increasing importance of environmental protection, green supply chain management has begun to replace the traditional supply chain. Green supply chain is a type of supply chain management where strategies for producing environmentally sensitive products or services are united. Companies not only reduce their environmental negative impact, but also increase efficiency, giving them a huge competitive advantage in innovation and processes. Therefore, green suppliers that adopt the green thinking philosophy have become an indispensable element of the chain. Therefore, green suppliers have become more sensitive to green supplier selection as issues covering the environmental approach, which is a valuable aspect of sustainability, become important. Today, a number of criteria exist for selecting suppliers for this environmentalist approach. Green supplier selection is a problem that involves both qualitative and quantitative criteria and needs to strike a balance between these criteria. The technology that develops in this problem, the requirements of our time, the criteria used in relations between firms are increasing every day with the adoption of the environmental approach and these criteria produce data quickly. This data, which we call Big Data, has different dimensions called 7V. The criteria that are matched in these aspects aim to achieve goals such as correct decision making, accountability, traceability of data results, easy detection of changes occurring and achieving the most meaningful results through multidimensional significance. However, big data is believed to provide vital decisions that can transform business processes through well-informed decisions that result in both material and intangible benefits such as increased profitability and performance, defect reduction and improved customer satisfaction. This is why the role of big data is inevitable on green suppliers.

Key Words: Green supplier selection, big data, sustainability.

Evaluate the Performances of Restaurants and Hotels Listed in BIST 100 Under Balanced Scorecard Perspective Via Plithogenic Decision Making Approach

Çağlar Karamaşa¹

¹Department of Business Administration, Anadolu University, Eskişehir, Turkey

Abstract

Performance analysis has been accepted as an essential issue for firms and managers. Papers related that has been increased in recent years. Balanced Scorecard method consisting four perspective provides a more comprehensive and more holistic view of the organization and that the model can be used according to any perspective chosen for a particular application. Balanced Scorecard demonstrates how managers can assess all the important criteria and achieve success in one area by spending or making sacrifices in different areas. Additionally balanced scorecard provides important advantages in terms of clarifying the processes, determining the problems within the enterprise and resolving the urgent needs by strengthening internal communication. In this study the performance of restaurants and hotels listed in BIST are measured under the balanced scorecard perspective by handling 4 main criteria. Plithogenic sets are used to obtaining criteria weights for firms. Plithogenic sets which are introduced by Smarandache in 2017 as a generalization of crisp, fuzzy, intuitionistic and neutrosophic sets can be defined as creation, development and progression of new entities from composition of contradictory or non-contradictory multiple old entities. Elements of plithogenic set are characterized by attribute values having contradiction degree values for obtaining more accurate results from the judgments of decision makers. For this purpose plithogenic aggregation operations based on plithogenic set are used to combine decision makers' views for analyzing firms under balanced scorecard perspective. Results show the most and the least essential balanced scorecard criteria for analyzing the performance of restaurants and hotels.

Key Words: Performance analysis, balanced scorecard, Plithogeny.

Iterative Robust Bayesian Estimation of Regression Modelling based on Ramsay-Novick Distribution

Mutlu Kaya¹, Emel Cankaya², Olcay Arslan³

¹ Department of Statistics, Sinop University, Sinop, Turkey

² Department of Statistics, Sinop University, Sinop, Turkey

³ Department of Statistics, Ankara University, Ankara, Turkey

Abstract

Creating robustness within a Bayesian estimation procedure has long been an attractive research area so as to provide protection against departures from error distribution assumptions or the subjectivity of prior beliefs. Utilizing heavy-tailed distributions for both is a possible way to achieve robust inferences. Ramsay-Novick (RN) distribution happens to be a candidate heavy-tailed distribution for this purpose. However, there is a lack of theoretical evaluation in practicing with this distribution for Bayesian regression analysis.

This study therefore aims to suggest an iterative computation approach for the estimators of bayesian regression model with RN distributed errors and the prior information represented by Student-t or Normal distribution. We here illustrate the method using a real-data application. DIC is used for the model performance assessment and the results are discussed from a robustness point of view.

Key Words: Robust Bayesian estimation, Ramsay-Novick distribution, student-t prior, heavy-tailed distribution, iterative estimation.

Parameter Estimation of Dirichlet Distribution Based on Entropy

B. Sahin¹, A. Evren²

¹Yildiz Technical University, Institute of Natural Science, Department of Statistics, Davutpasa, Esenler, 34210, Istanbul-Turkey

²Yildiz Technical University, Faculty of Science and Literature, Department of Statistics, Davutpasa, Esenler, 34210, Istanbul-Turkey

Abstract

Entropy is a concept that originated in thermodynamics, and later, via statistical mechanics, motivated entire branches of information theory, statistics, and machine learning. Maximum entropy is the state of a physical system at greatest disorder or a statistical model of least encoded information. Nowadays, maximum entropy formalism is closely related to Laplace's famous principle of insufficient reasoning. In this study, the parameters of Dirichlet distribution have been estimated by two maximum entropy methods (namely, by the ordinary entropy method, and parameter space expansion method) as an alternative to maximum likelihood, and method of moments. Dirichlet distribution as the multivariate generalization of beta distribution is especially important for modeling categorical distributions. Hence its applications vary on a wide range from modeling cell probabilities of a contingency table to modeling income inequality. Thus, it is commonly used as the conjugate prior of multinomial distribution in Bayesian statistics. We have shown that in estimating the parameters of Dirichlet distribution, ordinary entropy method, and parameter space expansion method give the same results as the method of maximum likelihood, implying the asymptotic normality of the estimators. With the help of some simulations, the parameters of Dirichlet distribution are estimated by method of moments which probably benefits from the law of large numbers due to the large number of random observations simulated. In the applications parameter estimates found by maximum likelihood method and, hence the two maximum entropy methods converged to those found by the method of moments. Thus, we emphasize that these two methods can be used alternatively in modeling Dirichlet distribution.

Key Words: Dirichlet distribution, Beta distribution, principle of maximum entropy, maximum likelihood estimation, method of moments.

Multi-Criteria Decision Making in Site Selection for Nuclear Power Plants: Multi-Attribute Utility Approach

Berna Tuncer¹, Ayşen Dener Akkaya²

¹ Department of Statistics, Middle East Technical University, Ankara, Turkey

² Department of Statistics, Middle East Technical University, Ankara, Turkey

Abstract

Nuclear energy is an important alternative energy source. However, the construction of nuclear power plants (NPP) requires the consideration of various factors related to environmental, economic, socioeconomic, health and safety. Accordingly, selection of the most suitable site for the construction of a nuclear power plant yields to a multi-criteria decision making problem. In illustrating the implementation of the multi-criteria decision making methodology, the two candidate NPP sites in Turkey, namely: Akkuyu and Sinop are considered. Seismic hazards, tsunami, air temperature, high wind events, surface transportation (railways and/or roads), population density, availability of water sources for cooling purposes, ecological properties of the corresponding sites and land use are the main factors that are considered during the analysis. Among them, seismic hazard will be considered as the main factor. Because of various sources of uncertainties, probabilistic seismic hazard methodology is used for this criterion. For the remaining factors, single-attribute utility functions are constructed for each criterion and using them, multi-attribute utility functions are developed for each site. At the end of the analysis, overall utility values are obtained so that suitability of the sites for the nuclear power plant construction can be measured and these two sites can be compared with each other.

Key Words: Nuclear power plant, multi-criteria decision making, utility functions, site selection.

The Impact of Medical Tourism on the Turkey Economic Performance

Berhan Çoban
İzmir Katip Çelebi University, İzmir, Turkey

Abstract

Nowadays, tourism has recently become a popular global activity for many different purposes. Tourism may spur economic growth via its impact on foreign currency earnings, investments, tax revenue, and employment opportunities especially for many developing countries like Turkey. Among the sub-sectors in tourism Medical or Health tourism has become one of the fastest growing sub-fields in the tourist sector in recent years.

The most important factors to this development are globalization, cost of surgery, access to international transport and the greater awareness of patients. Therefore, medical tourism is becoming more common. Turkey is heading toward positioning the country to be an attractive destination for medical tourism. Turkey is accepted as a health tourism destination country with JCI-accredited hospitals that are 44 in number currently. Thanks to these hospitals Turkey provides better quality treatment at lower costs when compared to Europe, the U.S. and other western countries.

The main purpose of this study is to examine medical tourism's role in Turkey's long-term economic variable using a time series econometric approaches namely cointegration, causality analysis. Tourism and macroeconomic data are affected by factors such as economic crisis, security problems and natural disasters as well as many other time series. These factors may cause structural breaks, heteroscedasticity in time series as well as affect the results of tests performance.

In this study, conventional and recently developed tests unit root and cointegration tests are performed to the existence of cointegration between medical tourism and economic variables.

The tourism and health tourism revenues, total international visitor arrivals data are used for medical tourism sector and for economic variable data of GDP, exchange rate etc. The study reveals strong evidence that there is long term relationship between medical tourism sector and macroeconomic series of Turkey under some structural breaks.

Key Words: Medical/ health tourism, economic growth, cointegration.

Automatic Determination of K Parameter Value of K-Means for Time Series Clustering

Fatma Ozge Ozkok, Mete Celik,
Computer Engineering Department, Erciyes University, Kayseri, turkey

Abstract

Cluster analysis is a widely used data mining technique for time series data in many areas such as biology, finance, security and health. Time series include large number of sequential points so analysis of them is a challenging and difficult process. K-means is simple and successful clustering algorithm which divides data into user-defined k partitions. However, the determination of the value of k parameter is not easy for each dataset. One of the most used method for determination of the value of clusters (such as, K) is using clustering validation indexes. There are lots of internal and external indexing methods that are used for finding suitable cluster numbers on characteristics of dataset. In this study we aim to determine the value of K parameter of K-means algorithm and so propose a new hybrid approach using four internal clustering indexes that are Dunn, Silhouette, C-index and Davies-Bouldin. Our experiments show that the proposed approach is promising for clustering time series data.

Key Words: Clustering, time series, k-means, automatic clustering, clustering validation.

Predicting Sepsis via Unsupervised Machine Learning Using Bedside Device Data in Hospitals

Gokhan Silahtaroglu¹, Zehra Nur Canbolat², Kadir İdin³, Nevin Yılmaztürk⁴

¹Management Information Systems Department, Istanbul Medipol University, Istanbul, Turkey.

²Management Information Systems Department, Istanbul Medipol University, Istanbul, Turkey.

³Department of Anesthesiology and Reanimation, Istanbul Medipol University, Istanbul, Turkey

⁴Pusula Enterprise Business Solutions- Research and Development Centre, Istanbul, Turkey

Abstract

Sepsis infection is one of the most important causes of deaths especially in intensive care units. It is seen as a serious global health crisis. A very early sepsis infection diagnosis has vital importance to prevent deaths. The objective of this study is to perform a new model of unsupervised machine learning using respiratory rate, arterial blood pressure systolic, heart rate, temperature, and some other device data. The necessary data for this study have been obtained from MIMIC-III v1.4 version of the international clinical database. The database contains patients' data collected between 2001-2012. Unsupervised machine learning has been performed via both Fuzzy-C and deep learning algorithms on patients' data diagnosed as sepsis positive and sepsis negative. The machine generated a label for each cluster at the end of the training session. Some labels represent sepsis positive and some sepsis negative. The training has both been supervised and validated via Xie Beni index, partition coefficient and partition entropy values. The dimensions of cluster representatives have been reduced to a two dimensional array by Principal Component Analysis in order to visualize the results of training on screen to be used in intensive care units at hospitals. It is shown that the model is robust enough to plot patients' data blended with training results on screen which is helpful for practitioners to make a diagnosis.

Key Words: Septic infection, unsupervised machine learning.

¹ gsilahtaroglu@medipol.edu.tr

² zncanbolat@medipol.edu.tr

Clustering of Vector Autoregressive Models: An Application on Air Pollution

Aytaç Pekmezci

Department of Statistics, University of Muğla Sıtkı Koçman, MUĞLA, TURKEY

Abstract

Vector Autoregressive (VAR) model is the statistical model used for estimating linear relationship between time series. The major information about this relationship is attained from parameters of VAR model. In this study, the relationships between time series of PM_{10} and SO_2 measured from 111 monitoring stations at Turkey are investigated and it is tried to detect the monitoring stations which behave similar in terms of this relationship. There are two interrelated objectives in here: i) getting information about all the stations by monitoring fewer number of the stations selecting as representative, ii) decreasing monitoring cost and information redundancy. This study is the first with regard to taking into account more than one air pollutants simultaneously. The procedure followed to achieve these objectives consists of four steps: i) estimating VAR models between SO_2 and PM_{10} time series, ii) determining optimal number of clusters by using Xie-Beni (XB) index, iii) clustering of the parameters of VAR models, iv) determining representative and redundant stations. As a result of the analysis, it is observed that the monitoring cost and information redundancy in PM_{10} and SO_2 monitoring stations at Turkey can be decreased at rate of 70%.

Key Words: Air pollution monitoring networks, vector autoregressive model, Granger causality test, unit root tests.

Time Series Forecasting via Dendritic Neuron Model Neural Network Trained by Particle Swarm Optimization

Ayşe Yılmaz¹, Ufuk Yolcu²

¹Department of Statistics, Giresun University, Giresun, Turkey

²Department of Econometrics, Giresun University, Giresun, Turkey

Abstract

Time series forecasting is a crucial problem encountered in a variety of fields. Although there are a great number of methods in the literature, they can be reviewed under two basic title as probabilistic and non-probabilistic methods. Probabilistic methods are generally based on statistical approaches. Non-probabilistic methods can be also divided into two basic group as fuzzy-based methods and computational based methods. Especially, as computational-based time series forecasting methods, different kinds of artificial neural networks (ANNs) have been widely and successfully used in the literature. While some of them use additive aggregation function, some of them use multiplicative aggregation function in the structure of their neuron models. Particularly, Sigma-Pi NN and Dendritic NNs, proposed in recent years, have also both additional and multiplicative neuron models, together. In this study, it is aimed that dendritic neuron model neural network (DNM-NN) is taken advantage of as a forecasting tool in financial time series forecasting problem. As part of this study, the training of DNM-NN is realized by particle swarm optimization. To evaluate the forecasting performance of DNM-NN, some real-world data sets have been analysed and the obtained results have been discussed with the results produced by some other time series forecasting methods including some traditional, fuzzy-based and computational-based methods.

Key Words: Dendritic neuron model, time series, particle swarm optimization, forecasting.

¹ ayse.yilmazz@gmail.com

² varyansx@hotmail.com

Financial Time Series Prediction via Convolutional Neural Network

Melih Kirisci¹, Ozge Cagcag Yolcu²

¹Ministry of Industry and Technology, Regional Development Administrations, Giresun, Turkey

²Industrial Engineering Department, Giresun University, Giresun, Turkey

Abstract

In recent decades a variety kinds of computational intelligence techniques have been widely used for financial time series prediction. Especially in the last studies deep learning based convolutional neural networks (CNN) and long short-term memory (LSTM) neural networks are getting popular day by day in time series prediction literature. In this study it is aimed that a CNN with a new structure/architecture is taken advantage of for financial time series prediction. For this purpose a CNN structure having three convolutional layers and five full connected layers has been created. In the convolutional layers, Relu and Elu activation functions have been preferred to get nonlinear relations. Input layer of created CNN is 1D and the size of input layer depends on lagged variables. To be able to evaluate the performance of proposed CNN structure in financial time series prediction, Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) has been analyzed and obtained prediction results are discussed with the results produced by some other state-of-the-art prediction tools such as different kinds of ANN, LSTM, fuzzy-based approaches, and some traditional methods.

Key Words: Convolutional neural network, financial time series, prediction, deep learning.

¹ melih.kirisci@dokap.gov.tr

² ozgecagcag@yahoo.com

Multiplicative Neuron Model Based on Threshold Value for Forecasting Problem of Time Series

Asiye Nur Yıldırım¹, Eren Bas², Erol Egrioglu³

¹ Department of Statistics, Giresun University, Giresun, Turkey

² Department of Statistics, Giresun University, Giresun, Turkey

³ Department of Statistics, Giresun University, Giresun, Turkey

Abstract

Multiplicative neuron models artificial neural networks, which are frequently used in time series forecasting problems and do not have the number of hidden layer unit problems, have an important place in artificial neural networks literature. In the studies about multiplicative neuron models artificial neural networks, the model obtained from training process is a single model. In this study, different from other studies, a threshold value is determined to obtain the output of the system of multiplicative neuron models artificial neural networks. According to obtained threshold value, it is determined which weights and biases are used for the calculation of the output of the network. For the training of multiplicative neuron model based on threshold value that means to determine the optimal weights and biases of the system the harmony search algorithm and particle swarm optimization algorithms were used separately. To analyse the performance of the proposed method Australian beer consumption time series data between the years 1956 and 1994 was analysed. The results obtained from these two models are compared with well-known methods in the literature. Besides, the performance of these two methods was evaluated among themselves.

Key Words: Artificial neural networks, multiplicative neuron model, threshold value, harmony search algorithm, particle swarm optimization.

A Comparison Study on Stock Exchange Forecasting

Onur Derya¹, Ali Zafer Dalar²

^{1,2} Department of Statistics, Giresun University, Giresun, Turkey

Abstract

Time series forecasting covers a wide range of real-life problems, which has important research value in various fields. In general, a time series is a collection of observations made in chronological order. Time series data has unique characteristics, e.g. a large amount of data, high dimensionality, and update constantly, so it is not easy to analyse and model them effectively. For decades, many researchers have been devoted to time series forecasting and have achieved important results in theory and applications. However, most of these studies do rely on mathematical equations or simulation techniques to describe the evolution of time series data. In the last ten years, artificial neural networks and deep learning methods have been using for time series forecasting.

In this work, we provide results of experiments exploiting different well-known approaches, namely autoregressive integrated moving average as a traditional method, multilayer perceptron as an artificial neural network method, and the long short-term memory as a deep learning technique to predict stock exchange time series. We compare the methods on stock exchange time series data via the best root mean square error (RMSE) value from one-step ahead forecasting.

Key Words: Forecasting, time series, deep learning, long short-term memory, ARIMA.

¹ Corresponding author: onur.derya@giresun.edu.tr

The Missing Link: Are Individuals with More Social Capital in Better Health? Evidence from India

Baris Alpaslan¹, Julide Yildirim²

¹Department of Economics, Social Sciences University of Ankara

²Department of Economics, TED University

Abstract

This paper extends a two-period Overlapping Generations (OLG) model of endogenous growth in which associations between human capital, social capital, and health outcomes are critically examined for a low income country, India. If individuals with higher level of human capital can build strong social ties and have more robust social networks, they are then less likely to have health problems and are therefore physically healthier. In an attempt to test the so-called relationship between the variables in question, a unique dataset, where micro-level data from the World Values Survey (WVS) and regional-level macro data from the Central Statistics Office of India were both utilized, was accessed. A three-equation model has been then estimated using the conditional mixed-process (CMP) method in order to explicitly address endogeneity issues. Our estimation results provide important insights into the theoretical thesis in several ways. Firstly, human capital has a favorable impact upon social capital, which in turn enhances self-reported health. Secondly, we provide a comparison of three main experiments: an increase in the share of public spending by region on education, social capital-enhancing activities, and health. The results confirm the positive effect of an increase in each form of government spending on outcome variables. Thirdly, the correlation coefficient between disturbances of these three equations turns out to be statistically significant, suggesting that there are unobserved factors, which can affect self-reported health, social capital and human capital variables.

JEL Classification Numbers: H51, H52, H59, I15, I25, O41, C30.

Key Words: Social and human capital, self-reported health, India, conditional mixed-process model.

¹ E-mail: baris.alpaslan@asbu.edu.tr

² E-mail: julide.yildirim@tedu.edu.tr

Predicting Patients Future Laboratory Test Result Patterns via A Novel Unsupervised Deep Learning Model

Gokhan Silahtaroglu¹, Zehra Nur Canbolat², Kadir İdin³, Nevin Yılmaztürk⁴

¹Management Information Systems Department, Istanbul Medipol University, Istanbul, Turkey.

²Management Information Systems Department, Istanbul Medipol University, Istanbul, Turkey.

³Department of Anesthesiology and Reanimation, Istanbul Medipol University, Istanbul, Turkey

⁴Pusula Enterprise Business Solutions- Research and Development Centre, Istanbul, Turkey

Abstract

Loads of tests have been conducted in hospitals everyday besides the device data which are generated constantly to be monitored by healthcare staff. All these data are used for the treatments of patients. Some of the test are not only expensive but also need expertise and a lot of time to be done. On the other hand, some are as cheap as free. Physicians often need those expensive and time consuming tests results urgently in order to make a diagnosis. However, this may not be possible every time. In this study, we have used 28.000 patients' laboratory and device data to train machine for predicting patients' future data patterns in a time series manner. The objective of the study is to predict and estimate patients test values which are expensive or difficult to produce in a short time. Parameters to be estimated have been chosen among others with respect to the difficulty to reach the test result and also importance of the parameter for an urgent diagnosis. A new machine learning model which blends unsupervised learning and deep learning has been designed to realise the study. Our results are promising enough to use the model to detect patients ours earlier before they are in a risky state.

Key Words: Diagnosis, unsupervised machine learning, time series, patients data.

¹ gsilahtaroglu@medipol.edu.tr

² zncanbolat@medipol.edu.tr

The Role of Social Capital in Environmental Protection Efforts: Evidence from Turkey

Julide Yildirim¹, Barış Alpaslan², Erdener Emin Eker¹

¹Department of Economics, TED University, Ankara, Turkey.

²Department of Economics, Social Sciences University of Ankara, Ankara, Turkey.

Abstract

The root causes of environmental problems are not only ecological. They are also linked to collective choices of individuals. Humans may control the environmental degradation by developing strategies through collective action. The role and importance of social capital in natural resource management has been recognized by the existing literature. Several studies provide empirical evidence that higher levels of social capital may positively affect individuals' behaviour towards natural resources management. This study is an attempt to investigate the environmental quality impacts of social capital and environmental local government spending taking spatial dimension into account for the time period 2008-2016, employing provincial data for Turkey. The variable for environmental degradation is particulate matter (PM10). An index of social capital has been constructed by employing principal component analysis, where the variables considered include provincial population growth rate, total number of doctors, entertainment and cultural expenditures, total number of higher education graduates. Following a traditional regression analysis, a general-to-specific approach has been adopted where spatial variations in the relationships have been examined utilizing the spatial econometric models.

Key Words: Environmental pollution, social capital, spatial econometric models.

Prediction of Failures in Plastic Extrusion Process with Deep Learning

Fatma Demircan Keskin¹, Ural Gökay Çiçekli², İsmail Doğukan İcli³

¹ Ege University/Department of Business Administration, Izmir, Turkey

² Ege University/Department of Business Administration, Izmir, Turkey

³ Ege University/Graduate Faculty of Social Sciences, Izmir, Turkey

Abstract

Predictive engineering is one of the most critical elements of smart manufacturing. This concept is based on collecting real-time and accurate data, analyzing this data with advanced analytical techniques, detecting the patterns and predicting the issues before they happen, and thus preventing failures.

Deep learning has gained considerable attention, especially in recent studies addressing extraction of information from process data, identification of patterns, and prediction of failures. In this study, the failure prediction problem in the plastic extrusion process will be addressed. In this process, a number of failure types, including edge tearing, die cleaning, die exchanging, filter cleaning can occur. Some of the critical parameters of the process are the temperature of dies, condition of equipment, line velocity, pressure, melt temperatures.

The application of the study will be carried out using actual measurement data on the failure types and factors collected from the plastic extrusion process of one of the leading insulation companies of Turkey operating in Izmir. Prediction performance of the applied methods will be evaluated by employing prediction accuracy, Type I, Type II errors, and area under curve criteria.

Key Words: Failure prediction, deep learning, plastic extrusion process, smart manufacturing.

¹ fatma.demircan.keskin@ege.edu.tr

² ural.gokay.cicekli@ege.edu.tr

³ dogukan.icli@gmail.com

Application of Statistical Process Control in Production Process: A Case Study in Cleaning Products Sector

Gülşen Aydın Keskin¹, Murat Çolak², Bengü Güngör³

¹Department of Industrial Engineering, Kocaeli University, Kocaeli, Türkiye

²Department of Industrial Engineering, Kocaeli University, Kocaeli, Türkiye

³Department of Industrial Engineering, Balıkesir University, Balıkesir, Türkiye

Abstract

Global competition forces businesses to meet customer needs much better, reduce costs and increase productivity. It is only possible for companies to maintain their existence and increase their market share in this competitive environment only by delivering the quality products and services they produce to their customers whenever they need them. Businesses can overcome competitive threats by focusing on their customers, applying continuous improvement in their operations. Therefore, quality, quality management and continuous improvement concepts are seen as a tool of competing with many competitors.

This study has been conducted in an enterprise which produces cleaning materials in Marmara Region. Defects in the work in process during production process deteriorate product quality, adversely affect production efficiency and increase production costs. In order to improve quality in enterprises, to create global competitive conditions and to reduce production costs, it needs to prevent poor quality. In this study, the parameters that cause production shutdowns due to the defects caused by work in process in production are examined in terms of process improvements by using Statistical Process Control Methods. In the study, checklist, pareto analysis, cause-effect diagram and Anderson Darling Tests, ANOVA One Way Tests, X-R control graphs, Process adequacy C_p , C_{pk} analysis methods were used. At the study, samples were taken during the production period of the related product and all quality properties of the work in process were tested. As a result of the study, process capability analysis has been done for weight, width resistance, length resistance and elongation values and; C_p , C_{pk} and PPM values were obtained. According to the C_p and C_{pk} values; it can make interpretation as the process adequacy is not suitable, the variability is reduced, and the process does not meet the specification value. Therefore, it was concluded that the process should be kept under control. In terms of PPM; production specification values of work in process should be checked. The accumulation in a certain area within the specification limits indicated that the manufacturer should only make the relevant improvements in the system operating at these limits.

Key Words: Statistical process control, defect of work in process, machine shutdown, process capability analysis

¹ gaydin@kocaeli.edu.tr

² colak.murat@kocaeli.edu.tr

³ bengu.gungor@balikesir.edu.tr

Evaluation and Estimation of Wind Speed with Data Mining Methods

Kadriye Ergün¹, Melike Sultan Karasu Aşnaz²

^{1,2}Department of Industrial Engineering, Balıkesir University, Balıkesir, Turkey

Abstract

Wind energy has a very important place among renewable energy sources, because wind is a natural, clean and sustainable power. Wind turbines are a system which transforms the movement energy in the wind into firstly mechanical energy and then into electrical energy. Obtaining electrical energy from wind turbines varies depending on several factors. One of these is wind speed, and energy is proportional to the cube of wind speed. For this reason, estimation of wind speed and detection of factors affecting wind speed have become very significant for energy efficiency, use of resources and right investment decisions.

In this study, data mining methods were used to estimate wind speed and show how factors affect energy production. We based on one year wind data. C4.5 algorithm, k-nearest neighbour algorithm, artificial neural networks, C-Means and association rules were applied to the obtained data by using Weka programme. In addition, select attribute was used to determine the relationships between the factors affecting wind speed.

Key Words: Wind energy, forecasting, data mining, artificial intelligence, classification.

Investigating the Effectiveness of Metaheuristic Approaches in Parameter Estimation via TOPSIS Method

Aynur Şahin¹, Nimet Yapıcı Pehlivan²

¹ Department of Statistic, Selcuk University, Konya, Turkey

² Department of Statistic, Selcuk University, Konya, Turkey

Abstract

Metaheuristic methods are known as effective approaches to solve some difficult optimization problems. These methods have been mostly used for the parameter estimation problem of some distributions in recent years. However, there are a few studies on deciding which metaheuristic method is better in the literature. In this study, TOPSIS method that is one of the Multi-Criteria Decision-Making methods has been proposed in order to compare the effectiveness of five different metaheuristic methods: Genetic Algorithm (GA), Differential Evolution (DE), Particle Swarm Optimization (PSO), Simulated Annealing (SA), and Artificial Bee Colony (ABC) in parameter estimation. Firstly, parameter estimation results for each method have been obtained with an extensive simulation study. Then, the TOPSIS method has been implemented by using these results to rank the considered metaheuristic methods in terms of solution quality, solution time, simplicity and robustness criteria. A real-life example is finally presented.

Key Words: Metaheuristic methods, parameter estimation, TOPSIS.

Determination of The Number of In-Port Transportation Vehicles by Simulation Modelling

Bengü G ng r¹, G l  en Aydın Keskin²

¹Department of Industrial Engineering, Balıkesir University, Balıkesir, Turkey

²Department of Industrial Engineering, Kocaeli University, Kocaeli, Turkey

Abstract

Efficient and uninterrupted operation of the ports and providing added value services to the customers have become a necessity brought by the competition. On the other hand, the waiting loads because of congestion in port effect on extending delivery times throughout the supply chain. In order to provide efficient service in ports, a performance measurement tool is needed to manage operations. There are basically four logistics processes in container terminals which are ship operations, transport, storage and handling of the container at the port site. The effectiveness of these main processes increases the competitiveness level of the specific port. At this point, optimization of the port processes is gained more importance because, each element affecting the basic processes contributes to overall system. Therefore, it is necessary to examine the operations and contribute the system at critical points. Moreover, port operators need to a flexible decision support tool that can easily operates logistic processes related to the cargo handling, in-port transportation and storage of the port, to measure the performance of terminal equipment, to provide timely information on the problems, and to compare different alternatives for finding best solution against the problems. Taking into account all of these reasons, simulation models can be such kind of useful tools since it has all the necessary features. Within the scope of this study, it was determined that the empty waiting times of the MAFI terminal tractors that take containers from SSG (Quay Crane) to RTG (Field Crane) are high in the port. In application phase, it was gradually examined and decided on how many MAFI should be assigned for an SSG. For each case, possible cycle times, vehicle usage rates, idle rates and operating rates were analyzed statistically in simulation program.

Key Words: Ports, container transportation, simulation modelling.

¹ bengu.gungor@balikesir.edu.tr

² gaydin@kocaeli.edu.tr

The Effect of Different Antibiotics Use on Prognosis of Patients who Hospitalized with Urinary Tract Infection: A Retrospective Study

Sevilay Kılınçkaya¹, Osman Gürdal²

¹ Süleyman Demirel Üniversitesi Isparta Sağlık Hizmetleri Meslek Yüksekokulu, Yaşlı Bakımı Anabilim Dalı, Isparta, Turkey

² Süleyman Demirel Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Isparta, Turkey

Abstract

The objective of this study was to determine the microorganisms that are responsible for patients with urinary tract infection (UTI), and to compare the parameters that affecting the prognosis and the duration of stay in hospital with several univariate, multivariate statistical methods, and finalize the conclusion.

In this study, an adult group was targeted, and they were given different antibiotics after collection of their blood and urine cultures during the hospital stay. We thus, investigated prognoses of the patients and their hospital stays via antibiotic usage. For this purpose, there were 124 people, who were aged between 18 and 95 year-old and hospitalized with the diagnosis of UTI between March 21, 2017 and March 21, 2018 in the Infectious Disease Service Department of Isparta City Hospital, Isparta.

Our result showed that the UTI was more common in women (54.8%) and was more frequent in ages between 76 and 95 years (44.4%). E. Coli pathogenic strains were dominant among the other genera. We found that the most effective antibiotics were realized as Cephalosporins among the others for the shorter length of stay in hospital (5.46 ± 3.30), although the Carbapenems (55.6%) were the most commonly prescribed by the physicians.

As a result, we retrospectively identified which type of antibiotics usage was appropriate for a better treatment of UTI, and therefore determined the length of hospitalization with statistical analyses.

Key Words: UTI, factor, prognosis, hospital stay.

Analyzing Hospital High Length of Stay Outliers

Pakize Yiğit¹

¹Istanbul Medipol University Medical Statistics and Informatics, Dr, Istanbul, Turkey

Abstract

The length of stay (LOS) outliers are an important indicator for hospital management and finance. DRG (Diagnosis Related Groups) is used for case mix system for hospital reimbursement and performance measurement. DRGs are highly correlated with length of stay of patients. The aim of the study was to analyze high LOS outliers with logistic regression, decision tree, and naïve bayes classifier in DRG data. The performance of these approaches was also compared. The data were collected from inpatients whose discharge period is from January to December in 2012 from twelve trainee hospitals in Istanbul, Ankara, Izmir. Length of stay outliers defined by the geometric mean plus two standard deviations. The variables were age, gender, comorbidity, DRG complexity, discharge status, type of operation. In all the inpatient episodes, it was found a proportion of 6.8% high LOS outliers.

Key Words: Classification, diagnosis related groups, length of stay outliers.

Receiver Operating Characteristic (ROC) Analysis in Clinical Diagnosis: A Diabetes Case

Ebru Dumlupınar¹, Osman Gürdal², Yusuf Bilen³

¹Department of Biostatistics and Medical Informatics in Adıyaman University, Adıyaman, Turkey

²Department of Biostatistics and Medical Informatics in Süleyman Demirel University, Isparta, Turkey

³Department of Internal Medicine in Adıyaman University, Adıyaman, Turkey

Abstract

²Clinicians and researchers in the field of health care make sound decisions, despite the incomplete information and uncertain results. Before these medical decisions, some performance evaluations must be made and the optimal threshold value should be calculated for an accurate classification for the disease. ¹Receiver Operating Characteristic (ROC) analysis used as an important tool for statistical calculations for making pre-evaluations. It is moderately implemented in the healthcare domain to determine how accurately a patient's diagnostic test can be distinguished whether he or she has a disease or free from it.

In this study, it was aimed to introduce the diagnostic evaluation criteria and ROC analysis to give information about its usage, and to make the application of the information given through clinical cases with different statistical programs. We, thus, to provide the fundamental conceptual framework and interpretation of ROC analysis help medical researchers to use it effectively.

For ROC analysis, we investigated diabetes as a clinical case. For diabetes, the different strengths of HbA1c, glucose, urea and creatinine tests were evaluated and examined. In the case of HbA1c test, which is used for diagnosis of diabetes, our ROC curve estimation showed that the highest Area under Curve (AUC) value was 0,746 with the lowest standard error 0,03 ($p < 0,001$). It is considered a fair discrimination. In general, there was no significant difference between AUC estimations obtained from MedCalc and LABROC according to the non-parametric and semi-parametric approaches of minor differences between AUC estimates, standard errors and confidence intervals. The highest difference was observed in the AUC estimation of creatinine tests ($p < 0,001$) and it can be stated that there is no significant difference among the others.

Key Words: ROC analysis, clinical case, diagnostic tests, LABROC.

¹ Metz CE. Basic principles of ROC analysis. Seminars in Nuclear Medicine 1978; 8:283-98.

² Pepe MK. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press 2003, ISBN 978-0-19-850984-4.

Considering Linear Constraints for Almon Two Parameter Ridge Estimation

Nimet Özbay¹, Selma Toker²

¹Department of Statistics, Çukurova University, Adana, Turkey

²Department of Statistics, Çukurova University, Adana, Turkey

Abstract

In linear regression model, Lipovetsky and Conklin (2005) proposed two parameter ridge estimator as an alternative to ordinary least squares estimator in the presence of multicollinearity. By setting additional linear constraints on the parameters, Şiray and Toker (2014) suggested restricted version of this estimator. To estimate distributed lag model which is our model in question Özbay (2018) defined Almon two parameter ridge estimator. It is an appealing estimator since its first biasing parameter recovers multicollinearity and the second one is responsible for quality of the regression fit. In this paper, we evaluate the linear restrictions for the Almon two parameter ridge estimation to introduce restricted Almon two parameter ridge estimator. Besides, an optimal biasing parameter as well as some properties of this new estimator is examined in theoretical. Then the theoretical findings are supported numerically.

Key Words: Almon ridge estimator, linear constraints, multicollinearity, distributed lag model, two parameter estimation.

Acknowledgements

This paper is supported by Çukurova University Scientific Research Projects Unit with Project Number: FBA-2018-10169.

References

- Almon, S. (1965). "The distributed lag between capital appropriations and expenditures", *Econometrica*, vol. 33, no. 1, pp. 178-196.
- Chanda, A.K., Maddala, G.S. (1984). "Ridge estimators for distributed lag models", *Communications in Statistics-Theory and Methods*, vol. 13, no. 2, pp. 217-225.
- Gültay, B., Kaçiranlar, S. (2015). "Mean square error comparisons of the alternative estimators for the distributed lag models", *Hacettepe Journal of Mathematics and Statistics*, vol. 44, no. 5, pp. 1215-1233.
- Hoerl, A.E., Kennard, R.W. (1970). "Ridge regression: Biased estimation for nonorthogonal problems", *Technometrics*, vol. 12, no. 1, pp. 55-67.
- Lipovetsky, S., Conklin, W.M. (2005). "Ridge regression in two parameter solution", *Applied Stochastic Models in Business and Industry*, vol. 21, pp. 525-540.
- Özbay, N. (2018). "Two-parameter ridge estimation for the coefficients of Almon distributed lag model", *Iranian Journal of Science and Technology, Transactions A: Science*, DOI:10.1007/s40995-018-0634-5.
- Üstündağ Şiray, G., Toker, S. (2014). "Restricted two parameter ridge estimator", *Australian&New Zealand Journal of Statistics*, vol. 55, no. 4, pp. 455-469.
- Vinod, H.D., Ullah, A. (1981). *Recent Advances in Regression Methods*, Marcel Dekker, New York, Inc.

¹Corresponding author e-mail: nturker@cu.edu.tr.

Modelling Disease Transmission Scenarios Using Field Data and Heavy Tailed Random Effects: The Zika Case

Zafer Bekiryazici¹, Tülay Kesemen², Mehmet Merdan³, Tahir Khaniyev⁴

¹Department of Mathematics, Recep Tayyip Erdogan University, Rize, Turkey

²Department of Mathematics, Karadeniz Technical University, Trabzon, Turkey

³Department of Mathematical Engineering, Gümüşhane University, Gümüşhane, Turkey

⁴Department of Industrial Engineering, TOBB University of Economics and Technology, Ankara, Turkey

Abstract

In this study, we investigate a deterministic compartmental model of Zika Virus transmission under random effects. Random effects enable the analysis of the random numerical characteristics of transmission, which cannot be modelled through deterministic equations. Data obtained from Zika cases around the world is used along with heavy tailed random effects to obtain new random variables for the parameters of the deterministic model. Finally, simulations of the model are carried out to analyse the random dynamics of Zika Virus transmission. Deterministic results are compared to results from the simulations of the random system to underline the advantages of a random modelling approach.

Acknowledgement: This work was supported by Research Fund of the Recep Tayyip Erdogan University. Project Number: FBA-2019-992.

Key Words: Zika Virus, Pareto Distribution, Random Differential Equation, Random Effect, Simulation

References

1. Kibona, I. E., & Yang, C. (2017). SIR model of spread of Zika virus infections: ZIKV linked to microcephaly simulations. *Health*, 9(08), 1190.
2. Merdan, M., Bekiryazici, Z., Kesemen, T., & Khaniyev, T. (2017). Comparison of stochastic and random models for bacterial resistance. *Advances in Difference Equations*, 2017(1), 133.
3. González-Parra, G., & Benincasa, T. (2019). Mathematical modeling and numerical simulations of Zika in Colombia considering mutation. *Mathematics and Computers in Simulation*.
4. Feller, W. (2008). *An introduction to probability theory and its applications* (Vol. 2, 2nd Edition). John Wiley & Sons.
5. World Health Organization, Zika Virus Fact Sheet, July 2018.
6. Arnold, B. C. (2015). Pareto Distribution. In *Wiley StatsRef: Statistics Reference Online* (eds N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri and J. L. Teugels).

¹ Corresponding Author: Asst. Prof. Zafer Bekiryazici, E-mail: zafer.bekiryazici@erdogan.edu.tr

Mixed Frequency Estimation by Smoothed Least Squares

Selma Toker¹, Nimet Özbay², Kristofer Månsson³

¹Department of Statistics, Çukurova University, Adana, Turkey

²Department of Statistics, Çukurova University, Adana, Turkey

³Department of Economics, Finance and Statistics, Jönköping University, Jönköping, Sweden

Abstract

Mixed data sampling (MIDAS) regression models introduced and analysed by Ghysels et al. (2005, 2006, 2007) incorporate data sampled at different frequencies and have more research interest nowadays. MIDAS approach is a parametric approach in which the shape of the lag distribution is determined with the use of some specific parametric functions. With these specifications, the parameters can be estimated by nonlinear least squares (NLS). However, the parametric functions may produce inaccurate approximations of the lag distribution in practice. In this context, Breitung and Roling (2015) suggested a nonparametric approach to define smoothed least squares (SLS) estimator. Our motivation in this paper is to supply some guidance for choosing the smoothing parameter of the SLS estimator. We examine several selection methods to reach a more appropriate level of smoothness in practice. In addition, the SLS estimator obtained by using different choices of smoothing parameter is compared numerically with the NLS estimator in the sense of prediction mean square error.

Key Words: Mixed data sampling, nonlinear least squares, lag distribution.

References

- Breitung, J., Roling, C. (2015). "Forecasting inflation rates using daily data: A nonparametric MIDAS approach", *Journal of Forecasting*, vol. 34, pp. 588–603.
- Ghysels, E., Santa-Clara, P., Valkanov, R., (2005). "There is a risk-return trade-off after all", *Journal of Financial Economics*, vol. 76, no. 3, pp. 509-548.
- Ghysels, E., Santa-Clara, P., Valkanov, R. (2006). "Predicting volatility: getting the most out of return data sampled at different frequencies", *Journal of Econometrics*, vol. 131, pp. 59–95.
- Ghysels, E., Sinko, A., Valkanov, R. (2007). "MIDAS regressions: further results and new directions", *Econometric Reviews*, vol. 26, no. 1, pp. 53–90.
- Mansson, K., Shukur, G., Kibria, B.M.G., (2010). "A simulation study of some ridge regression estimators under different distributional assumptions", *Communications in Statistics-Simulation and Computation*, vol. 39, no. 8, pp. 1639-1670.
- Özbay, N., Toker, S. (2019). "Efficiency of Mansson's method: Some numerical findings about the role of biasing parameter in the estimation of distributed lag model", *Communications in Statistics-Simulation and Computation*, DOI:10.1080/03610918.2018.1517215.

¹ Corresponding author e-mail: stoker@cu.edu.tr.

Unit-Lindley Mixed-Effect Model for Ratio Data

Hatice Tul Kubra Akdur¹

¹ Department of Statistics, Gazi University, Ankara, Turkey

Abstract

Linear mixed models (LMM) and generalized linear mixed models (GLMM) are commonly exploited for repeated measures data and longitudinal studies. GLMM and LMM take into account the dependence that usually occurs between observations in the same cluster or experimental subject whereas regression models that require independence between observations are usually misspecified especially in clustered or repeated measures datasets. Outcome variables in GLMM can be ratio such as migration, fertility and obesity rates which lie in the unit interval. Recently, unit-lindley distribution which derived from a transformation on the Lindley distribution has been proposed for modelling outcome variable in the bounded interval (0, 1) [1]. In this study, unit-lindley mixed model for bounded outcome in the unit interval is proposed. In the case of LMM, the likelihood function can be expressed in closed forms. However, the likelihood functions of GLMM cannot be derived in closed form and depend on integrals with respect to random effects [2]. Many techniques can be used to approximate log-likelihood function of GLMM for parameter estimation such as Laplacian approximation, marginal quasi likelihood method and adaptive Gaussian quadrature (AGQ) method. AGQ method and Laplacian approximation are applied to obtain maximum likelihood parameter estimations of unit-lindley mixed model.

Key Words: Unit-lindley, mixed-effect models, likelihood approximation, adaptive Gaussian quadrature.

References

- [1] Mazucheli, J., Menezes, A. F. B., & Chakraborty, S. (2019). On the one parameter unit-Lindley distribution and its associated regression model for proportion data. *Journal of Applied Statistics*, 46(4), 700-714.
- [2] Pinheiro, J. C., & Chao, E. C. (2006). Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15(1), 58-81.

Bitcoin Inefficiency Analysis

Mária Bohdalová¹, Michal Greguš¹

¹Department of Information Systems, Faculty of Management, Comenius University in Bratislava, Slovakia

Abstract

Nowadays, the Bitcoin market receives the attention of investors, policymakers, etc. This paper analyses the informational efficiency of the Bitcoin prices at a high-frequency level using Hurst exponent. We focus on both the Bitcoin return and its volatility. The strength of the long memory effect is analysed using the correlation between the increments of the returns or volatilities. In particular, we test the Efficient Market Hypothesis for Bitcoin in regulated cryptocurrency exchange Gemini market. We have used data at different frequencies (1 and 60 min and daily data) from 10/08/2015 to 07/07/2019. We propose fractal and statistical properties of the returns and volatility of the Bitcoin prices, too. Moreover, we have used overlapping and non-overlapping moving window analysis to examine the evolution of the informational efficiency of Bitcoin. Our study shows whether the shocks in Bitcoin prices are behaving as transient and whether persistence in the high-price period has been falling. The paper provides evidence of the presence of informational inefficiency in the Bitcoin market at higher frequency levels.

Key Words: Efficient market hypothesis, Hurst exponent, Bitcoin, high-frequency data

Catch-Up Performance of Turkish Manufacturing Firms and International Trade Effect

Aslihan Atabek Demirhan¹

¹Structural Economic Research Department- CBRT, Economist, Ankara, Turkey

Abstract

This paper investigates the productivity catch-up performance of Turkish manufacturing firms using a novel and detailed dataset. National frontier firms are observed to be larger, more export market oriented, and more productive firms. Estimation results reveal evidence for the statistically significant catch-up and learning effect for the manufacturing sector during 2013-2016 periods. Laggards are found to be benefiting more from regional frontiers in terms of productivity growth. Moreover, our empirical results show that the degree of trade integration accelerates the catch-up speed and learning of the firms from frontier.

Key Words: Firm-level data, productivity catch-up, panel data estimation methods.

Investigation of Effects of Some Cryptocurrencies on Bitcoin: Causality Analysis

Yüksel Akay Ünvan¹

¹Assoc. Prof., Ankara Yıldırım Beyazıt University, Faculty of Management, Banking and Finance Department, Turkey

Abstract

This study was conducted to see whether Bitcoin Cash, Ethereum, Litecoin and Ripple, 4 of the top 5 cryptocurrencies in the cryptocurrency market, have an effect on Bitcoin. Daily data between 23. 07.2017 and 22.09.2019 were used for the analyses to be carried out for this purpose. These data were taken from coinmarketcap.com and Eviews 9 program was used for analysis. Firstly, the general structure of these 5 variables and their relations between them were given. For these variables, stationarity analysis was performed with the ADF Test. Then, the lag length was determined in the VAR model for stagnated variables. After selecting the appropriate model, Johansen Cointegration Test was applied to the model. Due to the determination of cointegration equations, Vector Error Correction Model (VECM) was applied. In order to find causality relationships in this model, the Wald Test and VECM Granger Causality Test were performed. As a result of these tests, causality relationships from Bitcoin Cash, Ethereum, Litecoin and Ripple to Bitcoin were determined.

Key Words: Bitcoin, causality analysis, Cryptocurrency, VECM, Bitcoin cash, Ethereum, Litecoin, Ripple.

¹ aunvan@ybu.edu.tr

Drought Monitoring of Marmara Region using Novel Machine Learning Tools

O. Ozan Evkaya¹, Fatma Sevinç Kurnaz²

¹Matematik Bölümü, Atılım Üniversitesi, Ankara, Türkiye

²İstatistik Bölümü, Yıldız Teknik Üniversitesi, İstanbul, Türkiye

Abstract

Extreme weather events are harmful natural disasters that affect various regions and have certain environmental and economic results. Drought is one of the important and costliest disaster over the world, especially with the accelerated progress of climate change, its frequency of occurrence and negative impacts are rapidly increasing. It is crucial to initiate and sustain an early warning system to monitor and predict the possible impacts of future drought events. Vast number of drought indices exist in the literature in order to mitigate the effects of future dry periods and their effective projection is required to aid in the development of early warning system. Recently, with the rise of data driven models, various case studies are conducted by using Machine Learning (ML) algorithms instead of using pure statistical approaches. The main objective of this study is forming a comprehensive drought forecasting for Marmara Region, using weather stations. As an indicator, widely used drought index, Standardized Precipitation Index (SPI), is calculated for each location and various ML algorithms are implemented. For that purpose, novel and recent ML techniques are employed and the ability of new models are discussed. According to a pool of goodness of fit (GOF) tests, the performance of the considered models is compared and future projections are analysed for each station.

Key Words: Forecasting, drought, drought index, machine learning, future drought.

Modeling of BACE-1 Inhibitory Activity using MLR, ANN and SVM Techniques

Bashirulla Shaik¹

National Institute of Technical Teachers' Training & Research, Bhopal-462202,
Madhya Pradesh, India.

Abstract

The present study deals with the modeling of spirocyclic β -Site amyloid precursor protein cleaving enzyme 1 inhibitors (BACE1) using molecular descriptors which are responsible for the BACE-1 activity. The study has been undertaken by three techniques MLR, ANN, and SVM. A multiple linear regression (MLR) analysis has shown that physicochemical and 2D autocorrelation descriptors of these compounds are the major influencing factors of their activity and docking study has shown that compounds can form hydrogen bonds with the receptor and have effective steric interactions. The MLR model fits the train set with $R^2 = 0.857$ while in ANN and SVM with higher values of $R^2 = 0.874, 0.887$, respectively. SVM model shows improvement to estimate the BACE-1 activity of the trained data, while in test set ANN have higher R^2 value than those of MLR and SVM techniques. Using the MLR model, some new compounds have been proposed that have higher potency than the existing ones. Docking analysis was performed on these predicted compounds with the enzyme (PDB id: 4JOO), all the predicted compounds were found to have several hydrogen bonds with the receptor as well as to have some hydrophobic interactions.

Key Words: BACE-1, QSAR, ANN, SVM, Docking analysis.

¹ Email: bshaik@nitttrbpl.ac.in

Household Expenditures Pattern on Durable and Nondurable Goods and Services: A Comparison of Multiple Regression and Neural Network Model

Khadija Fatima¹, Muhammad Suleman², Zahoor Ahmed³, Abeeda Fatima⁴, Amna Riaz⁵,

^{1,2,5} Department of Statistics, University of Gujrat, Gujrat, Pakistan

³ Department of Statistics, University of Lahore, Sargodha, Pakistan

⁴ School Education Department, Govt. of Punjab, Faisalabad, Pakistan

Abstract

The purpose of this paper is to estimate the household (HH) expenditures patterns on durable and nondurable goods and services. Data is taken from Federal Bureau of Statistics on Household Integrated Economic Survey (HIES) 2015-16 Pakistan. For analysis purpose, multiple regression and neural network multilayer perceptron model has been used and also compared the performance of both. Annual expenditures on non-durable goods and services significantly affected by family size, no. of earners, proportion of out of school children's, proportion of out of school children's, HH assets and HH income. Annual expenditures on durable goods are significantly affected by no of earners, HH assets and HH income and remittances.

Key words: Household expenditures, durable and nondurable goods, HH assets, HH income and neural network model.

Stochastic Frontier Analysis to Measure the Impact on the Productive Efficiency of Lemon Producers in Michoacán, through the Incorporation of the Economic Costs of Insecurity.

Laura Plazola¹, Mariela Sandoval¹

¹ Quantitative Methods Department, University of Guadalajara, Zapopan, México.

Abstract

The aim of this work is to analyse the causal relationship among the costs of insecurity and the loss of productive efficiency of lemon producers in the Apatzingan Valley, Michoacán, México.

The method consisted of relating the costs of insecurity with the impact of productive efficiency through the parametric Stochastic Frontier Analysis (AFE) model. The results show the costs caused by extortion by organized crime, as the trigger for 11% of the technical inefficiency of the model; whereas, as a finding, the production retention mechanism generated an increase in technical efficiency of 12%.

Finally, these variations allowed a stochastic estimation of a 74% technical efficiency. Therefore, these findings may imply an increase in technical efficiency of 85%, if extortion is eliminated and the retention mechanism is maintained.

Key Words: Statistical applications, econometrics, stochastic frontier, insecurity.

¹ University Center of Economics and Administration, University of Guadalajara.

Asymptotic Results for Stationary Characteristics of Random Walk with a General Interference of Chance

Ozlem Ardic Sevinc ^{1,2}, Tahir Khaniyev ¹

¹Department of Industrial Engineering, TOBB University of Economics and Technology, 06560, Ankara, Turkey

²Department of Structural Economic Research, Central Bank of the Republic of Turkey, 06050, Altindag, Ankara, Turkey

Abstract

In this study, a semi-Markovian random walk process $(X(t))$ with general interference of chance is constructed and investigated. The key point of this study is the assumption that the discrete interference of chance has a general form. Under some conditions, it is proved that the process $X(t)$ is ergodic and the exact forms of the ergodic distribution and characteristic function of the process $X(t)$ are obtained. Then, weak convergence theorem for the ergodic distribution of the standardized process $W(t) \equiv X(t)/\lambda$ is proved when $\lambda \rightarrow \infty$ and limiting form for ergodic distribution of the process $W(t)$ is obtained. Finally, under some assumptions for the discrete interference of chance in general form, two term asymptotic expansions for all moments of the ergodic distribution are obtained. Additionally, coefficient of variation, skewness coefficient and kurtosis coefficient of the ergodic distribution are computed.

Key Words: Discrete interference of chance, semi-Markovian random walk, weak convergence, asymptotic expansion.

Mixed Lasso Estimator for Stochastic Restricted Regression Models

Huseyin Guler¹, Ebru Ozgur Guler¹

¹Department of Econometrics, Cukurova University, Adana, Turkey

Abstract

Ordinary Least Squares (OLS) is a well-known estimator to estimate parameters of a linear regression model. In case of an exact prior information about model parameters, Restricted Least Squares (RLS) will improve OLS estimator in terms of standard errors. However, Mixed Estimator (ME) outperforms RLS when the prior information is stochastic.

The recent developments in computers, information systems, and data science methods increased the importance and impact of Big Data analysis in different areas. A problem in Big Data analysis is to determine the important variables among a large set of independent ones to predict a dependent variable. In a regression setup, this problem can be solved using penalized estimators like Least Absolute Shrinkage and Selection Operator (Lasso), Elastic Net, and Bridge. These estimators can be used for model selection and estimation simultaneously.

In the recent literature, Lasso and Bridge estimators are extended to include linear restrictions to incorporate exact prior information about model parameters. The proposed estimators are called Restricted Lasso (R-Lasso) and Restricted Bridge (R-Bridge) and they can be used to select a model while incorporating linear restrictions. However, to the authors' knowledge, no such studies have addressed implying stochastic restrictions with penalized estimators.

In this paper, we propose a Mixed Lasso (M-Lasso) estimator to incorporate stochastic linear restrictions within a Lasso framework. Our proposed estimator can be used for model selection and estimation simultaneously while implying stochastic linear restrictions to improve estimates. We compare M-Lasso with R-Lasso, Lasso, ME, RLS, and OLS with a simulation study according to mean squared error and model selection performance. We also employ aforementioned estimators to estimate a consumption function analyzed in the literature for an application on a real dataset.

Key Words: Stochastic restrictions, Lasso, mixed estimator, model selection.

On the Limiting Behaviour of the Ergodic Distribution of Random Walk with a Generalized Reflecting Barrier

Başak Gever¹, Zülfiye Hanalioğlu²

¹Department of Industrial Engineering, Turkish Aeronautical University, Ankara, Turkey

²Department of Actuary and Risk Management, Karabuk University, Karabuk, Turkey

Abstract

In this study, a random walk processes with generalized reflecting barrier ($X(t)$) is considered. After proving its ergodicity under some weak conditions, the explicit form of the ergodic distribution is obtained. It is proved that the limit form of the ergodic distribution of the standardized process is converges to $R(x)$ as follows:

$$Q_X(\lambda x) \equiv \lim_{t \rightarrow \infty} P\{X(t) \leq \lambda x\} \xrightarrow{\lambda \rightarrow \infty} R(x) \equiv \frac{2}{\mu_2} \int_0^x \left\{ \int_z^\infty (1 - F_+(v)) dv \right\} dz$$

Here, $F_+(x)$ is the distribution function of the first ladder height (χ_1^+) generated by $S_n = \sum_{i=1}^n \eta_i$ and $\mu_2 = E(\chi_1^{+2})$. Moreover, $\{\eta_i, i = 1, 2, \dots\}$ random sequence represents the jumps of the process $X(t)$.

At the end of the study, in order to evaluate the asymptotic rate of the weak convergence, the following equality is obtained, when λ is large enough:

$$|Q_Y(x) - R(x)| \leq \frac{2\mu_1 m_1 (1 - \pi_+(x)) + m_2 (1 - F_+(x))}{\lambda m_1 \mu_2}$$

Here, $\pi_+(x) = \frac{1}{\mu_1} \int_0^x (1 - F_+(v)) dv$; $F_+(x) \equiv P\{\chi_1^+ \leq x\}$; $\mu_k = E(\chi_1^{+k})$; $m_k = E(\eta_1^k)$, $k = 1, 2$.

Key Words: Random walk, reflecting barrier, weak convergence, asymptotic rate.

Forecasting the Volatility of Stock Price Index

Eda Yalçın Kayacan ¹

¹Econometrics, Pamukkale University, Denizli, Turkey

Abstract

It has always been attractive to forecast models that provide the most realistic predictions of financial time series. In addition to many econometric methods to forecast time series, which are generally nonlinear, non-stationary and chaotic, methods based on deep learning models that take into account the aforementioned characteristics of the series are also preferred.

The main objective of this study is to investigate the differences between forecasting successes by predicting the BIST 100 stock market index by using an econometric model such as Generalized Autoregressive Conditional Heteroscedasticity Model (GARCH) and deep learning methods such as Long Short-Term Memory Networks (LSTM) and Recurrent Neural Networks (RNN). The study basically seeks to answer the question of whether there is a difference between deep learning and econometric estimation methods. Therefore, it is proposed hybrid models such as LSTM+GARCH and RNN+ GARCH for stock price volatility.

For the purposes specified in the study, it was used the data of stock closed prices of Borsa Istanbul-100, which period is from 01.01.2000 to 01.07.2019. Then, it was evaluated models that provide the best volatility forecasting using the models such as LSTM, RNN, GARCH, LSTM + GARCH and RNN + GARCH.

Key Words: Volatility, recurrent neural networks, long short-term memory networks, generalized autoregressive conditional heteroscedasticity model, hybrid model.

¹Econometrics, Pamukkale University, Denizli, Turkey.

Using Data Mining Methods for Prediction of Women's Birth Time in Turkey

Özer Bakar¹, Duygu İçen²,

¹Department of Actuarial Sciences, Hacettepe University, Ankara, Turkey

²Department of Statistics, Hacettepe University, Ankara, Turkey

Abstract

Data mining methods have become popular in recent years with the help of developing technology, ease of use, no need for any distribution information about the data, and reliability and easy interpretation of the results. In this study, data obtained from Turkey Demographic and Health Survey which conducted in 2008 that includes variables such as year of birth, marital status, year of marriage, education, child place and region, mother tongue, birth cohorts, age at marriage and age at first birth is used for classification in the consideration of data mining methods. The birth times of the women that means how many months after/before the first marriage given birth were grouped as pre-marriage, first year of marriage, between first and third years of marriage and third year after marriage. Afterwards, the grouped birth times were classified and predicted with the methods of; Decision Trees, Bayesian Networks, Support Vector Machines, K Nearest Neighbor, Artificial Neural Networks and Logistic Regression models using the SPSS Modeler 18.2. According to the success criteria of the methods Support Vector Machines and Logistic Regression were found the best estimator with the result of %85.7. The most effective variables were found as the age at birth and the age at first marriage.

Key Words: Data mining, classification, birth time, TDHS.

An Improved Search Engine for E-Commerce Websites Based on Word2vec Model

Bahadır Durak¹, Buse Dedeakayoğulları²

¹ Research & Development Center, OBASE, Istanbul, Turkey

² Research & Development Center, OBASE, Istanbul, Turkey

Abstract

Searching is the fundamental way to find any kind of information on the web; similarly it is a crucial functionality of e-commerce websites. High variety of products on e-commerce websites makes it difficult for customers to reach the desired product. Therefore, search engines in these websites have become an important factor affecting customer satisfaction. Traditional search methods generally use the exact word as it is to find the products and they are not efficient to find the other products with similar words. Moreover, most of the search methods, which consider similar words, capture text similarity only and ignore semantic relationship between words. Although there are many recent studies dealing with this issue, most of the search methods have been developed for English. Relatively little research has been undertaken in developing search methods to analyse and evaluate word similarity for Turkish.

In order to include contextual information provided by search words, a semantic similarity method based on word2vec is used in this study. Word2vec is one of the most popular techniques to establish shallow neural networks that use texts as inputs and transform them to vectors in the high dimensional vector space. By this method, we create a model in which the words are represented by vectors that allow us to measure their similarity. We trained word2vec model on a large set of products from various e-commerce websites. Moreover, Turkish morphological database is used to obtain the roots of words for more efficient model training. Calculated similarities between words are used to find the products by considering search words and product definitions together. In other words, word2vec model finds the similar products and defines how similar they are by evaluating the search words. Therefore, the search engine, which uses our search method, is also able to order the products before presenting them to the user.

A test set, which contains searched products and search words from an e-commerce website, is used to test the model and success rates are reported. These rates are compared with the success rates obtained from current traditional search engine. The results show that the proposed model based on word2vec improves the performance of the search engine significantly.

Key Words: Data analysis and modeling, data mining, data science, text processing, text analysis, word representation, word2vec, natural language processing, neural networks.

Comparing Forecasting Accuracies of ATA and Exponential Smoothing

Beyza Cetin¹, Idil Yavuz²

¹Statistics, Dokuz Eylul University, Izmir, Turkey

²Statistics, Dokuz Eylul University, Izmir, Turkey

Abstract

Forecasting is a crucial step in almost all scientific research and is essential in many areas of industrial, commercial, clinical and economic activity. There are many forecasting methods in the literature but exponential smoothing stands out due to its simplicity and accuracy. Simple exponential smoothing is used for data sets, which randomly fluctuate around a level, and Holt's linear trend method is a valuable extension of simple exponential smoothing that helps deal with linearly trended data. Despite the facts that exponential smoothing methods are widely used and have been in the literature for a long time, they suffer from some problems that potentially affect the models' forecasting accuracies. The proposed ATA model has emerged to overcome these problems and to provide accurate forecasts. In this study, the forecasting accuracies of ATA and exponential smoothing will be compared for both trended and un-trended data sets. The results given in this study are obtained using simulated data sets with different sample sizes and variances and forecasting errors for both short and long term forecasting horizons are compared. The results show that the proposed approach outperforms exponential smoothing for both types of time series data for both long and short term forecasting.

Key Words: ATA method, exponential smoothing, forecasting, simulation, time series

A Study on Industry 4.0 Awareness of Undergraduate Students by Using Structural Equation Modelling

Sevde Ceren Yıldız¹, Seniye Ümit Fırat²

¹Faculty of Engineering Department of Industrial Engineering, Doğuş University, İstanbul, Turkey

²Faculty of Engineering Department of Industrial Engineering, Marmara University İstanbul Turkey

Abstract

The fourth industrial revolution, Industry 4.0, appears in variety of fields such as manufacturing, service sector and daily life by its exponentially growing structure in a global scale. Industry 4.0 is constructing a real-time interaction between human beings, machines and products to develop a customized and digitalized smart manufacturing model with high elasticity with respect to customer demands. In local scale, although Turkey tries to keep up with the rapid development of Industry 4.0, because of the technological inadequacy and innovation, it falls behind of the world's current standards and development phase. According to Global Innovation Index 2018, Turkey placed at 50th place by decreasing seven ranks compared to 2017. Also another rank falling of Turkey is seen in the Global Competitiveness Index 2018 edition. Turkey was ranked 61st positioning among 140 countries, while 58th in 2017 edition of same index.

The aim of this study is to determine and evaluate Industry 4.0 awareness of younger generation (undergraduate students) in Turkey while Turkey losing ranks worldwide in a variety of indexes like technological competence and innovation. For this purpose, at the first stage of the research study a theoretical awareness model was established based on the literature review to form the basis of the survey. Then at the second stage, we have conducted a survey study for Turkish university students, to measure their perception and attitude of Industry 4.0. The main part of the applied questionnaire was a Likert type scale that based on Industry 4.0 awareness related questions. The items of the measurement scale has included the general frame and principles of Industry 4.0 umbrella based on the important concepts such as cyber physical systems, internet of things, smart factory, internet of services, 3-D printers, smart robots and augmented reality.

At the last stage of the research, the constructed awareness research model was evaluated statistically and determined the awareness level of younger generation. The collected data was analyzed by using Structural Equation Modelling that is a multivariate statistical method and determine the awareness level of younger generation. The findings and results were discussed in details of the Industry 4.0 main frame. At last, as a conclusion, a comparative study was realized on final results of Industry 4.0 awareness survey with Industry 4.0 related global indexes indicators.

Key Words: Industry 4.0, Industry 4.0 awareness, structural equation modelling, global innovation index.

Good Agricultural Practices and Technical Efficiency in Chili Production in Thailand

Wirat Krasachat

Department of International Business Administration, King Mongkut's Institute of Technology Ladkrabang.
Bangkok, Thailand

Abstract

In response to a growing awareness of quality, safety and yield of agricultural production, Thai government initiated the “Good Agricultural Practices” (GAP) program in 2003. This is a voluntary program where farmers may adopt it at will. Despite the government and NGOs’ efforts to assist farmers in adopting and handling the GAP technology, the majority of farmers still use conventional method, especially in chili production. This practice, if it continues, could result in higher healthcare cost for farmers and consumers. In addition, it could affect the opportunities for Thai farmers to export chili and raise their standard of living.

This paper investigates whether adopting GAP technology in producing chili indeed raises technical efficiency, compared to conventional method. Specifically, this paper uses a sample of 100 small farms in Thailand for the 2018 crop year to estimate a propensity score matching - stochastic production frontier. Our model includes four inputs: land, labour, machinery and equipments, and materials and one output: fresh chili. We also control for technology whether it is GAP or conventional farming as well as growing season (i.e. rainy vs dry season). This paper hypothesizes that GAP farms, on average, are more technical efficient than conventional farms.

Key Words: Chilli production, good agricultural practices, sample-selection model, stochastic production frontier, technical efficiency

The Use of Generalized Estimating Equations and Panel Causality Tests in the Evaluation of Countries' Economic Profiles

Harun Yonar¹, Neslihan İyit²

¹ Department of Statistic, Selcuk University, Konya, Turkey

² Department of Statistic, Selcuk University, Konya, Turkey

Abstract

Gross Domestic Product (GDP) which is the most important indicator of economic growth, has been widely used in economic evaluations of countries. Therefore, the relationship between GDP and macroeconomic variables should be evaluated with strong statistical and econometric analyses. However, the explainability of the GDP by using major macroeconomic variables and modeling of the economic profiles of countries is quite complex. The solution of this complexity is necessary for countries to develop correct and sound economic development strategies. In this study, it is aimed to model the countries' economies in the aspect of macroeconomic variables by using Generalized Estimating Equations (GEE), which is an extension of Generalized Linear Models. In addition, panel causality tests have been used to investigate causality between variables in longitudinal data. In the selection of the macroeconomic variables to be used in GEE models, the results obtained from causality tests have been used. The most important finding of this study is that the contribution of causality tests to GEE models is revealed. Moreover, the changes in GDP have been explained by the macroeconomic variables that are the causes of GDP through the GEE models created for each country.

Key Words: Generalized estimating equations, panel causality tests, gross domestic product.

Bi-Responses Mars Modelling through Earth Package for Regression Type Problems

Ecevit Eyduran¹, Meleksen Akin², Demet Canga^{3,*}, Ali Eylem Celik⁴, Sadiye Peral Eyduran⁵

¹Iğdır University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Quantitative Methods, Iğdır, Turkey

²Iğdır University, Faculty of Agriculture, Department of Landscape Architecture, Iğdır, Turkey

³Osmaniye Korkut Ata University, Bahce Vocational School, Osmaniye, Turkey

⁴Iğdır University, Faculty of Agriculture, Department of Animal Science, Biometry-Genetics Unit, Iğdır, Turkey

⁵Iğdır University, Faculty of Agriculture, Department of Horticultural Science, Iğdır, Turkey

Abstract

As an alternative to response surface method (RSM), multivariate adaptive regression splines (MARS) data mining algorithm is a nonparametric regression technique that allows analysts to examine the relationship between sets of responses and predictors for classification and regression type problems. There is still lack of information on bi-responses MARS modeling for regression type problems in literature; therefore, this research was conducted to develop an R script file on bi-responses MARS modeling for regression type problems by using “cbind function” in R free software. For this aim, bi-responses MARS modeling produced two prediction equations simultaneously and the developed R script file permitted to calculate goodness of fit criteria for each continuous response. Earth (enhanced adaptive regression through hinges) package of the R free software was installed and specified for bi-responses MARS modeling. Tuning parameters as number of selected terms (k) and degree of interactions were specified to lower GCV value as a selection criterion for the best predictive model.

The R script file developed on bi-responses MARS modeling for regression type problems will be a valuable source for researchers at various scientific fields in producing prediction equations for continuous responses.

Key Words: MARS, bi-responses, earth package, data mining, CART.

* demetcanga@osmaniye.edu.tr

New Goodness of Fit Tests for Unit-Lindley Regression Model

Deniz Ozonur¹

¹Department of Statistics, Gazi University, Ankara, Turkey

Abstract

In applied sciences, regression models are commonly used to describe associations between explanatory and response variables. Response variables with values restricted to interval (0,1) are common in social sciences, agronomy, psychometrics and other areas. For example, proportion of income spent on food is affected by the level of income and the number of persons in household. Unit-Lindley regression model is useful for situations where the response variable is continuous and restricted to unit interval [1].

Goodness of fit tests examine how well a sample of data agrees with a theory such as a given distribution. In the case of a generalized linear model, Pearson chi-square and deviance test statistics are commonly used in assessing the goodness of fit between the given generalized linear model and the data. Smooth goodness of fit tests are applied to some generalized linear models such as Poisson and Logistic regression models [2]. In this study, smooth goodness of fit tests are proposed to test distributional assumption for unit-Lidley regression model. These tests are then compared with existing tests in terms of type I error and power values by simulation study.

Key Words: Unit-Lindley regression model, smooth test, power of test, type I error.

References

- [1] Mazucheli, J., Menezes, A. F. B., & Chakraborty, S. (2019). On the one parameter unit-Lindley distribution and its associated regression model for proportion data. *Journal of Applied Statistics*, 46(4), 700-714.
- [2] Rippon, P. (2013). Application of Smooth Tests of Goodness of Fit to Generalized Linear Models. Ph.D thesis, Australia, University of Newcastle.

Robust Regression Method for Parameters of Power Lindley Distribution

Berivan akmak¹, Fatma Zehra Doęru²

¹Department of Statistics, Giresun University, Giresun, Turkey

²Department of Statistics, Giresun University, Giresun, Turkey

Abstract

The power Lindley (PL) distribution was proposed by [Ghitany, M.E., Al-Mutairi, D.K., Balakrishnan, N., and Al-Enezi, L.J. 2013. Power Lindley distribution and associated inference. Computational Statistics and Data Analysis, 64, 20-33] to analyze survival data sets. In this study, we consider estimating the shape and scale parameters of the PL distributions using the least squares (LS) estimation method. However, estimation results obtained from LS method are very sensitive to the outliers; therefore, we propose a robust regression estimation method as an alternative to the LS estimation method. To obtain robust estimators, we change the LS criterion with the robust criterion based on Tukey's and Huber's functions. We also provide a simulation study to compare the performance of robust estimators with LS estimators.

Key Words: Least squares estimation, power Lindley distribution, robust estimation.

Use of Bootstrap Aggregating (Bagging) MARS to Improve Predictive Accuracy for Regression Type Problems

Ecevit Eydurhan¹, Demet Canga^{2,*}, Harun Sevgenler³, Ali Eylem Celik³

¹Iğdır University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Quantitative Methods, Iğdır, Turkey

²Osmaniye Korkut Ata University, Bahçe Vocational School, Osmaniye, Turkey

³Iğdır University, Department of Animal Science, Agricultural Faculty, Biometry and Genetics Section, Iğdır, Turkey

Abstract

Bootstrap aggregating (Bagging) is an effectual method specified with several classification and regression methods to improve their predictive correctness at big data sets. Multivariate adaptive regression splines as an extension of recursive partitioning algorithm (CART) is a nonparametric regression technique developed to reveal high dimensional relationship between sets of the response variable(s) and predictors for both classification and regression type problems. In this research, bootstrap aggregating multivariate adaptive regression splines (Bagging MARS) as a hybrid approach obtained by combining the bootstrap aggregating (Bagging) with multivariate adaptive regression splines (MARS) was employed for a sample data set. MARS algorithm is applied for each bootstrap sample and an overall prediction is computed by averaging the results of the possible bootstrap samples identified for regression type problems. For the Bagging MARS algorithm, earth (enhanced adaptive regression through hinges) and caret (classification and regression training) packages of R studio program were used in order to present more powerful solution of regression problems. Predictive performance of the Bagging MARS technique was measured for various number of bootstrap samples i.e. 3, 5, 10, 15, 20, 25 and 50.

Consequently, it was argued that the implementation of the Bagging MARS algorithm may be a significant selection for improving powerful predictive accuracy under the investigation.

Key Words: Bagging, MARS, Bootstrap, CART.

* demetcanga@osmaniye.edu.tr

The Relationship Between the Graph Energies and Vulnerabilities

Erva Kol¹, Mehmet Ali Balcı², Ömer Akgüller³

¹Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

²Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

³Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

Abstract

In mathematics, the energy of a graph is the sum of the absolute values of the eigenvalues of the adjacency matrix of the graph and this quantity is studied in the context of spectral graph theory. In this study, we present the correlations between the several graph energies namely adjacency energy, Laplacian energy, and distance energy; and some vulnerability measures of simple graphs such that edge and vertex connectivity, and solitude number.

Key Words: Graph energies, vulnerability measures, correlation.

¹ ervakol@posta.mu.edu.tr

² mehmetalibalci@mu.edu.tr

³ oakguller@mu.edu.tr

On Solving the Generalized Assignment Problem via Hypergraphs

Kezban Afetsu Başdoğan¹, Mehmet Ali Balcı², Ömer Akgüller³

¹Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

²Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

³Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

Abstract

The goal of the well-known Generalized Assignment Problem is to assign tasks to agents such that the capacity of the agent does not exceed its limits as it minimizes the total cost. In this present study, we use hypergraphs to solve generalized assignment problem by looking for a cost minimizing solution to tasks assignment to the agents which are the individual hyperedges. Hence, we first determine the tasks as hyperedges, then obtain the vertex cover sets of the simple graph representation of the modelling hypergraph. We also apply presented methods to two different data sets.

Key Words: General assignment problem, hypergraphs, vertex cover set.

¹ afetsukezbanbasdogan@posta.mu.edu.tr

² mehmetalibalci@mu.edu.tr

³ oakguller@mu.edu.tr

The Cubic Rank Transmutation Method and Applications

Demet Sezer¹, Nuri Celik², Aydın Karakoca³

¹Department of Actuarial, Selcuk University, Konya, Turkey

²Department of Mathematics, Gebze Technical University, Gebze/Kocaeli, Turkey

³Department of Statistics, Necmettin Erbakan University, Konya, Turkey

Abstract

Well-known distributions may not be flexible enough when dealing with real life data sets. Therefore, the generalizations of distributions are one of the most prevalent method in statistical methodology to get more fitting models. In this study, some distributions used in the statistical applications are discussed by using cubic rank transmuted methodology proposed by Granzatto (2017) which allows to fit complex data sets with bimodal hazard rates. The characteristic properties of new distributions such as probability density function, distribution function, expected value and variance are examined. The parameters of the new distributions are estimated by using maximum likelihood method. In addition, real data sets taken from actuarial, insurance and finance fields will be modelled with new distributions and compared with other alternatives used in the literature.

Key Words: Cubic rank transmutation, statistical distributions, hazard rate, moments.

Hierarchies in Communities of FTSE100 from the Perspective of Brexit

Serdar Can Güzel¹, Mehmet Ali Balcı², Ömer Akgüller³

¹ Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

² Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

³ Department of Mathematics, Muğla Sıtkı Koçman University, Muğla, Turkey

Abstract

Nowadays, increase of the analysing stock markets as complex systems lead graph theory to play key role. For instance detecting graph communities is an important task in the analysis of stocks, and minimum spanning trees let us to get important information for the topology of the market. In this study, we use cross correlations in the daily closing prices of the London stock exchange FTSE 100 index companies from July 2010 to March 2018 in order to obtain minimum spanning tree structures. The aim of this paper is to study the hierarchical evolution of the market communities throughout the Brexit referendum. Hence, the overall sample is divided into two sub-periods of equal time windows of pre-referendum, and post-referendum to obtain communities and hierarchical structures. The findings show that there are significant changes in the hierarchical structures of communities throughout the Brexit referendum.

Key Words: Financial networks, network communities, cross correlation, minimum spanning tree.

¹ serdarcanguzel@posta.mu.edu.tr

² mehmetalibalci@mu.edu.tr

³ oakguller@mu.edu.tr



A Comparison of Regression Analysis Methods for Beer Consumption

Duygu Dede Şener¹

¹ Başkent University Department of Computer Engineering, Ankara, Türkiye

Abstract

Regression analysis is an important statistical technique used to investigate the information contained in the data. It has an increasing use in a variety of fields such as time series data modelling, forecasting, data mining and finance. In this study, an experimental analysis was conducted on a real dataset from Kaggle data repository to predict beer consumption of people lives in San Paulo, Brazil. To the best of our knowledge, there is no study which perform regression analysis on this dataset. The analysis was performed with using different types of regression techniques such as Lineer Regression (LR), k-Nearest Neighbours Regression (k-NNR), Support Vector Regression (SVR), Random Forests and Locally Weighted Random Forest (LWRF) to choose the best suited model for prediction of the beer consumption. To assess the prediction accuracy of the used methods, correlation coefficient (CC), root mean squared error (RMSE) and mean absolute error (MAE) were calculated between the observed and predicted values in terms of measurements taken with period of one year. According to the experimental results, LR has slightly better performance in prediction beer consumption with CC, MAE and RMSE with values 0.85, 1.98 and 2.34 respectively. SVR has close performance to the LR method. Besides this, LWRF has the worst precision performance among other methods. In order to assess whether differences between performances of regression techniques were statistically significant, a statistical test named Paired t-test was applied. Thus, it was observed that the results of LR method are statistically better than other methods.

Key Words: Regression, data mining, linear regression, k-nearest neighbours regression, support vector regression, random forest, locally weighted random forest.

Modified Two Parameter Estimator to Combat Multicollinearity: A Comparative Study

Selahattin Kaçıranlar¹, Nimet Özbay², Ecem Özkan³

¹ Department of Statistics, Çukurova University, Adana, Turkey

² Department of Statistics, Çukurova University, Adana, Turkey

³ Department of Statistics, Çukurova University, Adana, Turkey

Abstract

Typical chemical data tend to be characterized by many independent variables on relatively fewer observations. In addition, there is a high degree of near collinearity among the explanatory variables. It is common knowledge that under these conditions, ordinary least squares (OLS) estimations of regression coefficients may be very unstable, leading to very poor prediction accuracy. The purpose of this paper is to combine principal components regression (PCR) and two parameter estimation. An alternative procedure that combines both PCR and Liu estimation is also introduced. Furthermore, a Monte Carlo simulation experiment and real-life application are presented by using different estimators of the biasing parameters to compare the performance of the combination of two parameter estimation and PCR and some other existing methods. The results of both simulation study and real-life application show that the proposed estimators outperform other competing estimators depending on parameters.

Key Words: Multicollinearity, principal component regression, ridge regression, two parameter estimator.

The Government Revenue - Expenditure Nexus: Symmetric and Asymmetric Causality Tests

Şule Nazlı Arslan¹, Havvanur Feyza Erdem²

¹Department of Econometrics, Karadeniz Technical University, Trabzon, Turkey

²Department of Econometrics, Karadeniz Technical University, Trabzon, Turkey

Abstract

In the literature, the relationships between government revenues and expenditures are explained by four hypotheses: the Tax-and-Spend Hypothesis, the Spend-and-Tax Hypothesis, the Fiscal Synchronization Hypothesis, and the Institutional Separation hypothesis. The purpose of this study is to examine the relationship between government revenues and spending for Turkish economy. In this study, the data cover the period of 2006:M1-2019M3. In this study, symmetric causality relationships between variables were analyzed by using VAR/Granger test. However, the asymmetric causality relationships were tested by using Asymmetric Causality Test developed by Hatemi-J (2012). The empirical findings of this study indicate that both the Fiscal Synchronization Hypothesis is valid for the relationships between total expenditures and total revenues. However, when the relationship between total expenditures and tax revenues is evaluated, it is seen that the Tax-and-Spend Hypothesis is valid.

Key Words: Expenditure, revenue, asymmetric causality, Hatemi J.

¹sulenazlii.arslan@gmail.com, Karadeniz Technical University, S.B.E., Department of Econometrics, Trabzon/Turkey

²havvanurerdem@ktu.edu.tr, Karadeniz Technical University, İ.İ.B.F., Department of Econometrics, Trabzon/Turkey

The Validity of Easterlin Hypothesis in Turkish Economy

Ayşenur Üçeş¹, Havvanur Feyza Erdem²

¹Department of Econometrics, Karadeniz Technical University, Trabzon, Turkey

²Department of Econometrics, Karadeniz Technical University, Trabzon, Turkey

Abstract

According to Easterlin (1980), there is a significant statistically the relationship between cohort size and fertility rates. Easterlin (1980) implies that relatively large cohorts will exhibit lower fertility rates or relatively small cohorts will exhibit larger fertility rates. The purpose of this paper is to investigate the Easterlin hypothesis validity in Turkey. For this aim, the data cover 1991-2017 (annually). In this study, we first calculate twelve different relative cohort size for women, men, and total population, separately. Then we estimate the effect of relative cohort size and some control variables on the fertility rate by using stepwise analysis. The empirical findings of this study indicate that Easterlin hypothesis is not valid in Turkey.

Key Words: Fertility rate, relative cohort size, Easterlin hypothesis, stepwise regression analysis.

¹aysenur.uces@gmail.com,, Karadeniz Technical University, S.B.E., Department of Econometrics, Trabzon/Turkey

²havvanurfeyza@ktu.edu.tr, Karadeniz Technical University, İ.İ.B.F., Department of Econometrics, Trabzon/Turkey

A Modeling Study for Discrete and Continuous Variables by Using Parametric and Nonparametric Regression Analysis

Gözde Karakoç¹, Özlem Türkşen²

¹Ankara University, Faculty of Science, Mathematics Department, Ankara, Turkey

²Ankara University, Faculty of Science, Statistics Department, Ankara, Turkey

Abstract

Many of the real life problems have discrete and continuous data together. These data sets are called mixed data in the literature. The main step for analysis of mixed data set can be considered as modeling stage. In order to model the mixed data set parametric and nonparametric modeling methods can be used. For application of parametric modelling, some assumptions should be satisfied. However, in many cases it seems not possible. When parametric modeling can not be applied, nonparametric modeling methods should be preferred.

There have been several nonparametric modeling methods in the literature, e.g. k-closest neighborhood, decision trees, support vector machines (SVM) and artificial neural networks (ANN), Data Envelopment Analysis (DEA) and Nonparametric Regression Analysis. One of the most commonly used nonparametric modeling method is nonparametric regression analysis. The nonparametric regression method does not make strong modelling assumptions and does not depend on a fixed parametric model.

In this study, parametric and nonparametric regression analysis were used to model the mixed data set. For application purpose, a mixed data set was chosen from chemical engineering field and the obtained results were discussed.

Key Words: Mixed data, parametric regression analysis, nonparametric regression analysis.

Alignment Method: Concept and Application

Gizem Uyumaz¹, Gözde Sırgancı², Alperen Yandı³

¹Measurement and Evaluation in Education, Giresun University, Giresun, Turkey

²Measurement and Evaluation in Education, Bozok University, Yozgat, Turkey

³Measurement and Evaluation in Education, Abant İzzet Baysal University, Bolu, Turkey

Abstract

In studies comparing different groups using a measurement tool, measurement invariance between the groups should be examined. Measurement invariance is a validity problem and validity is an important issue that should be examined in every study. If measurement invariance is not provided, comparisons cannot be made across groups. Multiple group confirmatory factor analysis (MG-CFA) is one of the methods commonly used in the measurement invariance. However, MG-CFA fails when too many modifications are required. The existence of very large modification indexes indicates that a long model modification sequence is needed to achieve an acceptable model. The purpose of this study is to explain the basic concepts and processes of alignment method for MG-CFA presented by Asparouhov and Muthén (2014). The strength of this method is the ability to predict models appropriately for many groups. There are two different Bayesian Alignment estimation methods: Bayesian structural equation modeling (BSEM) and configural methods. The advantage of BSEM alignment estimation is that it can be used to solve estimation problems in groups where the number of people in the data set is small or where an estimation problem related to the data arises. While the method allows approximate measurement invariance, it compares the factor means and variances across the groups and provides information about the degree of measurement invariance. A detailed analysis can be performed to determine which measurement parameters are approximately invariant and which are not. In this study, there is an example for practitioners the determination of the measurement equivalence between Turkey and other countries. For this purpose, ST94 questionnaire data which is one of the psychological measurement tools used in PISA 2015 was used. Analyses were made in Mplus. The necessary procedures for measurement invariance are performed for Turkey and 56 other countries. Configural invariance is achieved (CFI = ,998, TLI = ,996). In the metric invariance stage, where the equivalence is impaired, modification indexes ranging from 62,910 to 4,633 were obtained. Alignment method was used to determine the equivalents for the 57 countries in which comparison was made considering the disadvantages of performing all modification indexes.

Key Words: Validity, measurement invariance, modification indexes.

Different Methods for Testing of Measurement Invariance with Many Groups

Gözde Sırgancı¹, Gizem Uyumaz², Alperen Yandı³

¹ Measurement and Evaluation in Education, Bozok University, Yozgat, Turkey

² Measurement and Evaluation in Education, Giresun University, Giresun, Turkey

³ Measurement and Evaluation in Education, Abant İzzet Baysal University, Bolu, Turkey

Abstract

Due to accessibility to international survey data, cross-cultural studies establish comparisons across a large number of groups. Measurement noninvariance suggests that the construct that measured means different in some study groups, and so the construct cannot be comparable across cultural groups. Thus, Measurement invariance (MI) is important testing to provide validity of these kind of studies. MI is testing with some traditional methods which are called exact MI methods based on the factor analytic and item response theory approaches. However, these methods have some methodological drawbacks while studying over many groups. These methods generally shows up in too many large modification indexes. And modification indexes approach does not guarantee to find the best fit model. Thus, MI generally fails over configural invariance model. Recent studies shows that rather than exact MI approaches approximate invariance approaches are more practical for testing MI with many groups. The aim of this study is to compare four methods for MI testing across large groups (multiple group confirmatory factor analysis, multilevel confirmatory factor analysis, Bayesian approximate MI testing, and alignment optimization) and identifying strengths and weaknesses of these approaches in terms of their conceptual framework and statistical procedures. The data used in this study were from the student questionnaire, code ST113 of Programme for International Student Assessment (PISA) 2015. There are 384,734 subjects in 57 countries. This study took place three parts. First ST113 questionnaire data was testing MI with four methods. Second, Monte Carlo study was conducted to simulation based on real data population parameter values that is obtain from first analysis. In final stage, simulated data sets that generated for various conditions were analysed with four methods. All analyses were conducted using Mplus version 7.1. Results are submitted for four different group sample size (N=100, N=200, N=500, N=2000). The efficiency of the 4 methods are investigated for various fit criteria. Overall, across four methods showed reasonable performance in identifying the level of invariance. While MG-CFA and ML-CFA evaluate exact invariance, The Bayesian MI and alignment optimization evaluate approximate invariance. When exact invariance methods compare with each other, MG-CFA failed metric invariance but ML-CFA provided it. On the other hand ML-CFA is sensitive as depends on number of groups. The Bayesian MI method figured out the issues of poor model fit under the strict exact invariance assumption. Alignment optimization has been suggested mainly for factor mean comparisons under approximate MI. Alignment is not recommended when many measurement parameters are noninvariant. Finally, all methods have some advantages and disadvantages under some conditions.

Key Words: Approximate measurement invariance, exact measurement invariance Bayesian, multilevel CFA, alignment optimization.

Institutional Quality and Income Inequality: Panel Cointegration Analysis in 12 Developing Countries

Seyit Ali Miçooğulları¹, Süleyman Değirmen²

¹ Department of Foreign Trade, Kilis 7 Aralık University, Kilis-Turkey

² Department of Economics, Mersin University, Mersin-Turkey

Abstract

Income inequality is regarded as one of the most important economic issues of our time. Although there is no consensus in the literature, economic, social and political factors as well as institutional factors have an impact on income inequality. In this respect, this study examines the relationship between income inequality and institutional quality of 12 developing countries. We employ panel unit root, cointegration and causality analyses for annual panel data from 2010 to 2015. In the long run, a dynamic relationship between variables is investigated by using Panel DOLS and Panel FMOLS coefficients.

The results of both Panel DOLS and FMOLS show that the increase in institutional quality of the 12 developing countries is an effective factor on income inequality. The coefficients obtained by both methods are very close to each other. The results of the study show that the 1-unit increase in institutional quality reduces the income inequality by 0.064% in the long-term. Finally, the short and long-term causality relationship between the variables was estimated by the vector error correction model (VECM) and the results showed that there is a significant causality in terms of institutional quality income inequality.

Key Words: Income inequality, institutional quality, cointegration analysis, vector error correction model, developing countries.

Examine of Mathematical Anxiety of Engineering Students by Factor Analysis

Şüheda Güray ¹, Tülay Korkmaz Devrani²

¹Faculty of Education, Başkent University Ankara, Turkey

²Faculty of Economics and Administrative Sciences, Başkent University Ankara, Turkey

Abstract

In this study, it was aimed to evaluate mathematical anxiety of Engineering Faculty students. In order to see the mathematical anxiety of the students we used Mathematical Anxiety Scale which is developed by Erol (1989) and the scale was adapted to our country with the last regulation and named as Mathematical Anxiety Scale (MANX). For this purpose of the study, Mathematical Anxiety Scale questionnaire was applied on a random sample taken from the students of the Faculty of Engineering at Başkent University and winter semester of 2018-2019. Consequently, we examined the results obtained from the sample with explanatory factor analysis and determined the sub-dimensions of the anxiety scale.

As it is very well known, factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. In other words, factor analysis searches for joint variations in response to unobserved [latent variables](#). The observed variables are modelled as [linear combinations](#) of the potential factors and [error](#) terms. Factor analysis aims to find independent latent variables. After factor analysis, the scale consisting of 45 questions was reduced to 8 sub-factors and these factors were redefined. On the other hand, we calculate the Cronbach's alpha coefficient which is measure the reliability/quality of the data is 0.91.

Key Words: Mathematical anxiety scale, factor analysis.

References

- 1) Baloğlu, M. (2004). Üniversite Öğrencilerinin Matematik Kaygı Düzeyleri Açısından Karşılaştırılması. XIII. Ulusal Eğitim Bilimleri Kurultayı, 6-9 Temmuz, İnönü Üniversitesi, Eğitim Fakültesi, Malatya.
- 2) Bekdemir, M. (2007). İlköğretim Matematik Öğretmen Adaylarındaki Matematik Kaygısının Nedenleri ve Azaltılması İçin Öneriler (Erzincan eğitim fakültesi örneği). . Erzincan Eğitim Fakültesi Dergisi, 9(2), 131-144
- 3) Bekdemir, Işık ve Çıkılı (2004) “ Matematik Kaygısını Oluşturan ve Artıran Öğretmen Davranış ve Çözümleri” Eurasian Journal Of educational Research. (16).88-94
- 4) E. Erkin, G. Dönmez, & S. Özel, Matematik kaygısı ölçeği'nin Psikometrik Özellikleri. Eğitim ve Bilim, 31(140), 26-33 (2006).
- 5) E. Erol, . Prevalence and correlates of math anxiety in Turkish high school students. Yüksek Lisans Tezi, Boğaziçi Üniversitesi, İstanbul.. (1989)

The Impact of Perceptions of Social Media Advertisements on Advertising Value and Brand Awareness: Research on Generation Y Instagram Users

İbrahim Halil Efendioğlu¹, Yakup Durmaz²

¹Department of Informatics, Gaziantep, Turkey

²Department of Business Administration, Gaziantep, Turkey

Abstract

With a rapid increase in the number of social media users over the last ten years, a new advertising domain has become available for companies. Brands that manage social media well in their advertising strategies can quickly influence consumer decision making and create awareness. However, in social media advertising, which is different from traditional advertising, the right kind of content should be produced and this content should be perceived correctly by consumers. On that basis, the main problem posited by the research is: “Does the perception of Instagram advertisements influence the advertising value and brand awareness from the point of view of Generation Y?” When all these are combined, social media advertisements that are correctly perceived and create awareness will be successful and effective. In this context, the departure point of this study is the effect of the perception of social media advertisements on advertising value and brand awareness. The study proposes a new conceptual model that combines the brand awareness and the advertising value model from Ducoffe’s (1995) Uses and Gratification Theory To achieve this, it is necessary to build rapport with consumers and to present correctly what they wish to see in advertisements by creating awareness. In view of the increasing importance of social media advertising, the study examines how consumer perceptions of social media advertisements affect advertising value and brand awareness. This study was conducted with Generation Y consumers on the basis of their Instagram habits, a popular social media app. For this purpose, face-to-face interviews were held with 665 participants from Generation Y who use Instagram. The collected data was analyzed with AMOS and SPSS statistical package programs using structural equation modelling. According to the analysis results, Y-generation’s perceptions of Instagram advertisements have both a positive and negative impact on advertising value, brand awareness and brand associations.

Key Words: Social media advertisements, consumer perceptions, advertising value, brand awareness.

¹ Gaziantep University

² Hasan Kalyoncu University

Data Dimensionality Reduction based on Variables Clustering

Arnaldo Rabello de Aguiar Vallim-Filho¹, Pedro Okidoi², Leandro Augusto da Silva³, Marco Vinicius Bhering de Aguiar Vallim⁴, Leilton Santos da Silva⁵

¹Mackenzie Presbyterian University, Computer Science Dept., Sao Paulo, Brazil

²Mackenzie Presbyterian University, Computer Science Dept., Sao Paulo, Brazil

³Mackenzie Presbyterian University, Computer Science Dept., Sao Paulo, Brazil

⁴Mackenzie Presbyterian University, Computer Science Dept., Sao Paulo, Brazil

⁵EMAE - Metropolitan Company of Water & Energy, Sao Paulo, Brazil

Abstract

One of the important aspects in data analysis is the question of the number of attributes (or variables) to study, i.e. the dimensionality of the dataset. This paper deals precisely with this issue, aiming to analyse a dataset, involving a large number of variables of an industrial operation, used to predict occurrences of equipment shutdown due to mechanical and / or electrical failure. Predictive models based on artificial neural networks predict these stops. As the number of variables is large, we seek to verify the possibility of grouping these variables involved in the problem. Thus, there is a clustering of variables, not clustering observations (instances), as usual. Having homogeneous clusters, the set of variables of each cluster could be represented by its centroid, thus reducing dimensionality. The paper proposes the determination of these centroids based on multiple regression models. Another important aspect in data analysis is the question of data transformation, in order to make it in adequate conditions for the analysis development. One such transformation technique is the normalization of the data mass, so that all variables under study are represented on the same scale. Normalized data minimizes effects of data units, especially when you have variables measured in very distinct units, such as an electrical variable measured in volts and a mass variable measured in grams. Testing multiple clustering techniques from different data normalization techniques is, therefore, another objective of this paper. The artificial neural network dedicated to equipment downtime prediction, based on the set of variables, is applied to the original and reduced datasets for comparison of results. These tests are developed for all alternatives under analysis. The article describes this methodology, the development of the experiments and the interesting findings obtained in the study.

Key Words: Data science, dimensionality reduction, clustering analysis, artificial intelligence, big data.

Neutrosophic Linear Regression Analysis of Construction Accidents in Turkey

Şüra Toptancı^{1*}, Nihal Erginel², İlgin Acar³

¹Department of Industrial Engineering, Eskisehir Technical University, Eskisehir, Turkey

²Department of Industrial Engineering, Eskisehir Technical University, Eskisehir, Turkey

³Department of Industrial Engineering, Eskisehir Technical University, Eskisehir, Turkey

Abstract

Millions of people in the world get injured or die every year as a result of occupational accidents. Besides, occupational accidents have a great impact on the economy since they also cause huge losses regarding production, productivity, sustainability and finance. In recent years, the occupational accident rates in construction companies have visibly increased with the expansion of the construction industry in Turkey. The construction sector, which contributes to the economic growth of developing countries with employment it provides and added value to the economy, is one of the most hazardous sectors due to existence of different construction projects, working conditions and high turnover in this industry. In order to understand the nature of occupational accidents and to decrease the number of accidents in the sector, the accident data analysis should be done for identifying and quantifying a set of variable that associated with the severity of construction accidents. Linear regression model as one of statistical methods is used to express the relationship on two or more related variables. However, ordinary linear regression model has some limitations under the conditions of inadequate observations, vagueness in the relationship between independent and dependent variables, difficulty to verify distribution assumption of errors and involved human judgements. To overcome these limitations, alternative methods for ordinary regression analysis have been developed in literature.

In this study, an occupational accident data of a construction company reported in 2017 are analysed. The main purpose of this study is to investigate the relationship between accident severity and risk factors related to occupational accidents under uncertainty environment by using Neutrosophic Linear Regression Analysis (NLRA) that consists of neutrosophic sets and multiple linear regression analysis. When the relationship between accident severity and risk factors of the accidents can be identified from analyzing result of NLRA, it is easy to take precautions against accidents and to reduce direct and indirect economic losses related to occupational accidents for the construction sectors. Therefore, construction firms or other sectors may benefit from this approach used in this study to improve their occupational safety management plans and programs.

Key Words: Uncertainty, neutrosophic sets, neutrosophic regression analysis, construction industry, occupational accidents.

¹ This work has been supported as part of the 1705F427 numbered Project approved by Eskisehir Technical University Scientific Research Projects Commission.

² This work is produced from the Ph.D. dissertation prepared by the first author.

³Corresponding author: sura_t@eskisehir.edu.tr

An Application of Cluster Analysis Based on Competitive Intensity and Market Turbulence of Firm

Sinem Peker¹

¹The Department of Mathematics, Yaşar University, Izmir, Turkey

Abstract

In this study, a survey was applied to the senior employees of the leading firms of Turkey and three groups were determined by twostep clustering method based on the competitive intensity and market turbulence of the firms. Considering the leading firms of Turkey, it was found that the average of the monetary resources allocated to the marketing department of the firms, which have high competition intensity and high market turbulence, is higher than the average of the firms whose market turbulence is low but the competitive intensity is a little high. In addition, it has been determined that the average of term orientation of firms, which have high competitive intensity and high market turbulence, is higher than the average of firms with low market turbulence but a little high competitive intensity.

Key Words: Clustering, TwoStep clustering, competitive intensity, market turbulence.

The Huber Functions and Their Application to a Classification Problem

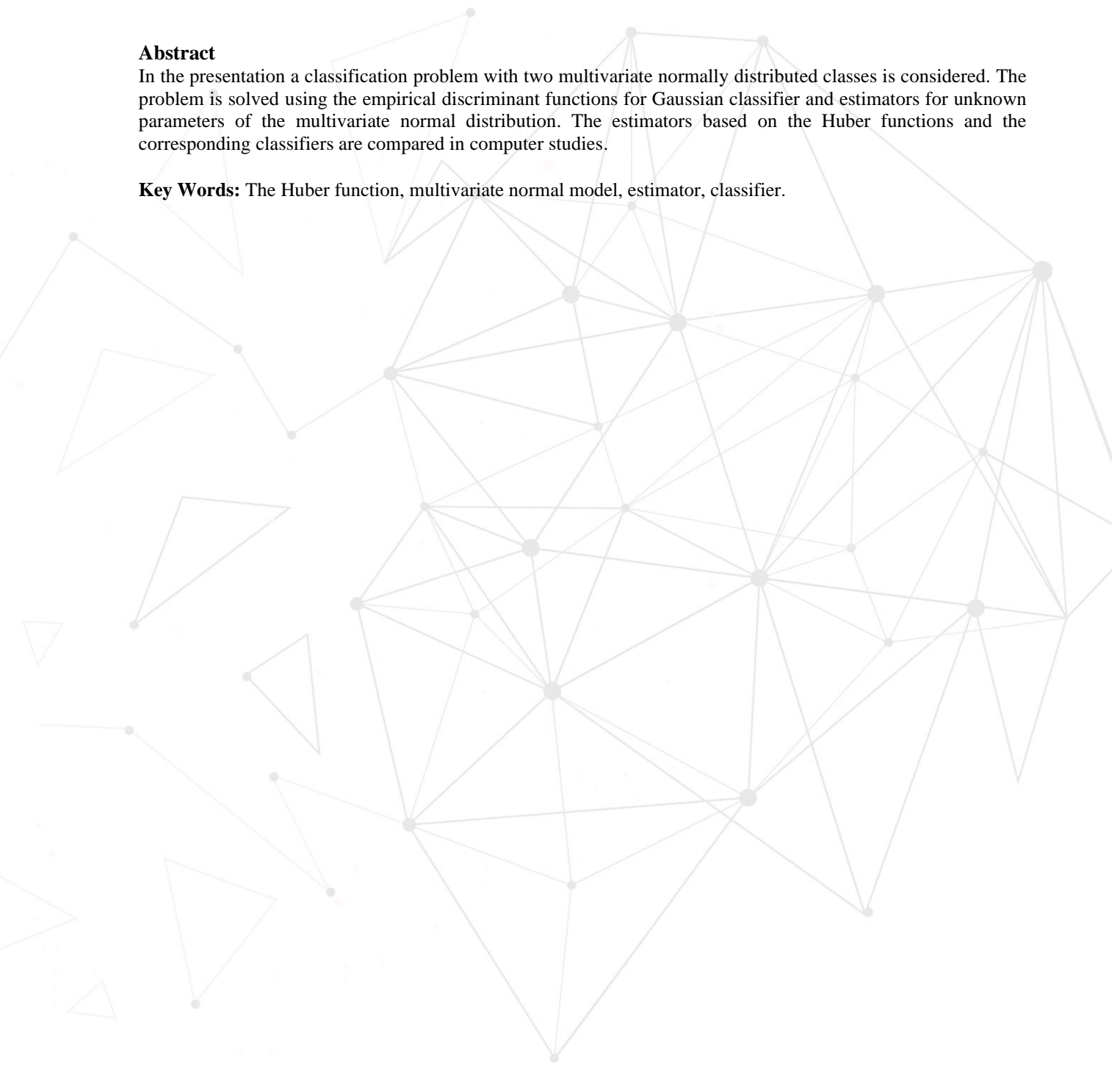
Henryk Gacki¹, Agnieszka Kulawik¹

¹Institute of Mathematics, University of Silesia in Katowice, Katowice, Poland

Abstract

In the presentation a classification problem with two multivariate normally distributed classes is considered. The problem is solved using the empirical discriminant functions for Gaussian classifier and estimators for unknown parameters of the multivariate normal distribution. The estimators based on the Huber functions and the corresponding classifiers are compared in computer studies.

Key Words: The Huber function, multivariate normal model, estimator, classifier.



Determining the Most Risky Production Area by Using Aggregated Criteria Importance Weights

Gülin Feryal Can¹, Pelin Toktaş²

¹Industrial Engineering Department, Başkent University, Ankara, Türkiye

²Industrial Engineering Department, Başkent University, Ankara, Türkiye

Abstract

This study proposes a new aggregated criteria weighting approach for Multi-criteria Decision Making (MCDM) problems. It is a well-known fact that different weighting approaches produce different importance weights for criteria. It is a valuable effort to consider all these different criteria weights in decision making. Different weighting results can provide decision makers (DMS) to interpret decision problem from different viewpoints. Additionally, it is necessary to know aggregated weighting result for DMs to reach a solution. To provide aggregated criteria weighting, this study suggests a new aggregation by considering different criteria weighting techniques' results. In this context, the most common used criteria weighting techniques selected from the literature as Standard Deviation, Criteria Importance Through Intercriteria Correlation (CRITIC), Entropy and Performance Selection Index (PSI) were performed to compute criteria importance weights in the study. Then, an aggregation operator was proposed to combine the weighting results produced by these four different techniques. Finally, Multi-Objective Optimization on the basis of Simple Ratio Analysis (MOOSRA) was utilized to rank alternatives considering aggregated criteria weightings. A case study related to determine the most risky production area in a firm which manufactures automobile glass was used to apply the proposed approach. Nine risk criteria as number of accidents that occurred during night shift, average age of workers having an occupational accident, average experience of workers having an occupational accident, the number of work accidents related to machines, the number of accidents related to operations, fatal accident number, injured accident number, the number of observed unsafe conditions and lost time were considered to order seven production areas as autoclave area, tunnel kiln area, dyehouse, cutting and processing area, ironworks area, bus oven area, vinyl room.

Key Words: Standard deviation, CRITIC, entropy, fuzzy, performance selection index, MOOSRA.

Creating Risk Based Quality Management Procedures in A Company Producing Medical Devices

Selin Sezer¹, Pelin Toktaş²

¹Quality Engineering, Başkent University, Ankara, Turkey

² Industrial Engineering, Başkent University, Ankara, Turkey

Abstract

At the present days, technological advances have reshaped the purview of medicine and many devices have been developed for diagnosis, treatment and rehabilitation. Very simple technologies have been used as treatment methods in the historical process and with the development of technology, more complex and high technology products have started to be produced and variety has increased. In accordance with these developments, there has been a cognitive complexity in time. In the course of time, general definitions have been developed by the competent authorities on a country-by-country or regional basis, with little differences, and legal regulations have been published. Although compliance with these regulations was based on conscience in the past, this compliance became compulsory in time and manufacturers were complied with these requirements.

When the establishment, operation and historical development of the Quality Management System in medical device manufacturing sites are monitored, legal obligations have been imposed in time to protect public health. Although the certification of medical equipment manufacturing facilities all over the world requires compliance with regional standards and legislations, the main purpose of all of them is to provide benefits to human health by establishing a quality management system in production. Compliance is a key to the success of an organization. At the same time, compliance with the legislation mandated by the competent authorities of different countries is critical for manufacturers to improve their sales networks.

In this study, ISO 13485: 2016 Quality Management System Standard for Medical Devices and 21 CFR Part 820 Quality System Regulation are compared and integrated to a medical device manufacturing company. As a result, it is observed that the most important stage is the risk assessment. Risk assessment methods are examined within the scope of related standards. Failure Mode Effect Analysis (FMEA) and Hazard and Operability (HAZOP) methods are integrated to each other and implemented to manufacturing processes. All risks are determined and necessary actions are taken. At the end of the study, all risks are reduced to acceptable levels.

Key Words: Medical device, risk assessment, quality management system, ISO 13485, 21CFR Part 820.

Vehicle Headway Modeling with Exponentiated Weibull Distribution using Ranked Set Sampling Method

Büşra Sevinç¹, Selma Gürler², Melek Esemem¹

¹The Graduate School of Natural and Applied Sciences, Dokuz Eylül University, İzmir, Turkey

²Department of Statistics, Dokuz Eylül University, İzmir, Turkey

Abstract

Vehicle headway distribution modeling is very important concept for intelligent transportation in traffic engineering. It is useful for the traffic signal optimization and flow modelling by characterizing the distribution of vehicles in a road. There are a variety of distributions that can be used for different traffic conditions. In this study, we use exponentiated Weibull distribution as a flexible model for headway data using ranked set sampling (RSS) and obtain maximum likelihood estimation of the parameters. We give a simulation study in R-software to compare the performances of SRS and RSS methods in terms of bias and mean squared error. Also, we give numerical results based on a real data application to show its flexibility.

Key Words: Vehicle headway, exponentiated Weibull, ranked set sampling, maximum likelihood.

Control Charts for the Compound Weibull Exponential Distribution under Truncated Life Tests

Canan Hamurkaroglu¹, Ayten Yiğiter²

¹Department of Actuary and Risk Management, Karabuk University, Karabuk, Turkey

²Department of Statistics, Hacettepe University, Ankara, Turkey

Abstract

Control charts are one of the important tools of statistical quality control. The quality characteristics of the product are monitored with the control charts to determine whether the process is statistically controlled and then the product is produced according to the specified specification limits. In general, control charts are examined in two classes for measurable and non-measurable properties. As attribute control charts for non-measurable features, the classic Shewhart p and np control charts are the most commonly used statistical control charts to monitor failed product ratio / number. When the quality characteristic is the lifetime of the product, the number of failures in the process can be monitored by the attribute control chart np for an appropriate distribution under the truncated life test. In addition, the performance of a control chart is given by its average run length (ARL). In-control ARL is a function of type I error, and out-of-control ARL is a function of type II error.

In this study, a quality control chart is obtained to monitor the number of failures under the truncated life test assuming that the product's lifetime has Compound Weibull-Exponential distribution. The control constants of the new chart are calculated for various quality parameters. Furthermore, the performance of the proposed schedule is given in ARL according to various process shift parameters.

Key Words: Control charts, average run length, truncated survival test, I. type error probability, II. type error probability, compound Weibull-Exponential distribution.

¹ e-mail: cananhamurkaroglu@karabuk.edu.tr

² e-mail: yigiter@hacettepe.edu.tr

Statistical Analysis of Wheat Flour Quality Parameters

Bariş Ergül¹, Arzu Altın Yavuz²

¹ESOGÜ Science and Letter Faculty Department of Statistics, Eskişehir, Turkey

²ESOGÜ Science and Letter Faculty Department of Statistics, Eskişehir, Turkey

Abstract

Wheat flour is of great importance in human nutrition. The quality of wheat flour varies significantly depending on climate, genetic variation, environmental and storage conditions. Bread, pasta and biscuits are usually made of wheat flour. The quality of these end products varies depending on the quality of the flour. In flour, quality is expressed by many measurable parameters. Measurement of these parameters can be done through physical, chemical, physicochemical and technological analyzes. In this study, logistic, robust logistics, logistic ridge and lasso logistic regression analyses were used to determine the variables used to determine the quality of wheat flour. The main purpose of logistic regression analysis is to try to explain the causality relationship between independent variables and dependent variable as with other regression methods with the help of the least variable. In logistic regression analysis, as in other statistical techniques, outliers may be found. Outlier values can be defined as values that differ greatly from the overall data. In logistic regression analysis, outliers in explanatory variables affect the probability of taking 0-1 values of the dependent variable. In addition, there is multicollinearity problem among many variables considered for the determination of wheat flour quality. Logistic ridge regression analysis was used as an alternative to the likelihood method to overcome this problem. In logistic regression, all explanatory variable coefficients are considered as values other than 0. The recently used Penalized Logistic Regression analysis adds a non-negative term to the likelihood estimator. Thus, the magnitude of the high dimensional coefficients of independent variables is controlled. Many penalty terms have been defined in the literature. LASSO regression is the most commonly used regression method. In this study, the above methods were used to determine the variables determining the quality of wheat flour and a comparison was made in terms of correct classification percentages.

Key Words: Flour, logistic regression, ridge, robust, Lasso.

¹ ESOĞÜ Science and Letter Faculty Department of Statistics, Eskişehir, Turkey

² ESOĞÜ Science and Letter Faculty Department of Statistics, Eskişehir, Turkey

Performance Evaluation of Non-Life Insurance Companies via Grey Relational Analyse and Analytic Hierarchy Process

Murat Kırkağaç¹, Övgücan Karadağ Erdemir²

¹Department of Insurance and Risk Management, Kütahya Dumlupınar University, Kütahya, Turkey

²Department of Actuarial Science, Hacettepe University, Ankara, Turkey

Abstract

In the insurance sector, which is quite competitive, insurance companies had better make performance evaluations periodically in order to incorporate potential insured persons and perform well. Performance evaluation is an analysis of whether the optimal output is achieved by using available resources effectively. In the performance analysis of insurance companies, certain financial ratios specific to insurance are utilized. Using these rates, insurance companies analyse their financial capabilities and their solvency margin. There are many methods suggested for performance evaluation. Data Envelopment Analysis is one of the most important methods of these methods. Furthermore, Principal Component Analysis, Malmquist-Total Factor Productivity, Non-hierarchical Clustering Analysis, Tobit Regression Model, Stochastic Boundary Analysis, TOPSIS, Analytical Hierarchy Process are also utilized. Recently, Grey Relational Analysis is often preferred for performance evaluation. In Grey Relational Analysis, which is one of the multi-criteria decision-making methods, grey refers to the situation where information is not fully known. This method is used to analyse the uncertainty in the multi-criteria decision problem. In this six-step method, the weight of the proportions utilized affects the results. At this point, the weight of financial ratios can be calculated by using the Analytical Hierarchy Process. In this method, decision makers try to make a decision for complex problems by modelling the relationship between the main objective, criteria, sub-criteria and alternatives in a hierarchical structure. Practicing performance analysis between specific periods rather than a single period is important for examining the change in performance on a yearly basis.

In this study, performance analysis of certain selected non-life insurance companies operating in Turkey for the last 5 years is carried out by incorporating Analytical Hierarchy Process into the Grey Relational Analysis. It is concluded that the weight of the financial ratios calculated with Analytical Hierarchy Process affected the Grey Relational Analysis results. Based on the performance results obtained with the help of these two multi-criteria decision-making methods, non-life insurance companies will be able to compare their performance to competing companies and over the years. According to the results of the comparison, the companies will have the opportunity to determine and improve the changes to be made in their assets, resources and management policies.

Key Words: Performance evaluation, financial ratios, multi criteria decision analysis, analytical hierarchy process, grey relational analysis.

Classification of Artworks with Deep Learning Methods: Ebru and Suminagashi

Aysin Tasdelen¹, Sebnem Ozdemir^{2 3}

¹Altınay Robotics, Istanbul, Turkey

²Management Information Systems (MIS), Istinye University, Istanbul, Turkey

³Computer Science and Artificial Intelligence Lab (CSAIL), Massachusetts Institute of Technology (MIT), Boston, USA

Abstract

Developments in deep learning methods have led to different applications in many areas. The use of these methods, especially in the field of art, has made it possible to perform studies that require a long, challenging and specific expertise, such as the transfer of the styles in the works, the authenticity of the artist in the work, and the ability to perform them more easily and faster. In this study, the classification of two art forms belonging to two different cultures, which are accepted as very similar because of their style, with deep learning techniques has been fulfilled. As a result, Barut Ebru, belonging to Turkish culture, and Simunagashi, belonging to Japanese culture, were classified with a success rate of over 90%. This result shows that even in the branches of art where the shapes are formed depending on the movement of the water, the originality of the culture can be determined by using deep learning methods.

Key Words: AI in Art, Deep Learning, Ebru, Simunagashi

¹ Aydin Tasdelen, Altınay Robotics, Istanbul, Turkey

² Sebnem Ozdemir, Management Information Systems (MIS), Istinye University, Istanbul, Turkey,
sebnem.ozdemir@istinye.edu.tr

³ Computer Science and Artificial Intelligence Lab (CSAIL), Massachusetts Institute of Technology (MIT), Boston, USA
sozdemir@mit.edu



A Numerical Approximation of Inversion Method to Generating Random Numbers

Orhan Kesemen¹, Buğra Kaan Tiryaki², Ayşe Pak^{3,*}

^{1,2} Department of Statistics and Computer Sciences, Karadeniz Technical University, 61080 Trabzon, Turkey

³ Department of Bioinformatics, Karadeniz Technical University, 61080 Trabzon, Turkey

Abstract

Two common approaches used in random number generation are the rejection and the inversion methods. The rejection method can be used for almost all distribution functions. However, it is seen as a weak method in terms of calculation time and success rate. On the other hand, the inverse method is a powerful generation technique, but it cannot be applied to all distribution functions. Therefore, a method which is as powerful as the inverse method, and which can be applied to every distribution function such as the rejection method, is proposed in this study. In the proposed method, the probability density function is divided into parts and the random numbers are generated on these each divided part by a linear approach. Especially, the most important approach in this study is to obtain the minimum error by using the optimal polygonal approach to the probability density function.

Key Words: Random number generation, rejection method, inverse method, numerical method.

¹ okesemen@gmail.com

² bugrakaantiryaki@gmail.com

³ aysebanu.pk@gmail.com



Sector-Wise Analysis of Cardinality Constraint Portfolio Optimization Based on a Hybrid Approach: Nonlinear Neural Network and Genetic Algorithm

Ilgin Yaman¹, Turkan Erbay Dalkilic²

¹Department of Statistics, Giresun University, Giresun, Turkey

²Department of Statistics and Computer Science, Karadeniz Technical University, Trabzon, Turkey

Abstract

The process of selecting a portfolio is still a very vital issue in the finance world. Harry Markowitz had proposed standard portfolio optimization in 1952. He had considered investors should maximize anticipated returns while minimizing the variance of return. That is variance-expected return trade-off. Because of the type of fitness function, the portfolio optimization problem is a kind of quadratic optimization problem. Besides, cardinality constraint is adding, which is a binary constraint, the problem becomes a mixed-integer quadratic optimization problem.

In this paper, we proposed a nested algorithm to solve that cardinality constraint portfolio optimization problem. The proposed algorithm is the combination of genetic algorithm and nonlinear neural network which solves cardinality constraint portfolio optimization problem effectively. Indeed, the genetic algorithm is used to solve the binary part of cardinality constraint portfolio optimization and the fitness function is used to solve with the nonlinear neural network. The network was represented in 2014 by Yan. The proposed neural network is based on solving primal and dual problems simultaneously. We had adjusted the neural network for portfolio optimization in 2019. In this study sector-wise analysis is done for cardinality constraint portfolio optimization problem. Five sectors are examined for the proposed algorithm which are Basic Materials, Consumer Cyclicals, Financials, Industrials & Technology, and Others. Istanbul Stock Exchange-all (ISE-all) data are used to solve the nonlinear neural network which is adapted to solve portfolio optimization.

Key Words: Nonlinear neural network, genetic algorithm, portfolio optimization, ISE-all.

Statistical Analysis Techniques of Determination of a Key Comparison Reference Value in International Measurement Comparisons

Sevda Kaçmaz¹, Hilmi Ege Subat²,

¹ TUBITAK - National Metrology Institute (UME), Kocaeli, TURKEY

² İstanbul Kültür University - Faculty of Science and Letters, Department of Mathematics and Computer Science, İstanbul, TURKEY

Abstract

The evaluation of key comparison data is a particularly significant task in the area of metrology, because of the relevance to global trade and because such comparisons are intended to test the principal techniques in the field. Within the such evaluation is the determination of a key comparison reference value (KCRV) and its associated uncertainty, and the degrees of equivalence of and between national measurement standards. This paper presents different statistical methods as Weighted Mean Value, Monte Carlo, Median and Linear Generalized Least-Squares Estimation to determine a comparison reference key value. And also the application of these methods in metrology is explained with an example, advantages and disadvantages are discussed.

Key Words: Key comparison reference value, weighted mean value, Monte Carlo, median and linear generalized least-squares estimation, uncertainty and metrology.

Forecasting Tourist Arrivals with Economic Regressors: A Grid Search on Combinations of Different Individual Models

Emrah Gülay¹, Korkut Bekiroglu², Okan Duru³

¹ The Department of Econometrics, Dokuz Eylül University, Izmir, Turkey

² The Department of Engineering Technology, SUNY Polytechnic Institute, Utica, USA

³ The School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

Abstract

Tourist arrivals forecasting has significance importance on the future developments of tourism industry to speed economic growth rate. Therefore, a great number of studies have attempted to forecast tourist arrivals by using a bundle of techniques based on statistical or econometric models that depend on historical data to forecast tourist visit in the coming future by assuming ceteris-paribus the economic environment. With accurate forecasts in tourism industry, the government's and industry player's decision-making processes include less risk and uncertainty by using tourism sources efficiently and economically.

Turkey is one of the popular destinations for all tourists all over the world. For foreign tourists most of the time what makes Turkey attractive and visited is not only natural beauties and summer tourism, but also her cultural and historical inheritance. According to United Nations World Tourism Organization (UNWTO), with 39.9 million incoming tourists, Turkey ranked sixth in the world. With increasing number of foreign visitors, the tourism revenue has been consistently increasing in Turkey. As is known, tourism income is one of the most important factors that contribute to the Turkish economy. Herein, the accurate forecasts of tourist arrivals will help to implement medium and long-term tourism strategies in terms of Turkish government policies. This study brings out a predictive analytics framework to the issue of selection of individual models by combining forecasts to improve forecasting accuracy.

In ensemble learning methods, many models are utilized to have a better prediction performance comparing with the utilization of one specific model or approach. A considerable amount of literature has been published on ensemble learning in machine learning literature since, in machine learning, the user can generate a particular infinite number of model and then these models are being averaged to create multiple models. In most cases, ensemble model averaging gives better results than a single model by using a different type of error metric. However, the primary challenge in forecasting literature is "overfitting" since if the weighting of the model coefficient or models could not be adjusted appropriately to contain only the main dynamics of the process, the overfitting might result in really bad forecasting performance. Therefore, in this paper, we address a sparse model combination for ensemble learning. We develop a self-calibrating high-level ideal forecasting algorithm combining various models forecasting proposed in econometrics, fuzzy logic, neural networks and other domains of research.

Key Words: Forecasting, grid search, ensemble learning, fuzzy, overfitting, artificial intelligence, univariate time series models, combining forecast.

ARIMA Model for Monthly Crude Oil Forecasting in Iraq

Qais Mustafa Abdulqader

Petroleum Geology and Sediments, Duhok Polytechnic University, Duhok, Iraq

Abstract

In this paper, the procedure of Box-Jenkins of Autoregressive Integrated Moving Average (ARIMA) has been applied for analysing and forecasting the exports of crude oil in Iraq by taking (72) observations of the monthly exports from 2013 to 2018. The data were divided into two parts, the first (60) observations was for training and the next (12) observations prepared for making a comparison between forecast and real values. Several suitable models of time series have been obtained and built and some of the performance measures have been used for the purpose of comparison between models. Results of the analysis concluded that the ARIMA (0,1,1) model is adequate to be used to forecast the monthly exports of crude oil in Iraq. The forecasted and real values of the year 2018 were close to each other. The forecasting process has done for the next 24 months. During the period 2019 to 2020, the exports of crude oil will reach to (148281000) barrel per month.

Key Words: Box-Jenkins, ARIMA models, time series forecasting, Crude oil.

Debt and Economic Growth Nexus in the US: Evidence from Asymmetric Causality¹

Nebile Korucu*

*Department of International Trade, Istanbul Kultur University, Istanbul, Turkey

Abstract

The relationship between economic growth and debt has been studied for decades. This (casual) relationship can be different according to debt types, regions and countries. In this sense, the main motivation of this study is to find out if there is a casual relationship. Moreover, we aim to analyze the nonlinearity between economic growth and debt for the US. This asymmetric relationship is analyzed by vector error correction model. Preliminary results show that there is an asymmetric causality for debt and growth.

Key Words: Debt, growth, asymmetry, panel vector error correction model, Turkey.

JEL Classification: O47, O51, H63.

¹ This study is supported by TUBITAK 2219 Postdoctoral Research Grant and done during the author's visit at Boston College, Boston, MA, the United States.

Investigation of the Effects of Research Supports on Academic Performance of Turkish Universities by Using Correlation, Regression and Path Analyses

Esra Satıcı¹, Bulut Tunca², Özge Akkuş³

¹Audit and Risk Management, Social Sciences University of Ankara, Ankara, Turkey

²Public Expenditure Analysis Department, Presidency of Strategy and Budget, Ankara, Turkey

³Statistics, Mugla Sıtkı Koçman University, Muğla, Turkey

Abstract

Ranking universities in terms of their academic performance has been analysed for a long time. Today, many centers have developed sorting criteria based on different indicators for the world universities. In our country, University Ranking by Academic Performance (URAP) research laboratory annually scores the world universities on the basis of quality of publications, quantity of publications and international research collaboration and releases the top 2500 universities around the world. In 2018, 87 universities from Turkey entered the that ranking list.

The aim of this study is to investigate the effect of the research supports given by TUBITAK to the academic performance of the Turkish universities. In addition to the research supports, it is thought to have an effect some other auxiliary variables such as the number of research application centers of the universities, the number of doctoral programs, the number of students in these programs, and the number of academic staff. For this purpose, correlation, regression and path analyses are used to obtain both the direct and indirect effects of these variables on academic performance.

Key Words: Regression analysis, path analysis, path coefficient.

Lifelong Learner: Question-Answering Chatbot

Ezgi Özen¹, Adil Alpoçak²

¹ Computer Engineering Master Student, Dokuz Eylul University, İzmir, Türkiye

² Computer Engineering, Dokuz Eylul University, İzmir, Türkiye

Abstract

Conversational agents have become a significant parts of our lives. They can be useful in many areas such as customer services or personal use in messaging apps on smartphones, aside from using chatbots for entertainment purposes. A chatbot is a computer program designed to simulate a human-like conversation with users. Although chatbots are very useful and popular nowadays, they still have some weaknesses. Most systems use only natural language processing and machine learning techniques for building chatbots. Yet, they do not have the capability to learn continuously and interactively. They do not improve themselves during the conversation.

In this paper, we propose a question answering chatbot system that is able to learn new knowledge during the conversation. We use Natural Language Processing to interpret the user's sentences. Then, the system classify these sentences into either a statement or a question and generate a related response. After that, we use Google Knowledge Graph API to answer user questions if the chatbot's knowledge base does not have the related answer. From the user sentence, the system extract knowledge via using Stanford OpenIE package. The knowledge is stored as a semantic triple relation, consisting of subject, predicate and object. The chatbot is modeled to ask and learn like humans when it encountered with a new concept such as new word or a relation. Our system answers to questions from different users via Knowledge Base or Google API results. In conclusion, a continuous and interactive learning chatbot design looks promising.

Key Words: Question answering, chatbot, lifelong learning, knowledge mining, interactive learning.

Generalization and Compounding Statistical Distributions, Goals and Benefits: A Review

Ehsan Fayyazı¹, Serpil Kılıç Depren²

¹ PhD candidate, Department of Statistics, Faculty of Arts and Sciences, Yildiz Technical University, Istanbul, Turkey.

² Assoc.Prof.Dr., Department of Statistics, Faculty of Arts and Sciences, Yildiz Technical University, Istanbul, Turkey.

Abstract

Regarding the practical problems in statistics, the term generalizing distributions has been identified to be as important as other its counterparts. The underlying assumptions concerning the generalizing distributions is the primary use of forms related to the mathematical functions followed by the induction of two dimensions, namely, location, scale or inequality parameters. In 1997, the early practices of generalization were put forth by induction of shape parameter(s), which were continuously pursued for the last two decades. To address the challenging conditions pertaining to series and parallel structures, there were attempts to employ an integrated form of discrete and continuous in 1998. Accordingly, the purpose of the current study is to shed light on the compounding univariate distributions along with their extensions and classes. To this end, initially several available compound classes are reviewed and then some new classes are suggested. Also, the current research work discusses the latest perspectives toward the construction of generalized and compounding classes. Ultimately, the study offers several guidelines for the future researches.

Key Words: Generalized classes, compound G-classes, Marshall-Olkin class, Zografos-Balakrishnan class.

¹ Corresponding author: ehsan.fayyazi.s@gmail.com.

Alpha Power Transformed Two-Parameters Rayleigh Distribution: Properties and Inference

Hakan Kızılkaya¹, Cenker Biçer²

¹Turkish Military Academy, National Defense University, Ankara, Turkey

²Statistics Department, University of Kırıkkale, Kırıkkale, Turkey

Abstract

In the present paper, we introduce a new probability distribution model called alpha power transformed two-parameter Rayleigh (APTTR) for modeling the positive real-valued and skewed data. In addition to being more flexible than the Rayleigh and two-parameter Rayleigh, the introduced distribution possesses important features of both the Rayleigh and two-parameter Rayleigh distributions in some special cases. By the study, some basic characteristics of the APTTR distribution such as moments, characteristic function, moment generating function, median, and hazard rate function are obtained. To estimate the parameters of the APTTR distribution, the different estimators are obtained by utilizing the maximum likelihood, moments, least-squares, and maximum spacing methods. Also, by extensive Monte-Carlo simulation studies performed at different sample sizes and different parameter values, the estimation performances of the maximum likelihood, moments, least-squares and maximum spacing estimators of the APTTR distribution parameters are compared according to the bias and MSE criteria.

Key Words: Alpha power transform, Rayleigh distribution, statistical inference, point estimation, simulation.

¹ Hakan Kızılkaya, Turkish Military Academy, National Defense University, e-mail: hkizilkaya@kko.edu.tr

² Cenker Biçer, Statistics Department, University of Kırıkkale, e-mail: cbicer@kku.edu.tr

Statistical Errors in Specialization in Medicine Thesis

Ayşegül Yabacı^{1,2}, Fatma Ezgi Can², Fisun Kesin², Robab Ahmadian², Ilker Ercan²

¹Biostatistics and Medical Informatics, Bezmialem Vakif University, Istanbul, Turkey

²Biostatistics, Uludag University, Bursa, Turkey

Abstract

When researchers, publishing in scientific journals in the field of Medicine, do not have sufficient statistical knowledge, they can make mistakes in the planning, design, execution, analysis and presentation of the data. Since the 1960s and 1970s, many authors, seeking to draw attention to these errors and deficiencies in methods and statistics in scientific studies, have published their researches and suggestions on this subject many times. The elimination of deficiencies in this subject is important not only for the scholars who carry out scientific research, but also for the physicians who need accurate and complete scientific studies along with the progress of evidence-based medicine in current clinical applications and follow the current developments in medical science from scientific journals.

The aim of this study is to evaluate the thesis of specialization in medicine in terms of statistical errors made and, thus, to contribute to the production of quality scientific publications by ensuring that scientific publishers in the field of Medicine are sensitive and careful about statistics when doing their work. In this study, we investigated 88 thesis which are defended from 6 different universities are obtained from the database of YOK. The investigation is conducted in terms of “Errors related to p values”, “Errors related to tests”, “Mathematical notation errors”, “Statistical symbol errors”, “Inappropriate interpretation”, “Presentation of the statistical method analysis and results in the incorrect section of the manuscript”, “Errors in summarizing data”, “Incomprehensible statistical terms”, and “Errors in statistical terminology”. All thesis had at least one statistical error of the 88 theses of specialization in medicine.

Key Words: Statistical errors, specialization in medicine thesis, statistical review.

Log - Two - Parameter Xgamma Distribution: Properties and Inference

Esra Kızılkaya¹, Cenker Biçer²

¹ Statistics Department, University of Kırıkkale, Kırıkkale, Turkey

² Statistics Department, University of Kırıkkale, Kırıkkale, Turkey

Abstract

In the current study, a new two parameters probability distribution family called log-two-parameter Xgamma distribution is introduced. The introduced distribution is a possible alternative to the famous lifetime distributions for modeling the skewed lifetime data. The log-two-parameter Xgamma distribution provides the features of Xgamma and two-parameter Xgamma distributions in some special situations. By the study, some base features of the log-two-parameter Xgamma distribution such as moments, characteristic function, moment generating function, median, survival function and hazard rate function are obtained. Furthermore, the statistical inference problem is investigated by using different estimation methods such as the maximum likelihood, moments, least-squares, and maximum spacing methods. Through the Monte-Carlo simulation studies performed on the different sample sizes, estimation performances of the obtained estimators by this study are compared by using the bias and mean squared errors (MSE) criteria.

Key Words: Xgamma distribution, lifetime data, statistical inference, point estimation, simulation.

¹ Esra Kızılkaya, Statistics Department, University of Kırıkkale, e-mail: esra_zeyno@outlook.com.

² Cenker Biçer, Statistics Department, University of Kırıkkale, e-mail: cbicer@kku.edu.tr.

Statistical Evaluation of the Cultural Difference Problem Experienced by Refugees in Health Services

Arzu Altın Yavuz¹, Barış Ergül², Şenay Kızılkum³

¹ESOGÜ Science and Letter Faculty Department of Statistics, Eskişehir, Turkey

²ESOGÜ Science and Letter Faculty Department of Statistics, Eskişehir, Turkey

³T.C. Eskişehir Governorship, Eskişehir, Turkey

Abstract

In recent years, millions of people have been forced to leave their homelands and seek refuge in foreign countries for various reasons, especially wars. Migration, which we define as individual or mass displacement movement, has affected societies to a great extent. As a natural extension of the outcome of migration, cases of refugees or asylum seekers have had as much impact as migration itself. Turkey due to its geographical location, the ongoing civil war in Syria has been affected as a result of the long-standing refugee crisis.

To start coming to Turkey and the Syrian people's culture in Turkey to remain in their culture has led to some problems. Even from this perspective, the issue of immigration and refugees leads to major changes globally. In addition, the high number of refugees in the country ensures that the discussions remain on the agenda. In this study, a questionnaire on language and cultural differences of Syrian refugees under temporary protection was designed. Syrian refugees, problems with the health care they interact while they benefit from health services in Turkey has been asked to examine. Multivariate statistical methods were used to reveal the problems faced by Syrian refugees due to language and culture differences.

Key Words: Syrian, language and cultural differences, migration, multivariate statistics, refugee.

¹ ESOĞÜ Science and Letter Faculty Department of Statistics, Eskişehir, Turkey

² ESOĞÜ Science and Letter Faculty Department of Statistics, Eskişehir, Turkey

³ T.C. Eskişehir Governorship, Eskişehir Turkey

An Approach of Feature Selection Consensus for the Gas Sensors' Drift Data

Çağatay Berke Erdaş ¹, Gülnur Begüm Ergün ², Selda Güney ²

¹Department of Computer Engineering, Başkent University, Ankara, Turkey

²Department of Electrical and Electronics Engineering, Başkent University, Ankara, Turkey

Abstract

Although, chemical sensors' usage areas are expanding day by day, they are suffering from the drift. This situation brings many ideas to the literature, yet it is still a problem to deal with. In order to compensate drift effect, calibration methods are utilizing. However, the motivation of this study is not data drift correction, classifying the dataset despite the drift effect. For serving this purpose, we studied on a comprehensive dataset which was consisted in three years. As a classifier, we considered three different algorithms: k-nearest neighbors (k-NN), random forest (RF) and support vector machine (SVM). On the other side, we tried to reduce dimension of the features. Due to reach this aim, we operated all features into five feature selection algorithms: Relief selection, chi-square selection, correlation-based feature selection (CFS), information-gain based selection (InfoGain) and gain-ratio-based selection (GainRatio) individually. The results have shown that the classification accuracy is very promising even though the corrupted data. Besides, it is an indication that it is possible to achieve that promising classification accuracy with only a few features. This output of the process may be a strong motivation for the future works regarding metal oxide gas sensor applications.

Key Words: Classification, dimension reduction, feature selection, metal oxide, sensor drift.

The Ranking of Cities for Household Consumption Expenditures in Turkey with RANKING-KEMIRA-M Approach

Pelin Toktaş

Industrial Engineering, Başkent University, Ankara, Turkey

Abstract

One of the most important indicators of welfare of a country or society is household consumption expenditures. These expenditures are the factors that closely concern the growth performance of the economy. In this study, integrated RANKING-KEMIRA-M approach that combines Modified Kemeny Median Indicator Rank Accordance (KEMIRA-M) and ranking methods for criteria weights is proposed to sort the different regions of Turkey in terms of household consumption expenditures. According to KEMIRA-M procedure, criteria are divided into two sets as necessary consumption expenditures (NCEs) and luxury consumption expenditures (LCEs). The expenditures for food and non-alcoholic beverages, clothing and footwear, housing, transportation and communication are determined as NCEs. Other expenditures for alcoholic beverages, cigarette and tobacco, furniture, household appliances and homecare services, health, entertainment and culture, education services, restaurants and hotels, various goods and services are given as LCEs. The data are obtained from 2016 Household Budget Survey of Turkish Statistical Institute. Cities of Turkey are divided into 26 regions which constitute the alternatives of the study.

Five different economists are selected as decision makers (DMs) to rank criteria sets independently. The ranking preference for a DM which minimizes sum of distances to the priorities given by all DMs is chosen as median components priority. Then, criteria weights reflecting median priority components are determined. This study recommends using 3 different ranking methods to calculate criteria weights: Rank sum, rank reciprocal and rank exponent methods. The weight pair for NCs and LCs is selected to minimize the sum of absolute value for weighted normalized mean differences of two criteria sets for 26 different regions of Turkey. Then, the final ranking of regions is obtained.

Key Words: KEMIRA-M, ranking methods, household consumption expenditure, MCDM.

Evaluation of Feature Reduction and Classification Methods for Activity Recognition

Çağatay Berke Erdaş¹, Didem Ölçer¹

¹Başkent University, Department of Computer Engineering, Ankara, Turkey

Abstract

In recent years, intensive research on the monitoring and recognition of human activity has been carried out on models of classification of data collected via motion sensors. In addition to the aforementioned models, a new smartphone based online human activity recognition (HAR) system is presented for the classification of activities. In this study, we present the different classification algorithms performances and the effect of Principal Component Analysis (PCA) method, which is dimensionality-reduction method by transforming a large set of variables into a smaller one, on the performance of algorithms. We experiment benchmark the basic human activities and postural transitions (HAPT) recognition dataset [1].

In this approach, Random Forest (RF), Multilayer Perceptron (MLP), Naïve Bayes (NB), and Classification and Regression Tree (CART) algorithms are using to obtain best performance. Algorithms are validated through with a 10-fold cross validation technique and the results are obtained as accuracy.

While the use of PCA results in a smaller decrease in the performance of RF and CART algorithms; on the other hand, the performance of NB and MLP algorithms led to an increase. The improvement in NB, which assumes all variables are not correlated to each other, was an expected result since the features became independent after applying PCA. The results showed that RF achieves 96.73% accuracy and the most successful performance with PCA was achieved with 95.81% with MLP.

Key Words: Activity recognition, random forest, multilayer perceptron, naive bayes, cart, principal component analysis.

Reference

[1] Reyes-Ortiz, J. L., Oneto, L., Ghio, A., Samá, A., Anguita, D., & Parra, X. (2014, September). Human activity recognition on smartphones with awareness of basic activities and postural transitions. In International conference on artificial neural networks (pp. 177-184). Springer, Cham.

Some Properties Provided by Fourier Transform with Distribution Function in Statistics

Seçil Yalaz¹,

¹Dicle University, Faculty of Science, Department of Statistics, Diyarbakır, Turkey

Abstract

Fourier transform, which is frequently used in mathematics, can be realized with both distribution function and density function in statistics. However, in general, density functions may not be known, not integrable, or the convergence rate may be much slower when considering asymptotic properties. Considering such situations, the Fourier transform using the distribution function is recommended in statistics. This study includes some statistical properties of Fourier transform performed with distribution function which is more convenient to use under certain conditions. The identifiability of Fourier transform performed by distribution function was examined, information was given about whether it provides the central limit theorem and its relationship with other distributions was examined.

Key Words: Fourier transform, distribution function, identifiability, central limit theorem, relations with other distributions.

¹ syalaz@dicle.edu.tr.

Examining the Response of Olfactory Stimulus in EEG Signals

Begüm Kara Gülay¹, Neslihan Demirel¹, Alper Vahaplar¹, Çağdaş Güdücü²

¹ Department of Statistics, Dokuz Eylül University, Izmir, Turkey

²Department of Biophysics, Dokuz Eylül University, Izmir, Turkey

Abstract

Electroencephalography (EEG) is the recording of electrical activity along the scalp produced by the firing of neurons within the brain. EEG is portable, non-invasive, relatively cheap and provide signals with a high temporal resolution. Many researchers have been trying to explore some more effective ways to analyze and classify EEG signals for identification of neurological diseases. In this study, the response of olfactory stimulus is examined. Data used in the study consists of recordings from 64 electrodes of different individuals selected among both control and patient groups. Each individual has different number of olfactory stimulus. The aim of the study is to detect a possible abnormal processing of the olfactory stimuli using statistical methods and presenting the preliminary results.

Key Words: EEG, olfaction, statistical learning.

Research of Stochastic Properties of a Population Model with Fokker Planck Equation Under Allee Effect

Çağatay Eskin¹, Özgür Gültekin^{2,3}

¹Department of Physics, Middle East Technical University, Ankara, Turkey

²Department of Physics, Istanbul University, Istanbul, Turkey

³Department of Statistics, Mimar Sinan Fine Arts University, Istanbul, Turkey

Abstract

Research on population models provides us rich theoretical tools for understanding the universal properties of systems that display chaotic behavior in different fields of science, as well as practical applications for population dynamics. For example, stability analysis of logistic population models have been an interesting research area for many years since it contributes to wide range of fields from evolutionary analysis to system biology, psychology to modeling of social events, chemical kinetics to kinetic theory of gases, modern cosmology to quantum chaos, and from artificial intelligence research to cryptology. In addition to this, recently, we see that, population models that describe populations with Allee effect have been extensively studied in the literature.

It is expected that for Logistic population models in high densities, as the density of the population increases, per capita growth rate of the population decreases. With this usual result, in low densities, as the density of the population decreases, the decrease in the growth rate of population is known as Allee effect. If the per capita growth rate is negative when the population size falls below a certain threshold and the process results in extinction, this is strong Allee effect. However, if there is no such threshold as the population density decreases and the growth rate remains low but positive at low population density, this is weak Allee effect.

In this study, firstly a population model under the influence of Allee effect will be discussed in terms of the limits of strong and weak Allee effect and the stability characteristics of the deterministic model by the means of its potential functions characterizing the population variation. Then, in-population and out-of-population fluctuations will be added to the population model to investigate their effects on the time evolution of population. In other words, noise terms whose statistical properties are clearly given will be added to the model. Thus, population change can be examined with a more realistic model considering stochastic properties. Stationary probability distribution function of population will be calculated by the solution of Fokker-Planck equation. In this way the effect of noise on stationary probability distribution function and the mean transition time between the steady states of the population will be discussed under strong or weak Allee effects for different parameters in the population model.

Key Words: Allee effect, population model, logistic equation, Fokker-Planck equation.

A Satellite Images Based Sampling Approach in An African City

Bayram Samet Şahin¹, Hakan Demirbüken², M. Akif Bakır³

¹Hacettepe University, Institute of Population Studies, PhD Student, Social Research Methodology, Ankara, Turkey

²UDA Consulting, Project Director, Ankara, Turkey

³Gazi University, Department of Statistics, Ankara, Turkey

Abstract

The spatial sampling methods have been widely employed for household surveys especially in conflict regions. However, these methods are also being used recently for household surveys where there is limited information available on the target population which would serve as sampling frame.

In the context of many African countries, administrative records are incomplete and even household or person register system does not exist. Besides, full household listing procedure is costly and can be risky in some African countries because of security reasons as well. Therefore, survey statisticians explore alternative methods using a combination of satellite maps and area-based sampling in these situations. There are many alternative methods which integrate satellite images, listing and gridding of PSUs. In this study, a sampling methodology from a poverty survey conducted in one of the capital cities in Africa is presented with details. The methodology is based on grid sampling methodology where a uniform grid was overlaid on the city map and PSUs were selected from each stratum. Within each selected PSUs, randomly selected households, SSUs, were interviewed.

Overall, the adopted methodology provided a reliable and consistent poverty estimates compared to other large-scale representative household surveys. The method has several strengths from practical aspect: it reduces the workload of field team, minimizes their discretion in sample selection, minimize enumerators' bias and allows for random selection with known probabilities. However, there are some weakness of the methodology such as the satellite images may be difficult to update, demarcation of boundaries is not easy to handle, and overlapping buildings into different grids. The paper also discusses the weight calculations and estimations techniques.

Key Words: Spatial sampling, household survey, sampling weights, GIS for sampling, grid sampling.

Data Science and Big Data Analytics Education - Serbian perspective

Dragan Vukmirovic¹, Tijana Comic², Zoran Radojicic³

¹Faculty of Organizational Science, Belgrade University, Belgrade, Serbia

²GOPA mbH, Bad Homburg, Germany

³Faculty of Organizational Science, Belgrade University, Belgrade, Serbia

Abstract

Data Science and Big Data Analytics have come up against significant and somewhat unplanned growth, caused by the Big Data revolution, which is reflected in the proliferation of complex data (in its format, not only by volume) in science, industry, government and society in general.

Only on the GitHub platform is currently registered more than 600 programs in Data Science, Big Data Analytics and related fields at over 200 universities around the World (<http://datascience.community/colleges>). There has been a significant increase of undergraduate programs being conducted at research institutions and liberal arts colleges. The vast majority of Data Science and Big Data Analytics Education are master's degree and certificate programs that are conducted in traditional way and/or online as distance learning. PhD programs in these areas are still relatively rare and their expansion may be expected in the near future.

However, regardless of the wide range of education programs at all levels of study, it is justified to question whether discipline, data science and Big Data Analytics are still in its infancy. By examining the current diversity of curricula and programs, this disciplines have yet to go through an important period of experimentation that should result in some level of standardization in order to best educate new generations of data scientists.

While professors tailor programs, no less confusion is among students - the results of a research in Serbia showed that students of applied mathematics and statistical programs are not interested in technical fields, and vice versa - those with pre-dominant technical studies do not want to devote additional attention to mathematics and statistics. However, practice has no time to wait - the sector of information and communication technology is in the expansion and urgently need experts from these areas.

This paper include: Consideration of basic educational concepts in Data Science and Big Data Analytics; Comparative analysis: Serbia – World; Serbia: Status and Perspective; Dilemmas and Recommendations for: data scientists, higher-education institutions and decision makers at the highest level – Government

Key Words: Data science, big data analytics, education, curricula.

Preemptive Goal Programming for the Daily Shipment Planning of an Automotive Lubricant Company

Deniz Türsel Eliiyi¹, Sel Özcan¹, Uğur Eliiyi²

¹ Izmir Bakircay University, Department of Industrial Engineering, Izmir, Turkey

² Dokuz Eylul University, Department of Computer Science, Izmir, Turkey

Abstract

In this study, the daily shipment planning problem of a company producing automotive and industrial lubricants is addressed. The problem involves a single distribution plant in Izmir, delivering products to geographically dispersed customers within Turkey. The objectives of the problem are to meet the total daily demand on time as much as possible, and to minimize the total transportation cost. The distribution routes to the customers, the vehicle types to be used on the distribution routes, and the products to be transported on each vehicle should be optimized based on the stated objectives. As the company uses third party logistics providers for the distribution of its goods, the return journey routes of the vehicles do not need to be considered. For this reason, this daily shipment planning problem is considered as a multi-objective variant of the Open Vehicle Routing Problem. We propose three preemptive goal programming models for solving this multi-objective optimization problem, in which the customer requirements, order splitting and the capacity constraints are considered. The proposed mathematical models are compared over real-life instances of the problem. This study is supported by the Scientific Research Projects Coordination Office of Izmir Bakircay University.

Key Words: Distribution problems, vehicle routing, multi-objective optimization, goal programming.

The Use of AHP-TOPSIS Method in Technopark Preferences of Technology Companies: Istanbul Case

İsmail Durak¹, Hakan Murat Arslan², Yahya Özdemir³

¹Department of Business Administration, Duzce University, Duzce, Turkey

²Department of Management Information Systems, Duzce University, Duzce, Turkey

³Department Of Computer Technologies, Yalova University, Yalova, Turkey

Abstract

Technoparks are innovation and technology center established to enhance the collaboration and cooperation between the university and companies. Companies working in technoparks have a high interaction with other companies. Communication of companies with other companies within the technopark, spending time together, making projects, participating in conferences contribute positively to the development of these companies and thus the country. In this study, it is aimed to determine the criteria that the companies in the technoparks in Istanbul take into consideration in their technopark preferences and to select the most appropriate technopark based on these criteria. To this purpose, current data about the technoparks in Istanbul are used. In the application part, Analytical Hierarchy Process (AHP) and Technique for Order Preference by Similar Solution (TOPSIS) methods are used separately in the selection of the most appropriate facility location of the companies in Istanbul Technoparks. The relevant data are obtained from the managers of the companies in the technoparks. These managers are also considered as decision-makers. Taking into account the general structure of the technopark member companies in Istanbul, it is determined that the most suitable facility location is the A1 (ITU) alternative as a result of the analyses carried out by two different methods. The results of the analysis are shared with the relevant managers. For later studies, hybrid or fuzzy logic based multi-criteria decision making (MCDM) methods can be used.

Key Words: AHP and TOPSIS methods, technopark preference, facility location, Istanbul.

Two-Stage Solution Framework for a Workload Balancing and Workforce Scheduling Problem: A Comparative Study

Sel Özcan¹, Uğur Eliiyi², Deniz Türsel Eliiyi¹

¹ Izmir Bakircay University, Department of Industrial Engineering, Izmir, Turkey

² Dokuz Eylul University, Department of Computer Science, Izmir, Turkey

Abstract

This study investigates the operational workload balancing and workforce scheduling problem of a leading international company in the oil and gas sector. The company is planning to launch a new product in all its gas stations in Turkey. The company has around 600 gas stations, dispersed all over the country. For marketing purposes of the new product, the brandings of all gas stations also need to be renewed. A multi-stage discovery, branding manufacturing, and construction work are required to be performed for this purpose at each station. A worker crew is assigned for handling this process countrywide. The first problem is involved with balancing the workload (in terms of gas stations) among the worker groups within the crew. For this problem, we propose an integer programming formulation and optimize the worker group assignments to the gas stations. The second problem focuses on obtaining a permutation of the gas stations to be visited by each worker group, where the objective is to minimize the total distance travelled by each group while completing the required work on time. We define this second problem as a multi-period, multi-depot, multiple traveling salesman problem, which incurs the decision of the set of nodes visited by each worker group in non-conflicting schedules, where each group is located at different depots. To solve the resulting balancing and scheduling problem of the company, we propose a two-stage solution approach. In the first stage, the gas stations are partitioned into a given number of disjoint regions, whereas in the second stage the schedules of the decomposed regions are determined. A computational study is carried out with real data and significant improvements are achieved with the proposed solution approach.

Key Words: Workload balancing, workforce scheduling, traveling salesman problem, optimization.

Machine Learning Based Feature Extraction for Determination of Balance Disorders

Serhat Ikizoglu¹, Ezgi Demir², Batuhan Atasoy³

¹Department of Controls and Automation Engineering, Istanbul Technical University, İstanbul, Turkey

²Department of Management Information Systems, Piri Reis University, İstanbul, Turkey

³Department of Mechanical Engineering, Piri Reis University, İstanbul, Turkey

Abstract

Balance disorders are the type of illnesses which are mainly caused from the disorders at the cerebellum or the otoliths. There are studies for the feature extraction processes, mainly Parkinson's disease. Some measurements have been held via speech signals, gait analyses, etc for Parkinson's disease, ALS and other type of illnesses.

In this study, apart from Parkinson's disease, other type of illnesses have been examined. In order to perform this, time series of force data, which have been obtained from the sole of the feet, have been used. Then the data has been analyzed, and then any feature to classify the type of disorders have been aimed to succeed. Afterwards, using machine learning techniques, the type of balance disorders have been aimed to be classified. Finally, by using chaotic approach based methods, the results obtained from machine learning methods have been compared.

As a first stage, the study is progressing to find one or more features to classify these disorders.

Key Words: Balance disorders, machine learning, chaotic analyses methods, feature extraction.

Supervised Learning Based Apparel Recommendation System for Online Shopping System

Selver Ezgi Küçükbay^{1,2,*}, Nafiye Kübra Turhan^{1,#}

¹Middle East Technical University, Department of Computer Engineering, Ankara, Turkey

²Başkent University, Department of Computer Engineering, Ankara, Turkey

Abstract

Recommendation systems are necessary for people to find products that are of interest to them. In today's world, finding a proper apparel is a challenging work especially for people who have no time for searching on the Internet. In this study, we present a recommendation system for online shopping by using the Dress Sales dataset. The utilized dataset contains attributes of dresses and recommendations according to their sales. We construct a supervised learning based structure to compare performance of different classifiers. According to the models, recommendation system aims to identify if a dress should be recommended or not. At the end of the project accuracy results of the classification techniques on proposed system are compared.

Key Words: Data analysis, data mining, recommendation systems, support vector machines, neural networks, naïve bayes, decision tree.

* seyalniz@baskent.edu.tr

kubra.turhan@metu.edu.tr

Predicting Loss Aversion Behaviour with Machine Learning Methods

Omur Saltik¹, Ridvan Soyu², Suleyman Degirmen³

¹Dept. of Economics, Mersin University, Turkey

²Dept. of Computer Science, Toros University, Turkey

³Dept. of Economics, Mersin University, Turkey

Abstract

The behavioural economics literature has shown from the 1950s onwards that economic agents have not always made decisions to maximize a particular utility function, as Von Neumann and Morgenstern (1944) have stated. It has been observed that economic agents, or as we can interested in micro-level decision makers, exhibit certain biases and anomalies (Allais Paradox, Reflection Effect, Loss aversion, St. Petersburg Paradox etc.) under uncertainty, and thus, may have different risk for consumption decisions, investment decisions, reward, incentive and penalty systems (The Nudge Theory et. al). It is an important effort to determine the macro effects of decisions taken by individuals at the micro level within the economy. At this point, behavioural economics and neuro-economics use the empirical findings and parameters of cognitive theories obtained from clinical experiments beyond the assumptions of classical financial models to model the risk-lover, risk-averse or risk-neutral behaviours of individuals. Artificial intelligence applications, founded by Thomas Bayes in 1800s and continued with Markov Chain in 1900s and experienced a rapid development with control theory and neural network after 1960s, have been integrated into social sciences, natural sciences and health sciences through various machine learning algorithms especially with the invention of backpropagation method in 2006. In this context, research question is which model and parameters will be applied based on activation functions, a very important role in machine learning rules used frequently in social sciences in recent years. These models and parameters are determined as a result of the experimental studies applied in the field of behavioural economics and neuro-economics. The most of the time, these parameters and models based on these decisions made by individuals by calculating a probability and reward amount presented to them have some kind of biases and anomalies. It is worthwhile to develop the small actual sample data into greater sample data through machine learning rules in order to predict the general tendency of population against a reward, accept a penalty, or take a risk. Therefore, in this study, we carried out an experiment with 28 participants to determine an important cognitive phenomenon called the Loss Aversion covered by the Prospect Theory (Kahneman and Tversky, 1979) and then, it predicts some random participant decisions with machine learning rules. In this single trial experiments, participants were asked to make 256 decisions under certain level of risk that contains some monetary rewards along with probabilities for each trial. The data produced by 28 people in the study were collected and arranged appropriately to use in machine learning algorithms and divided into random groups for training and testing data. Predictions has been made by supervised machine learning rules. To do this, the study benefits of python program and related studies to measure the prediction power of the model: the Mean Error Squares information criterion as a loss function.

Key Words: Behavioral economics, prospect theory, loss aversion, bias and anomalies, machine learning.

¹ omursaltik09@gmail.com

² ridvansoyu@gmail.com

³ suleymandegirmen@gmail.com

Predicting Bubbles in the S&P 500 Index with Macroeconomic Indicators: A Machine Learning Approach

Fatma Başıoğlu Kabran¹, Kamil Demirberk Ünlü²

¹Department of Finance, Banking and Insurance, İzmir Kavram Vocational School, İzmir, Turkey

²Department of Statistics, Ankara University, Ankara, Turkey

Abstract

In this study, we are interested to predict the bubbles in the S&P 500 index with macroeconomic and financial variables by employing a machine learning algorithm, namely Support Vector Machine (Vapnik, 2013). Support Vector Machine (SVM) is a non-parametric and non-linear binary classification technique which already proved itself in financial time series forecasting. In the literature, a bubble is often defined as a situation where the asset price exceeds its intrinsic value. Bubbles can be used by policymakers as early warning signals to take measures in order to prevent the possible financial crises. Therefore, a large number of studies in the literature have tried to identify the bubbles in the stock market indices while a growing number of studies have explained them by macroeconomic and financial variables such as inflation rate, unemployment rate, interest rate and dividend yields. Our study consists of two parts. The first part is devoted to identify the bubbles in the S&P 500 index using a widely recognized right tailed unit root test, generalized sup augmented Dickey Fuller test (C. B. Phillips, Shi, & Yu, 2015a). Then, SVM is employed to explain the previously identified bubbles by macroeconomic and financial variables. We conclude our study by comparing the performance of the SVM with traditional parametric logistic regression model.

Key Words: Bubbles, early warning, machine learning, support vector machines, macroeconomic indicators.

References

- C. B. Phillips, P., Shi, S.-P., & Yu, J. (2015a). Testing for multiple bubbles: historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*(56), 1043-1078.
Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Credit Scoring with using Alternative Model Techniques

Zeynep Bal¹, M.Aydın Erar¹

¹ Department of Statistics, Mimar Sinan Fine Arts University, İstanbul, Turkey

Abstract

Credit is the main operating income of the banks so the credit risk that may arise from this activity is the most critical risk area on which banks focus. Within the scope of managing this risk area, banks develop customer specific risk models through their analytical units. In this context, the models used in credit scoring have the most important share in the banks making the decision to provide loans to the consumer.

In this study, it was aimed to find the most appropriate model and technique for the probability of default. Logistic regression, logistic regression with transformed variables, support vector machine and random forest methods have been performed. As a result, random forest and transformed logistic regression gives better results.

In the analysis conducted during the credit scoring, one of the most important Bank's data was used within the framework of the acceptances signed within the scope of data privacy and usage. According to the confidentiality agreement, only the results of the analysis are included in the study. SAS Enterprise Miner, SAS Enterprise Guide and R studio programs were used for modelling and analysis.

Key Words: Credit scoring, machine learning, support vector machines, random forest, logistic regression.

Algebraic Methods for the Problem of Fitting Concentric Ellipses to Digitized Observations

Ali Al-Sharadqah¹, Lorenzo Rulli²

^{1,2} Department of Mathematics, California State University Northridge, Northridge, California, U.S.A.

Abstract

Various objects appear concentric in images, such as a compact disk, the iris of human's eye. Therefore, the problem of fitting coupled geometric objects to digitized data plays a vital role in biometric securities and industrial robotics. There are two common approaches to fitting geometric shapes to data: the geometric (iterative) method and the algebraic (non-iterative) methods. The geometric method is a nonlinear iterative method that can be obtained by minimizing the sum of the squares of Euclidean distances of the observed points to the ellipses. The geometric method is regarded as the most accurate method, but it is computationally expensive and needs a good initial guess in order to have a better convergence rate. Each algebraic method minimizes the sum of the squares of the algebraic distances with a constraint imposed on parametric space.

While fitting a single ellipse to data has been studied intensively in the literature, the research on the problem of fitting coupled geometric objects is still immature. Therefore, we study here the statistical properties of the existing methods for this problem by establishing general mathematical and statistical frameworks for this problem. Due to the nature of this geometric problem, the coordinates of each true (but unobservable) point must satisfy a functional relationship, therefore the problem belongs to the functional Errors-in-Variables models. Since geometric estimation plays a vital role in image processing, our adopted statistical assumptions are tailored to those applications. That is, we use the small-sigma model to study the asymptotic behavior of estimators. We also prefer *geometrically consistent* estimators over a consistent one. Using our rigorous perturbation analysis, we derive the theoretical accuracy of each estimator, and as such, we compare their performances based on their variances and their biases. We also propose unbiased estimators that outperform all existing methods. Both real and synthetic data are used to validate our methodology and the outperformance of our methods.

Key Words: Geometric estimation, geometrically consistent estimators, statistical efficiency, KCR, unbiased estimators.

¹ Email: ali.alsharadqah@csun.edu.

² Email: lorenzo.rulli.817@my.csun.edu.

Two Way Mixed Design: Symmetry Assumption and Missing Data

Sevgi Demircioğlu

Department of Mathematics, Çanakkale Onsekiz Mart University, Çanakkale, Turkey

Abstract

In the two way mixed effects model the assumption of independence of the random main and interaction effects is called the symmetry assumption. When it is violated the sum of square for the fixed main effects is not distributed as constant times chi-squared variable. Therefore, the F-test for the hypothesis of no fixed effects is not valid. In this case Hotelling's T^2 test is suggested.

In violating the symmetry assumption and having two step monotone missing data we show how to test the hypothesis of no fixed effects.

Key Words: Symmetry assumption, two step monotone missing data, Hotelling's T^2 test, chi-squared approximation.

Approximation of Continuous Random Variables for Evaluating Reliability of Complex Stress-Strength Models

Alessandro Barbiero¹

¹Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, Milan, Italy

Abstract

The term “stress–strength model” usually refers to a component (or system) with an intrinsic random strength that is subject to a random stress during its functioning, so that it works only when the strength is greater than the stress. The probability of this event occurring is denominated as the reliability parameter.

If the distributions of both strength and stress, usually modeled as independent continuous random variables, are known, the reliability parameter can be determined by using ordinary transformation techniques; however, this approach becomes cumbersome, if not impossible to apply, when stress and strength depend upon several stochastic factors through a known but complex functional relationship. In this case, the reliability parameter can be computed merely by Monte Carlo simulation, which can require a lot of computation time. Approximation-by-discretization techniques represent an alternative and pragmatic solution to this problem, which drastically cuts down computation time while retaining a satisfactory degree of accuracy.

In this work, we review and adapt several discretization techniques for continuous random variables that have been proposed so far in the literature and apply them to the computation of the reliability parameter of stress-strength models, describing their features and comparing their performances also with reference to practical problems.

Key Words: Approximation, discretization, Monte Carlo simulation, reliability parameter.

Adaptation of the r-d Class Prediction Method to Linear Mixed Models

Özge Kuran

Department of Statistics, Dicle University, Diyarbakır, Turkey

Abstract

Linear mixed models (LMMs) are statistical models containing both fixed and random effects and so, they provide flexibility in fitting models with various combinations of fixed and random effects. These models are often used to analyze data in a broad spectrum of areas including clustered data such as longitudinal data and multilevel data.

To estimate the unknown parameters of the LMMs, Henderson et al. (1959) developed the Henderson's predictors which are respectively named as the best linear unbiased estimator (BLUE) of fixed effects and the best linear unbiased predictor (BLUP) of random effects.

Generally, the variables of design matrix for fixed effects are assumed as linearly independent. However, in practice, we may encounter with strong or near to strong linear dependencies between the variables of design matrix for fixed effects. Then, the problem of multicollinearity is said to exist.

In the existence of multicollinearity, the variances of BLUE may be quite large and BLUE may be far away from its true value. As a result of this case, the Henderson's predictors may be unreliable.

The oldest techniques used to combat multicollinearity between the parameter vector of fixed effects are the principal components regression (PCR) (see Massy (1965)) and Liu estimation (see Liu (1993)).

Kaçıranlar and Sakallıoğlu (2001) combined the PCR estimator and the Liu estimator of Liu (1993) and introduced the r-d class estimator in the linear regression models. And then, Özkale and Kaçıranlar (2007) did the r-d class estimator comparisons in the sense of the matrix mean square error (MMSE) criterion.

Our primary aim in this study is to widen Kaçıranlar and Sakallıoğlu (2001) and Özkale and Kaçıranlar (2007) ideas in the linear regression models to LMMs and we suggest the r-d class predictors in LMMs. Superiorities of the linear combinations of the predictors are done in the sense of MMSE criterion. Numerical examples and a simulation study are done to illustrate the findings.

Key Words: Multicollinearity, best linear unbiased predictor, principal components regression predictor, Liu predictor, r-d class predictor.

Multi Criteria Group Decision Making with Neutrosophic Sets

Ayşegül Tuş¹, Gülin Zeynep Öztaş¹, Abdullah Özçil¹, Tayfun Öztaş¹, Esra Aytaç Adalı¹

¹Business Administration, Pamukkale University, Denizli, TURKEY

Abstract

Real life problems involve uncertainty and fuzzy set theory is a tool to deal with uncertainty. Since, the membership degree of a given element in a fuzzy set does not express the corresponding degree of non-membership, fuzzy sets remain incapable to represent uncertain data. However, in intuitionistic fuzzy sets the degree of the uncertainty depends on the degrees of membership and non-membership. The decision making process depends not only on information that is either true or false, but it also depends on indeterminate information, which is represented by an ignorance value between truth and falsehood. Neutrosophic sets deal with the indeterminacy which expresses the percentage of unknown parameters involving with the problem. In fact the neutrosophic sets are a generalization of the “fuzzy” and “intuitionistic fuzzy” sets. So the aim of this paper is to show the applicability of neutrosophic sets for a real case multi criteria group decision making problem.

Key Words: uncertainty, neutrosophic sets, group decision making.

The Relationship Between the Stock of Car Loans and the Growth of the Automotive Sector from a Times Series Analysis

Burcu Barlas¹,

¹Banking and Finance, Süleyman Demirel University, Isparta, Turkey

Abstract

In Turkey, the automotive sector is seen as one of the most rapidly growing sector. Among all sectors, the automotive industry has been in the first place in exports for 12 years. According to 2016 year-end data, the share of automotive industry production in manufacturing industry production is 8% and its share in GDP is 3%. Turkey constitutes 5.5% of total employment in the automotive sector employees. The share of direct and direct employment of the automotive sector in industrial employment is 15%. Every 13 seconds a car is manufactured in Turkey. The share of the automotive sector in Turkey is 5.6% of total tax revenues.

This study aims to determine whether a causal relationship between the stock of the automobile loans with the growth of the automotive sector.

Automobile loans are loans used to finance purchases of vehicles. Low interest rates on the market to make automobile bills attractive to customers who do not intend to buy automobiles or do not want to renew their vehicles. In other words, demand for vehicle purchases is expected to increase in environments with low credit interest rates. Therefore, when determining production strategies for automobile safety, automobile loans may be a variable in this regard, among other factors. Elsewhere, there's an environment that can be managed from there, and we'll be able to direct in this area. On the other hand, the macro-prudential policies can be implemented in Turkey also argued over the volume of loans might affect the demand for automobiles in Turkey

This study will mention in the automotive sector in Turkey and the cointegration issues with non-stationary time series. By using monthly data for the years 2008-2018, the stock of the automobile loans, production of cars and the number of motor vehicles will be trying to determine whether there is a correlation or long-term causal relationship between them.

Key Words: Banking sector, car loans, automobile loans, automotive sector, time series analysis, causality analysis.

¹ burcu.barlas@vakifbank.com.tr

Neutrosophic ARAS Method for Selection Problems

Esra Aytaç Adalı¹, Tayfun Öztaş¹, Abdullah Özçil¹, Gülin Zeynep Öztaş¹, Ayşegül Tuş¹,
¹Business Administration, Pamukkale University, Denizli, TURKEY

Abstract

Multi Criteria Decision Making (MCDM) contains different methods which are useful in almost all problems related with decision making. These problems deal with the evaluations of a set of alternatives in terms of a set of decision criteria which are expressed both in objective and subjective ways. In addition, it is difficult to measure qualitative criteria in a quantitative way, resulting incompleteness and uncertainty in data. There are numerous tools to handle the uncertainty such as Fuzzy Set (FS), Intuitionistic Fuzzy Set (IFS), Neutrosophic Set (NS), and so on. In this paper the selection problem is solved with ARAS method, one of the MCDM methods, by using the neutrosophic sets. It is concluded that the Neutrosophic ARAS method is preferable for dealing with incomplete and inconsistent information in MCDM problems.

Key Words: Multi criteria decision making, ARAS method, neutrosophic set, selection problem.

A New Heuristic Algorithm for the Label Printing Problem

Emre Çankaya¹, Ali Ekici¹, Okan Örsan Özener¹

¹Department of Industrial Engineering, Ozyegin University, Istanbul, Turkey

Abstract

In this paper, we study the Label Printing Problem (LPP) which has applications in the printing industry. In LPP, the demand for a set of labels is satisfied by printing the labels using templates with multiple slots. Given a fixed number of templates, the decisions in LPP are determining (i) the assignment of labels to the slots of the templates (which we call template designs), and (ii) the number of prints made using each template design. The objective is to satisfy the demand with minimum waste. We consider two variants of LPP where (i) each label can be assigned to the slot(s) of a single template, and (ii) each label can be assigned to the slot(s) of multiple templates. To address LPP, we propose a novel sampling-based construct-improve heuristic where we first generate “good” template designs and then choose the ones to be used and determine the number of prints made through a set covering-type mathematical model. Then, we improve the solution using some improvement ideas that utilize a strengthened linear integer model for the problem. Using the instances from the literature, we show that the proposed heuristic provides better results compared to the benchmark algorithms. We also find optimal solutions for some of the instances from the literature using the strengthened linear integer model.

Key Words: label printing problem; construct-improve heuristic; set covering; mixed-integer linear model

¹ Department of Industrial Engineering, Ozyegin University, Istanbul, Turkey

Curve Fitting for Environmental Measurements by Artificial Intelligence Techniques

Ayşe Nurcan Kapucu¹, Ceyhun Kapucu², Oğuz Akpolat³

¹Muğla Sıtkı Koçman University, Graduate School of Natural and Applied Sciences – Ecological Sciences, Muğla, Turkey

²Muğla Sıtkı Koçman University, Department of Enformatics, Muğla, Turkey

³Muğla Sıtkı Koçman University, Faculty of Science - Chemistry Department, Muğla, Turkey

Abstract

The aim of this paper was to create a prediction model for solar irradiation using open sourced meteorological data. A feed-forward neural network with back propagation was employed to build the predictive model. A large combination of parameters including learning algorithms, transfer functions, and different network topologies were used to customize the neural network. The data is a part of publicly available real outdoor measurement data from National Renewable Energy Laboratory (NREL). The validation of the model has been verified by defining predictive errors using normalized mean squared error (NMSE) and defining prediction accuracy using correlation coefficient (R) metrics. The R value of this model was 92.03%. The worst model within the parameter combinations has showed 80.82% prediction accuracy. The results showed that the distribution of neurons in the hidden layers is an important factor among other parameters. As the R value of the model with the worst parameter combinations was the prediction accuracy of 80.82%, that of the model with the best ones was 92.03%. It could be understood that the most important factor among the parameters to affect the prediction accuracy would be the distribution of neurons in the hidden layers.

Key Words: Artificial intelligence, solar irradiation, curve fitting.

¹ nk369@hotmail.com

² ckapucu@gmail.com

³ oakpolat@mu.edu.tr

Estimating Optimal Value for the Shrinkage and Biasing Parameters in Liu-Type Logistic Regression based on Particle Swarm Optimization

Nuriye Sancar¹, Deniz Inan²

¹Department of Mathematics, Near East University, Nicosia, North Cyprus

²Department of Statistics, Marmara University, Istanbul, Turkey

Abstract

Binary logistic model is one of the commonly used regression models in many fields of science, which expresses the relationship between a dichotomous response variable and one or more covariates. However, in the presence of multicollinearity between the independent variables, ML estimation of the regression coefficients will be unstable and imprecise with inflated standard errors of the maximum likelihood estimator. As a remedy to the negative consequences of this problem, Liu-type logistic estimator, $\hat{\beta}_{k,d}$ was defined as an alternative to the ridge logistic estimator by Inan and Erdogan (2013)¹. This biased estimator has two parameters: k, which is a shrinkage parameter and d, which is a biasing parameter. These parameters, (k,d) are obtained in two steps in the existing works. The biasing parameter d is obtained after selecting the shrinkage parameter k, such that the mean squared error (MSE) of $\hat{\beta}_{k,d}$ is minimized. However, the optimum value of the parameter d depends on unknown parameters because it is not possible to compute the MSE of $\hat{\beta}_{k,d}$ numerically. Different estimators of the parameters are utilized to estimate the optimum value of d. Thus, the choice of the parameters (k,d) in this way is not optimal. In this study, a novel simultaneous method based on the particle swarm optimization algorithm is developed to estimate the parameters (k,d) of the Liu-type logistic estimator. To aim this, an objective function that solves the multicollinearity problem and minimizes the bias of the model, as well as improves predictive performance of the model is built up. The performance of the proposed method is shown with a simulation study and real data set application.

Key Words: Liu-type logistic estimator, multicollinearity, maximum likelihood estimator, ridge logistic estimator, particle swarm optimization.

¹ Inan, D., and Erdogan, B. E. (2013). Liu-type logistic estimator. Communications in Statistics-Simulation and Computation 42, 1578-1586.

A Traffic Accident Prediction Model Based on ANN

Beyzanur Cayir Ervural¹,

¹ Department of Industrial Engineering, Konya Food and Agriculture University, Konya, Turkey

Abstract

Due to the increasing population and economic developments in the country, the number of vehicles causing numerous traffic accidents has increased as well as traffic congestion on the road network. Despite many measures, traffic accidents still remain at a serious level.

In road safety planning and policies, accident prediction models are used to obtain a realistic, guiding and accurate model. According to the prediction results, road safety measures can be taken extensively and some technical efforts are expected to prevent major accidents. In recent years, artificial neural networks (ANNs) have been widely used in accident modelling and road safety studies.

In this study, traffic accidents prediction model for Turkey over the period of 1990-2018, considering the following determinants of traffic accidents was investigated using neural networks: population, road lengths, number of vehicles and the number of accidents.

Key Words: Traffic accident, prediction, artificial neural networks.

¹ Corresponding author

Improvement of Official Statistics by Applying the Concept of Big Data - Case Study: Serbia

Tijana Comic¹, Dragan Vukmirovic², Aleksandar Djokovic³

¹GOPA mbH, STAT, Bad Homburg, Germany

²Faculty of Organizational Science, Belgrade University, Belgrade, Serbia

³Faculty of Organizational Science, Belgrade University, Belgrade, Serbia

Abstract

The rapid development of information and communication technologies, the expansion of the Internet and social networks has led to an explosion of information - the appearance of large amounts of data that are practically available to everyone, indicating the necessity of introducing innovations in the production processes of official statistics. Most of these data are dispersed on global network without order and structure, and the challenge that researchers are facing with is to collect and process those data in a proper way. The central problem that is being addressed in this paper is exploring possibilities of improving the system of official statistics using Big Data, through the application of new methods and research techniques, which facilitates the process of analyzing large amounts of data and improves the overall performance of the official statistics system.

The key scientific contributions of the paper are reflected in the improvement of the methodological procedure of the statistical research process and the formal description of the models and methods that enable their application. The proposed Big Data model is flexible, expandable, provides good performance, and allows the integration of various data source that include non-standard data formats as well. For the purposes of testing the applicability of the model on the territory of the Republic of Serbia, a research was carried out that confirmed the initial assumptions, which created the conditions for the development and implementation of the model.

Key Words: Official statistics, big data, statistical research, model, indicators, sustainable development.

Circular Analyses of Dates on Patients with Gastric Carcinoma

Karaibrahimoglu, Adnan¹, Ayhan, Seren²

¹ Asst. Prof., Biostatistics & Medical Informatics Dept, Faculty of Medicine, Suleyman Demirel University, Isparta, TURKEY

² Post-graduate student, Biostatistics & Medical Informatics Dept, Faculty of Medicine, Suleyman Demirel University, Isparta, TURKEY (Presenter)

Abstract

Although circular data analysis has been put forward many years ago, its importance has recently started to appear. Especially in the environmental, geographical and biological sciences began to be used frequently. Although it is not widely used in the field of health, some new studies are encountered in the literature. Dates have great importance in cancer diseases. Survival analyses are performed according to dates such as diagnosis, operation, follow-up and progression of stages. However, the date variables themselves are not analyzed. The aim of this study is to evaluate the descriptive statistics of diagnosis, operation and final follow-up dates in gastric carcinoma patients by circular analysis methods. This is a retrospective study. In the dataset, there are many variables, but only four variables – dates and gender- were considered to analyze in the dataset. The analyses were performed by RStudio programming. Totally 502 gastric carcinoma patients admitted to a university hospital between the years 2000-2011 were enrolled in the study. Of the patients, the ratio of women was 33.5%. The mean month of diagnosis date was found nearly November (~10.86) for female and May (~5.17) for male patients. The mean month of operation date was found March (~3.24) for females, and July & August (~7.79) for males. The mean month of final follow-up date was found as February & March (~2.61) for females, and May (~4.85) for males. Moreover, the mean day of the week for diagnosis date was found Thursday (~5.50) for both female and male patients. However, the mean days of operation date were different between genders, and found as Monday (~2.47) for females and Tuesday (~3.03) for males. Similarly, the mean days of final follow-up dates were different for genders (Monday for females and Wednesday & Thursday for males). The fitting of distributions of all variables were checked, also, according to von Mises, Rayleigh and Kuiper's tests. When the days and months were analyzed by classical descriptive statistics, the results were obtained completely different from the circular analyses results. Therefore, the dates and times should be analyzed in certain diseases to give an idea for physicians. It can be suggested that circular data, which indicates direction, angle, date or time should be taken into consideration in health sciences.

Key words: Circular data, gastric carcinoma, von Mises distribution.

Importance of Fractal Structures in Modern Biology and Discussion of Forward Kolmogorov Equation with Fractional Derivative Approach

Esra Yazıcıoğlu^{1,3}, Özgür Gültekin^{2,3}

¹Department of Biology, Istanbul University, Istanbul, Turkey

²Department of Physics, Istanbul University, Istanbul, Turkey

³Department of Statistics, Mimar Sinan Fine Arts University, Istanbul, Turkey

Abstract

Euclidean geometry does not completely describe nature at the present time. Countless physical and biological phenomenon in nature such as clouds, coastlines, mountain formations, molecular structures of crystals, plants can be described more accurately by fractal geometry. Fractal structures, basically formed by repetition and show features such as full self-similarity, semi-self-similarity, statistical self-similarity and complex geometric objects with fractional dimensions. Mathematically, a fractal can be defined as a set in which the Hausdorff dimension exactly exceeds the topological dimension.

Living systems exhibit fractal characteristics from micro-scale to organism size and even to collective lives. Therefore, we can say that modern biology research is directed towards fractal structures in the modeling of biological processes and beings at both molecular and system biology levels. However, the importance of statistical understanding in the formalization of natural laws increases the interest of natural scientists in statistical and stochastic processes. For example, diffusion is a very important process for biology. On the other hand, the normal distribution models of diffusion are not universal. Biological phenomena can be modeled more realistically with “abnormal convection” events that do not fit the Gaussian distribution, where fractal properties are also important. One of the mathematical tools used in modeling diffusion is the Fokker-Planck equation, also known as the advanced Kolmogorov equation. The Fokker-Planck equation is a partial differential equation which gives time evolution of the probability density of a system under the influence of stochastic behavior.

In this study, firstly the basic properties of fractal structures, their formation by random processes and their current use in biology will be discussed. We propose to use the fractional Fokker-Planck equation, which we have written using the fractional derivative approach, in the modeling of systems with fractal structure and compare the solutions of the fractional Fokker-Planck equation with the solutions of the classical Fokker-Planck equation. We fundamentally argue that the use of fractional equations in simulations of natural phenomena is not a mathematical exercise but also a more realistic approach to understanding deviations from normality.

Key Words: Fractal, Fokker-Planck equation, fractional Fokker-Planck equation, biological system.

Examination of Incidences of HIV Infection and Tuberculosis in Some Selected European Countries with Statistical Process Control

O Toluk¹, I Ercan¹, H Akalin²

¹Department of Biostatistics, Faculty of Medicine, Uludag University, Bursa, Turkey

²Department of Infectious Diseases and Clinical Microbiology, Faculty of Medicine, Uludag University, Bursa, Turkey

Abstract

Statistical process control (SPC) is a method of analysis developed for examining the data of a process and which facilitates understanding of that process. By monitoring the course of diseases over time with SPC charts, upward and downward trends can be determined, and as a result of this, control efforts can be given direction.

The HIV infection is one of the most important known risk factors for the disease of tuberculosis. The risk of developing tuberculosis disease is 20-30 times greater compared to HIV negative individuals.

The aim of this study is to investigate whether or not the Human Immunodeficiency Virus (HIV) infection and Tuberculosis (TB) were brought under control in some selected European countries between the years 1990-2016, as well as to examine the association between HIV and TB in the years when they went out of control. Moreover, the relationship between these diseases and selected The Organisation for Economic Co-operation and Development (OECD) criteria will be examined.

In our study, within the scope of which the time course of HIV and TB in some selected European countries was examined with SPC based on years, the HIV and TB incidence data on the World Health Organization (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS) databases between the years 1990 and 2016 were utilised. The mid-year population figures of the European countries were taken from the WHO database. The European countries included in the study were selected based on the World Health Organisation's European Region classification.

In our study, we determined that HIV and TB have been brought under control in Western European countries like France, Britain, Germany, Spain and Italy. However, in Central European countries like Hungary, Poland and Turkey, TB is under control, whereas HIV is out of control. In Ukraine, which represents the Eastern European countries, both infections are out of control.

When an assessment is made in terms of factors that affect bringing HIV under control, it is striking that there are significant epidemiological and HIV Treatment Cascade differences between countries. When policies are being created especially for bringing the HIV infection under control, the policies of countries which have brought this infection under control also need to be examined.

Key Words: Statistical process control (SPC), HIV incidence, Tuberculosis (TB) incidence, Laney P control chart.

A Two-Population Model with Renewal Process for Measuring Longevity Basis Risk

Selin Özen¹, Şule Şahin²

¹Department of Actuarial Sciences and Risk Management, Karabuk University, Karabuk

²Institute for Financial and Actuarial Mathematics, University of Liverpool, Liverpool & Actuarial Sciences, Hacettepe University, Ankara

Abstract

Basis risk is a serious concern when hedging longevity risk with index-based instruments. Basis risk arises from demographic mismatches between the portfolio's specific population and reference population associated with the hedging instrument. Although several multi-population models have been developed for assessing basis risk, only a few incorporates the mortality jumps. The papers on mortality jump modelling are used the Poisson process for their jump frequencies. Due to their low probability and high impact nature, the timing and the frequency of these jumps are unknown, but the history of the events can give information about their future occurrences. In this paper, we assume that the mean time between their arrivals no longer constant and a new approach to modelling jumps by using the renewal process is proposed. The proposed model is used for the reference population. Then we investigate the specific population's mortality dynamics and we find that cohort and period effects are important for the specific population. Our aim is to develop a more realistic framework for assessing basis risk. England and Wales males and Continuous Mortality Investigation assurance male lives data are used for illustrations. We find that the proposed model fits the historical data better.

Key Words: Longevity risk, basis risk, renewal theory, mortality jumps.

The Effects of Financial Innovation on Economic Growth in The Banking Sector: Case of Regions and Provinces

Aynur Karaçoban¹, Ömür Saltık², Süleyman Değirmen³
^{1,2,3} Mersin Üniversitesi, İktisadi İdari Bilimler Fakültesi

Abstract

The study focuses on the banking sector which has an important position for the financial markets to highlight the relationship between financial development and growth by being analyzed on the basis of regions and provinces. In parallel with the development of technology, banks focus on innovative technology derived products and services. In this context, the study demonstrates the relationship between financial innovation for the banking sector and economic growth for Statistical Region Units Level-1 (12 regions) and Turkey's 81 provinces. In the study, yearly data between 2010 and 2017 were used. First of all, a new innovation index was created by using Principal Component Analysis method because of the excess amount of innovation variables and to represent the related variables in as few dimensions as possible. Then, as one of the dynamic panel data methods, Arellano-Bond GMM first differences estimation method and Pooled Data method were applied as model. It was found that the coefficient of financial innovation variable for the banking sector was found significant and positive for the regions and provinces in the GMM model. Pooled data analysis covers both GDP level and GDP growth ratio as dependent variable. According to the results, the coefficient of the innovation variable, when the dependent variable GDP as considered in level, was found as significant in 8 regions, for 51 provinces, and positive. According to GDP ratio results; Its coefficients were found to be significant in 4 regions, significant in 31 provinces and positive. These results support that the innovation-based products and services under the leadership of technological developments, globalization, and information offer positive contributions to the economy in the banking sector for some regions and provinces in Turkey. These empirical findings confirm Schumpeter's idea and endogenous growth theories.

Key Words: Financial innovation, banking sector, regions-provinces, dynamic panel GMM method, pooled data method.

¹ aynurkaracoban44@gmail.com

² omursaltik09@gmail.com

³ suleymandegirmen@gmail.com



Net External Position, Financial Development, and Banking Crisis

Aytül Ganioglu¹

¹Structural Economic Research Department, Central Bank of the Republic of Turkey, Ankara, Turkey

Abstract

Does the external position of a country that is conditioned on financial development impact the likelihood of a systemic banking crisis? We address this question using data from 149 developing and advanced countries from 1970 to 2011, as well as a variety of statistical tools. Our findings are twofold. First, we find that the net external position of a country significantly affects its likelihood of a systemic crisis depending on the level of financial development. Conditional on low to moderate financial development, countries can lower the risk of banking crises significantly by maintaining a net foreign creditor status. Second, we find that the level of financial development raises a country's crisis risk significantly while its impact depends on the net asset position. This indicates a potential amplification effect in which countries with more developed and complex financial systems that are also debtor countries have a higher potential of incurring a systemic banking crisis.

Key Words: Banking crisis, net external position, financial development, probit.

¹ E-mail: aytulganioglu1@gmail.com

The Gravity Model of International Immigration: Evidence from OECD Countries

Ali Acaravcı¹, Seyfettin Artan², Cemalettin Kalaycı³, Sinan Erdoğan⁴

^{1,4} Economics, Hatay Mustafa Kemal University, Hatay, Turkey

² Economics, Karadeniz Technical University, Trabzon, Turkey

³ Labour Economics and Industrial Relations, Karadeniz Technical University, Trabzon, Turkey

Abstract

The term of migration that has started with history of humanity, have been debated both in politics and economics and has security, sociologic and institutional aspects. The migration movement that aims to reach developed or developing countries has went on at low density since the past but has gained momentum after the mass protest movements in the Middle East and North Africa. This study investigates the international migration in 22 Organization for Economic Co-operation and Development (OECD) origin and destination countries for 2005–2016 by using gravity model. Empirical results show that difference of per capita real income between destination and origin countries, contiguous and common language raise international migration flows. On the other hand, difference of population and distance between destination and origin countries are associated with a have negative effect. The evaluation of economic, social and institutional determinants of international migration will provide better understanding of migration and design well policy proposals for the problems associated with migration.

Key Words: International migration, gravity model, cross-section, panel, ols, ppml.

¹ This work was financially supported from the Mustafa Kemal University, Scientific Research Projects Coordination Office (Project Number: 18.M.022).

A Promotion Planning Model with Product Substitution and Promotion Fatigue

Dilek Tüzün Aksu¹, Bahadır Durak², Buse Dedeakayoğulları³

¹ Department of Industrial and Systems Engineering, Yeditepe University, Istanbul, Turkey

² Research & Development Center, OBASE, Istanbul, Turkey

³ Research & Development Center, OBASE, Istanbul, Turkey

Abstract

In a competitive environment, retailers make strategic pricing decisions to maximize their profits. One of the important decisions that affect purchasing behaviour of customers is price promotions. Deciding the magnitude, timing and frequency of promotions are the main challenges of promotion planning. Discounted prices may increase the demand of a particular product; on the other hand, high price levels may lead the customer to consume substitute products with lower prices. Therefore, demand of a particular product is affected not only by that product's own price, but also by the price of its substitutes. Price promotions are widely used in the retail industry to retain customers and increase demand. However, a promotion may also cause the customers to stockpile the product at discounted prices and reduce the demand for it in the following periods, giving rise to a phenomenon called *promotion fatigue*. Consequently, the effectiveness of a promotion campaign for a product highly depends on its timing, as well as the coordination with other campaigns for substitute products. We propose a methodology to coordinate the promotions of a group of substitute products sold by a retailer over a planning horizon such that the total profit over the horizon is maximized.

Our solution methodology consists of two stages. First, we use a log-linear regression model to estimate the demand of each product as a function of its own price and its interaction with the prices of substitute products. We also include the previous prices of the product in the model to account for the impact of promotion fatigue. Other factors such as trend, seasonality and special occasions are also taken into consideration. Next, we generate a promotion plan that maximizes the total revenue based on the demand estimates from the first stage. This plan provides the timing and the magnitude of price discounts for all products in the group over the specified planning horizon. Since concurrent promotions of two substitute products result in a reduced benefit, we assume that at most one product can be offered at a reduced price in each period. This assumption is in line with the general practice in the retail industry.

We conduct experiments on real data to test performance of the proposed methodology. Computational results demonstrate that our approach produces high quality solutions within reasonable time and has the potential to be implemented in practical settings to create a competitive advantage for the retailers.

Key Words: Promotion optimization, linear regression, retail operations, promotion fatigue, promotion planning, product substitution, pricing, revenue maximization.

Human Identification with Deep LSTM by Using Biometric Walking Patterns

Çağatay Berke Erdaş¹, Didem Ölçer¹

¹Başkent University, Department of Computer Engineering, Ankara, Turkey

Abstract

In recent years, recognition of the identity of the person in the machine recognition literature has briefly increased the popularity of the recognition area. Most relevant identification studies use feature extraction to achieve higher classification performance. Moreover, these studies often use traditional machine learning algorithms for classification. In this study, we focus on the deep Long Short-Term Memory (LSTM) neural network for the classification. Deep LSTM is a version of LSTM that has multiple hidden LSTM layers. An alternative solution that works faster and requires less neuron can be achieved by increasing the depth of the network [1].

Used dataset are collected from 22 participants walking in nature on a predefined path by using three-dimensional accelerometers [2]. Accuracy, sensitivity, specificity and F-measure criteria were used as the classification performance criteria and a 10-fold cross-validation approach was used for the evaluation of the study.

Based on the results, it can be assumed that the proposed deep LSTM approach can classify raw data with high performance.

Key Words: Human identification, LSTM, deep LSTM.

References

- [1] Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 6645-6649). IEEE.
- [2] Casale, P., Pujol, O., & Radeva, P. (2012). Personalization and user verification in wearable systems using biometric walking patterns. Personal and Ubiquitous Computing, 16(5), 563-580.

Different Approaches to Reducing Bias in Classification of Medical Data by Ensemble Learning Methods

Adem Doğaner

Department of Biostatistics and Medical Informatics, Kahramanmaraş Sutcu Imam University, Kahramanmaraş, Turkey

Abstract

Thanks to machine learning methods and development of deep learning, information extraction and classification in large data sets have started to yield remarkably successful results. Although a single classifier displays a high performance depending on the development of algorithms, it may still encounter bias and variance problems during learning process, which eventually results in overfitting and underfitting problems. One of the methods developed for the improvement of classification performance and the elimination of bias and variance problems is ensemble learning. It mainly aims at classifying data using multiple classifiers rather than a single classifier in order to obtain a higher classification performance. In addition to using multiple classifiers, ensemble learning also involves different methods such as boosting, bagging and stacking which will reduce bias and variance.

In the present study, ensemble learning methods were used in order to classify patient and control groups with a high performance and low bias. Different classifiers were included in the ensemble learning model to obtain the highest performance and lowest bias error. Hemogram values were used in order to classify patient and control groups. Additionally, different sample sizes were used in order to measure the effect of the sample size on the classifier performances. Local outlier factor (LOF) was used to find out outliers in the data. Grid Search algorithm was used for parameter optimization. Standardization (z-transformation) was applied to the data. Accuracy, sensitivity, specificity and AUC were used for performance assessment.

In conclusion, it was found out in the present study that classification performance obtained from ensemble learning methods was higher compared to methods using a single classifier. On the other hand, it was observed that the increasing number of samples did not contribute to the classification performance and reduction of bias error after a certain point. As for reduction of bias, methods based on boosting method displayed a higher performance compared to other methods.

Key Words: Machine learning, deep learning, ensemble learning, boosting.

Reduce Unexpected Airline Diverts: Modelling with Neural Network

Hazal Berve Doğan¹, Derya Gözen¹, Hakan Maral², Tahir Hanalioğlu³

¹STM - Savunma Teknolojileri Mühendislik ve Ticaret A.Ş., Ankara, Turkey

²ASELSAN A.Ş., Ankara, Turkey

³TOBB Ekonomi ve Teknoloji Üniversitesi, Ankara, Turkey

Abstract

In this study, a decision support system is designed in order to minimize the number of flights that are diverted unexpectedly. The aim is to reduce the expenses that arise when the aircraft is not able to land on the targeted airport due to the unfavorable weather conditions, such as rescheduling the timetable, overuse of aircraft fuel than planned, passengers' accommodation and ticket reissue. In order to reduce such temporal and financial losses caused by diverted flights, decision to take off or not is made before departure, while the decision to land or not is made during flight, after a brief analysis based on weather data of target airport. For the aircraft to land on target airport as scheduled, it is crucial that the weather forecasts for visibility range, ceiling and wind speed are within the limits of the safe flight requirements. Considering the significance of this decision regarding by finance, there is a need for a decision support system that is capable of boosting the process through optimal decision-making by forecasting airport weather conditions with using Neural Network. In the first step, attribute selection was made by examining the data set. The feature with the highest effect (> 90%) in the whole data set was determined as low visibility range by principal component analysis (PCA). As the data set is meteorological data, short and long term memory (LSTM) network structure, which is a kind of repetitive neural networks (RNN), In the next step we worked on two different scenarios in LSTM network. These scenarios are; in order to identify the next step, one step shift and five step shifts to represent the steps preceding the step to be detected. In both scenarios, it is aimed to estimate the future data for 3 hours. The network structure parameters were kept the same for both scenarios and the number of epoch was set to 200. When the results for the two scenarios were compared, it was observed that both scenarios could follow the visibility range prediction trend. When the RMSE values were examined, it was observed that the RMSE value of the five-step shift scenario was 0.018 while the RMSE value of the one-step shift scenario was 0.063. If it is predicted that the RMSE value will decrease when the number of epoch is increased, the number of cycles is kept constant and the confusion matrix of the five step shift scenario is extracted in order to both decrease the two scenarios in parallel and prevent overfit. Among the 130000 data provided by the airline, there are 84 divers. According to the 5 step shift scenario, it was observed that the decision support system decided to fly for 30 out of 84 flights. This result shows that the decision to fly at the beginning of 30 divers. This means a significant reduction in costs.

Key Words: Divert, weather forecast, neural network, deep learning, rnn, lstm, regression, flight divert, aviation, decision support system.

Financial Distress Prediction using Support Vector Machine with Optimal Parameter and Feature Selection

Seyyide Doğan ¹, Şenol Altan ²

¹ Faculty of Economics and Administrative Sciences, Karamanoğlu Mehmetbey University, Karaman, Turkey

² Faculty of Economics and Administrative Sciences, Ankara Hacı Bayram Veli University, Ankara, Turkey

Abstract

Prediction of financial distress, which affects the whole economic system, is crucial to both macroeconomics and to shareholders, investors, and lenders in the micro-level.

In recent literature, a number of successful financial early warning systems have been developed with machine learning techniques. The aim of this study is to develop a useful and meaningful prediction model with Support Vector Machine (SVM) to evaluate whether Turkish firms face financial distress. Different models have been proposed to increase the classification accuracy and stability of the method. In order to achieve this aim, both SVM parameters have been optimized and feature selection stage has been applied. A Grid-search technique which is an easy and effective tool for parameter optimization and three different statistical techniques were chosen for feature selection. These are independent sample t-test, Principal Component Analysis and Logistic Regression Analysis.

Experimental results reveal that all models with and without feature selection are useful and have a very high prediction accuracy rate. However, the feature selection phase contributes to the success of the SVM. We also compare the performance of these techniques, it was observed that Logistic Regression Analysis and t-test outperforms the other.

Key Words: Financial distress, support vector machine, grid-search, feature selection.

Distribution the DJSTOXXE50 Index

Yasemin Ulu¹

¹University of Michigan, Dearborn

Abstract

In this paper using data from 1999-2005 on 5-minute intraday returns, we construct a model free estimate of the daily realized volatility for the Eurozone DJSTOXXE50 index. We compute the unconditional volatility distribution of the DJSTOXXE50 index by a nonparametric kernel estimation method. Our results indicate that the unconditional volatility distribution of the DJSTOXXE50 returns are leptokurtic and highly skewed to the right. The logarithmic standard deviations seem to be approximately Gaussian. Our results are inline with Andersen, Bollerslev, Diebold Evens (1999) for individual DJIA equity return volatility and Andersen, Bollerslev and Cai (2001) for Japanese index, Nikkei 225.

Key Words: Realized volatility, Kernel density estimation, unconditional density, high-frequency returns.

¹ email: yasemin.ulu@gmail.com

Spatial Statistical Analysis for Geodetic Points of a Simulated Fault Plane through Geographical Information Systems

Özlem Türkşen

Statistics Department, Ankara University, Ankara, Turkey

Abstract

Forecasting of earthquake occurrence time is an important problem all around the world. The spatial statistical analysis of geodetic points helps to analyse the fault plane area and presents prior information to make forecasting with minimum error. In this study, it is aimed to determine the spatial patterns of geodetic points which are measured on a simulated fault plane area by using spatial statistical methods through Geographical Information Systems (GIS). The GIS is a computer system that analyses and displays spatial and geographical data. In the scope of the study, spatial central tendency, spatial pattern analysis and spatial autocorrelation of geodetic points were evaluated. The spatial central tendency of geodetic points was analysed by using spatial descriptive statistics, e.g. spatial mean center, spatial median center and spatial standard deviational ellipse. In order to define the spatial pattern of the geodetic points, the quadrat analysis and the nearest neighbor analysis were performed. And also, spatial autocorrelation of the geodetic points was calculated by using Moran's I index and Anselin Local Moran I statistics. It was seen from the analyses results that the spatial distribution of the geodetic points is uniform according to the results of quadrat analysis and nearest neighbor analysis which take into account only the location information. However, it was examined that the spatial pattern of the geodetic points were clustered according to the spatial autocorrelation methods since the geodetic points of the surface displacement were taken into account. It can be said that applying the GIS provides easiness to define and to visualize the spatial statistics of simulated fault plane geodetic points.

Key Words: Spatial statistical analysis, geographical information systems (GIS), fault plane, geodetic points.

Assessment of Longevity Risk on Pension Funds: Credibility Approach

Bükre Yıldırım Külekci¹, A. Sevtap Selçuk-Kestel²

¹Middle East Technical University, Institute of Applied Mathematics, Ankara/Turkey bukre@metu.edu.tr

²Middle East Technical University, Institute of Applied Mathematics, Ankara/Turkey skestel@metu.edu.tr

Abstract

Over the last 100 years, there has been a steady decline in mortality rates. With the unexpected decrease in death rates; variability in the age of death declined and deaths are concentrated to the older years of life, therefore, causing deficits in the pooled funds of insurance and pension providers, which are built upon the underestimated rates. To correctly measure the effect of mortality rates on the financial stability of these providers, longevity risk should be taken seriously. This study aims to investigate the future mortality and longevity risk of three countries with different age structures. Turkey, Germany, and Japan are chosen with respect to their expected life and population distributions. Lee-Carter mortality model is used on the historical census data to forecast future mortality rates. The longevity risk on a hypothetical portfolio is assessed based on static and dynamic mortality table approaches. To determine the impact of longevity risk retrieved from the use of a stochastic mortality model and a pension product, annuity, is taken into account. The net single premium, variance and the risk measures for the annuity are quantified under the proposed set up for the selected countries. Additionally, a credibility approach is proposed to establish a reliable estimate for the annuity net single premium between the static and dynamic approaches.

Key Words: Longevity risk, Lee-Carter model, Bühlmann credibility, pension, annuity.

The Effects of a Promotion on Store Traffic, Conversion, and Customer Expenditures: Evaluation with a Prediction Approach

Ignacio E. Inostroza-Quezada¹, Leonardo D. Epstein²

¹Rutgers Business School, Rutgers University, Newark, New Jersey, USA

²School of Business and Economics, Universidad de los Andes, Santiago, Chile

Abstract

Store managers conduct promotions to increase at least one of three measures of store performance: arrival traffic, conversion probabilities, and customer expenditures. These are the total sales' components that Lam *et al.* (2001) identify.

The talk presents a new approach to evaluate the effects of a promotion on each of these components: the approach develops a joint 3-variate time series model, where the variables are the arrival counts, the conversion counts and the average expenditure in one-hour time-bands. The model incorporates explicitly the dependence among the three components.

The approach uses data outside the promotion period along with regression models with independent sets of covariates for location and scale parameters, to build counterfactual baseline predictions for each sales component for the promotion period. These baselines assume that the promotion has not occurred. To evaluate the promotion effects on each component, the approach compares the observations to the corresponding baseline predictions during the promotion.

This approach has at least three strengths: first, it does not require building a model for the promotion effect. Second, it controls for covariate effects, such as time of the day, day of the week, etc. Third, it measures the effects of the promotion directly on the quantities of interest, arrival counts, conversion probabilities, and expenditures, and not on the scales of parameters which are often difficult to translate into effects on observables.

An illustration with data from an actual store combines arrival data from processed video images and sales recorded at sales registers.

Key Words: Counterfactual baseline, price promotion, multivariate stochastic processes, store performance.

Nonparametric Statistical Methods for the Comparison of Two Treatments in the Presence of Competing Risks

Jingjing Lyu^{1,#}, Jinbao Chen^{1,#}, Yawen Hou,² Zheng Chen,^{1,*}

¹ Department of Biostatistics, School of Public Health (Guangdong Provincial Key Laboratory of Tropical Disease Research), Southern Medical University, Guangzhou 510515, China.

² Department of Statistics, College of Economics, Jinan University, No.601, West Huangpu Avenue, Guangzhou 510632, China.

Abstract

Competing risks data arise frequently in clinical trials, and a common problem encountered is the overall homogeneity between two groups. In competing risks analysis, when the proportional subdistribution hazard assumption is violated or when two cumulative incidence function (CIF) curves cross, the most common currently used testing methods, e.g., the Gray test and the Pepe and Mori test, may have a significant loss of statistical testing power. In this article, we propose a new testing method based on the absolute difference in the area under the CIF curves. This method captures the difference over the whole time interval for which survival information is available for both groups and is not based on any special assumptions regarding the underlying distributions. This method was also extended to test short-term or long-term effects. We also consider a combined test and a two-stage procedure based on this new method that considers all possible alternatives, and a bootstrap resampling procedure is suggested in practice to approximate its limiting distribution. An extensive series of Monte Carlo simulations is conducted to investigate the power and the type I error rate of the methods. And from our simulations, our proposed ABC, Comb and TS tests have a relatively high power in most situations. Besides, The methods are illustrated using three different datasets, namely, data from a pediatric cancer trial, a malignant melanoma trial and an acute lymphoblastic leukemia trial, that have different situations of CIFs.

Key Words: Competing risks, crossing cumulative incidence function curves, nonproportional subdistribution hazards, nonparametric tests, Monte Carlo simulations, long-term test.

co-first authors

* corresponding author, e-mail: zheng-chen@hotmail.com

Modification of S Test Statistics for Randomized Block Designs with Autocorrelated Longitudinal Data

Melike Kaya Bahçecitapar¹

¹Department of Statistics, Hacettepe University, Ankara, Turkey

Abstract

Page-type tests and generalized Jonckheere-type tests are frequently used nonparametric test statistics for testing ordered alternative hypotheses in randomized complete block designs (RCBD). Modification of S (MS) test has been recently proposed as an alternative to generalized Jonckheere test in RCBD. In this study, Page, generalized jonckheere (GJ) and MS test statistics are adapted for non-normally distributed longitudinal data in randomized block designs which are autocorrelated in different structures. Longitudinal data occur when repeated measures are taken over time on same subjects. MS test statistics is considered as an alternative test to Page and Jonckheere-type test statistics previously examined for randomized block designs with independent measures. In addition, unlike previous studies, MS test statistics is examined for longitudinal data over time in randomized block designs with Laplace and t-distributed errors which are also autocorrelated in compound symmetry, AR1 and first-order moving average (MA1) structures. Simulation-based sample size calculation is also proposed using Monte Carlo simulations for MS, Page and Jonckheere test statistics in randomized block designs under different assumptions for distribution and autocorrelation between longitudinal data over time, which are not having closed forms in the literature to determine sample size of this kind of dependent cases in randomized blocks.

Key Words: Longitudinal data, nonparametric tests, randomized block designs.

Novel Approaches for Choosing Optimum Cut-Off Value

Naime Meric Konar¹

¹Department of Biostatistics and Medical Informatics, Kirsehir Ahi Evran University, Kirsehir, Turkey.

Abstract

ROC Curve Analysis is mainly used for two purposes, namely determining the diagnostic performance of biomarker(s) and obtaining the optimal cut-off value to distinguish diseased and non-diseased groups. Studies regarding cut-off value selection has become increasingly popular in recent years. In this study we aimed at deciding the optimum cut-off value with the help of some well-known statistics including Diagnostic Odds Ratio (DOR), F1-Score, Overall Accuracy, Mean Absolute Deviation (MAD), Median Absolute Deviation (MEAD) and Coefficient of Variation (CV) as well as the clinical utility measures Decision Curve Analysis (DCA) and Net Benefit (NB) metrics. We suggest to choose the optimum threshold value as the one which has the maximum of these aforementioned statistics along with the one having maximum Net Benefit. Moreover we have some suggestions regarding which statistics should be taking into account when determining the optimum cut-off value. Publicly available Mayo Clinic Primary Biliary Cirrhosis Data was used for illustration. Aforementioned statistics were computed in order to determine the optimum threshold value for each biomarker in the study to discriminate dead and transplanted or alive subjects. In overall, this study offers new approaches on choosing the optimal cut-off value when there are threshold alternatives in the study.

Key Words: cut-off, ROC Curve, diagnostic odds ratio, F1-Score.

Prognostic Model Animated Film Festival - VAFI

Lovorka Gotal Dmitrovic¹, Hrvoje Selec¹, Mario Lesina¹

¹University North, University Centre Varazdin, Croatia

Abstract

Purpose: The International Animated Film Festival - VAFI was the hallmark event of the Croatian city Varazdin. The Film-Creative Studio VANIMA Association since 2010 organizes VAFI - an international animated film festival for children and young people. The festival has an emphasis on the filmmaking of children and young people with the aim of encouraging media literacy. Throughout these ten years, VANIMA often has problems with financial resources for organizing the festival. The purpose of this paper is to identify and discuss the factors, with the emphasis on the funding needed for the festival's growth.

Design/methodology/approach: The case of the VAFI festival is analysed using VANIMA's data sources on: festival duration, the number of submitted films, the number of reported films, the number of selected films, number of official projections, the number of other programs, number of participants and budget for 10 years. Using descriptive statistics, parameters were determined, while correlation between parameters was determined using Pearson's and Spearman's correlation coefficient. The cause-effect diagram (Ishikawa diagram) was produced, with the result (effect) of a "successful festival". Based on it, a mathematical and computer model was developed, evaluated by statistical non-parametric tests.

Findings: The VAFI festival is complex system according to system theory. The results of the study reveal that the festival depends upon a combination of three components, these are: number selected films, number of official projections and the number of other programs. Every of the three components are depending on budget.

Originality/value: Very few studies have researched event using modelling as a research method, although the modelling is a recommended priority within the field of festival studies. This study presents a developing predictive model which is adaptive according event and contributes to this area of knowledge and provides relevant information for organizations and host cities.

Key Words: VAFI, animated film festival, prognostic model, VANIMA, modelling, model valuation.

Classification of OECD Countries with Socio-Economic and Education Indicators

Ebru Özgür Güler¹, Dilek Veysikarani¹

¹Department of Econometrics, Cukurova University, Adana, Turkey

Abstract

Research on education reveals that education is the most effective factor in socio-economic development. In other words, the performance of a country in the development process is closely related to the effectiveness of the education system. On the one hand, education fulfills the function of providing labor force with the quality and quantity needed in the development process. On the other hand, it contributes to the follow-up, development and transfer of the countries to modern production technologies with the function of producing and disseminating information. Education, as the driving force in economic and social development, enables countries with well-educated labor, namely human capital, to become more active and develop in the global market.

The purpose of this study is to classify 35 OECD countries, in which Turkey is a member, considering socio-economic as well as PISA indicators, which is regarded as an education level indicator. For this purpose, by applying both hierarchical and non-hierarchical clustering analysis, countries with similar characteristics are grouped and homogeneous subgroups are formed. In addition to that, countries are compared by considering different characteristics and various suggestions related to subgroups are included.

Key Words: PISA, socio-economic development, OECD, clustering analysis.

Theoretical and Practical Comparison of the Frequency Based Classification Methods: Education Sample

Cengiz Gazeloğlu¹, Sinan Saraçlı²

¹Department of Statistics, Suleyman Demirel University, Isparta, Turkey

²Department of Statistics, Afyon Kocatepe University, Afyonkarahisar, Turkey

Abstract

The purpose of this study is frequency based classification methods on both theoretical and real data and to provide the advantages and disadvantages of related methods comparatively. ZeroR, OneR, NaiveBayes and Decision Tree method were used in the study. In Addition, the theoretical basis of these methods is explained in detail. The WEKA program was utilized while analyzes were carried out in the study.

When all the results are evaluated in general, it was found out that it is the decision tree classification method that managed to estimate the grades best in both courses, whether it is a Portuguese language course or a mathematics course. The correct classification ratio of the decision tree classification method for the mathematics lesson and the Portuguese lesson has been computed approximately as 83 %. ZeroR method is the least accurate estimation rate method. This rate is 31 %.

Key Words: Data mining, classification, machine learning, education.

ATAforecasting: An R Package for Automatic Seasonal or Non-seasonal Time Series Forecasting

Hanife Taylan Selamlar¹, Ali Sabri Taylan²

¹Department of Statistics, Dokuz Eylul University, Izmir, Turkey

²The Union of Chambers and Commodity Exchange of Turkey, Ankara, Turkey

Abstract

ATA method is a new univariate time series forecasting method, which provides innovative solutions to issues faced during the initialization and optimization stages of existing methods. ATA forecasting performance is superior to existing methods both in terms of easy implementation and accurate forecasting. It can be applied to non-seasonal or deseasonalized time series, where the deseasonalization can be performed via any preferred decomposition method. The R package “ATAforecasting” was developed as a comprehensive toolkit for automatic time series forecasting. It focuses on modeling all types of time series components with any preferred ATA methods and handling seasonality patterns by utilizing some popular decomposition techniques. The ATAforecasting package provides for researchers modeling seasonality with STL, TBATS, stR, TRAMO/SEATS and Box Cox Transformation and analyzing the any time series with simple ATA and additive, multiplicative, damped trend ATA methods and level fixed ATA trended methods. It offers functions for researchers and data analysts to model any type of time series data sets without requiring specialization. However, an expert user may use the functions that can model all possible time series behaviours. The package also incorporates types of model specifications and their graphs, uses different accuracy measures that surely increase the ATA Method's performance.

Key Words: Time series, forecasting, accuracy, ATA method, STL, TBATS, stR, box-cox transformations.

¹ hanife.taylan@deu.edu.tr.

² alisabritaylan@hotmail.com

Ranking of Turkish Universities Based on Academic Performance by Using Fuzzy c-means Clustering Algorithm

Ayşen Apaydın¹, Fatih Tank¹, Furkan Başer¹

¹Department of Actuarial Sciences, Ankara University, Ankara, Turkey

Abstract

In recent years, based on an increasing worldwide desire for comparative information, students, universities and governments take a great interest in knowing the position that a university has in comparison with other universities. Moreover, in order for ensuring quality in higher education and increasing global competitiveness through various policies and actions there is a remarkable rise in systems for comparing and ranking universities across the world.

In this paper, Turkish universities were clustered and ranked by using the proposed algorithm based on fuzzy c-means clustering. Universities were ranked according to their academic or research performance over the period of 2013–2017, ranking indicators include number of articles, number of articles per faculty, number of citations, number of citations per faculty, total number of scientific publications, total number of scientific publications per faculty, number of Ph.D. students, percentage of Ph.D. students, number of students per faculty. According to the numerical application carried out for 109 state and 48 foundation universities, it was determined that proposed method gave remarkable results.

Key Words: University ranking, academic performance, scientometrics, fuzzy c-means clustering.

An Experimental Design Application in Chemistry for Optimization of Assay Conditions for tRNA

Oğuz Akpolat¹

¹Muğla Sıtkı Koçman University, Faculty of Science - Chemistry Department, Muğla, Turkey

Abstract

A design of an experiment is generally to fictionalize an execution of an experimental process *and* the term of experimental design strategy usually refers to a two-stage modeling. The *first of all* named as working strategy is the determination of the experimental execution model *and the last one is that* a linear or a non-linear mathematical model for response surface function to define the relationship between the experimental factors. As the simplest one of the scan designs is full factorial design Central Composite Design is used for optimisation of the reaction parameters influenced to the response of the reaction.

Here it was selected optimization of assay conditions for tRNA in vitro using Central Composite Design, the influence of three factors, namely (1) pH, (2) enzyme concentration and (3) amino acid concentration as an experimental design application and improved an experiment strategy in order to optimize the effective operation factors on this chemical reaction, mathematical solution techniques of the response surface functions to define the relationships between the factors and the experimental result affected by the factors, computer programming for the optimization of the reaction parameters and usage of some related software, statistically evaluation for the experimental results. As examining of the coefficients of the response surface equation and its graphics, we can say that the most effective parameter on the esterification process is enzyme concentration alone and together with pH.

Key Words: Experimental design, statistics, optimization, response surface, tRNA.

¹ oakpolat@mu.edu.tr

Smoothing Techniques in Solving Absolute Value Equations

Nurullah Yilmaz¹, Ahmet Sahiner²

¹Department of Mathematics, Suleyman Demirel University, Isparta, Turkey

²Department of Mathematics, Suleyman Demirel University, Isparta, Turkey

Abstract

In this study, we deal with the problem of absolute value equations which correspond to one of the non-smooth optimization problems. Two different type of smoothing techniques which are local and global smoothing techniques are applied to smooth out the non-smooth problem. A new algorithm based on the smoothing functions is developed to obtain the numerical solution of the problem. Finally, numerical experiments are presented to show the efficiency of the new algorithm.

Key Words: Non-smooth optimization, smoothing function, absolute value equation.

¹ e-mail: nurullahyilmaz@sdu.edu.tr

² e-mail: ahmetsahiner@sdu.edu.tr

Approximated Convex Bezier Curve for the Continuation Global Optimization Approach

Ahmet Sahiner¹, Idris A. Masoud Abdulhamid², Nurullah Yilmaz³

¹Department of Mathematics, Suleyman Demirel University, Isparta, Turkey

²Department of Mathematics, Suleyman Demirel University, Isparta, Turkey

³Department of Mathematics, Suleyman Demirel University, Isparta, Turkey

Abstract

The main idea in smooth global optimization is to modify the related function to a smooth function by decreasing the number of local minimizers.

In this study, we offer a new method for continuous global minimization problems. This method is based on constructing a transformation aimed to convert the given objective function into an auxiliary function with less number of the local minimizer by suggested control points treated with bezier curves. Accordingly, an algorithm is designed to find the global minimizer of the auxiliary function. The numerical results we have obtained by this method demonstrate the efficiency of the new method and related algorithm yielding the highest possible accuracy.

Key Words: Global minimization, auxiliary function method, smooth optimization.

¹ e-mail: ahmetsahiner@sdu.edu.tr.

² e-mail: idrisatea@gmail.com.

³ e-mail: nurullahyilmaz@sdu.edu.tr

A New Smoothing Technique for Total Variation Minimization in Image Denoising

Ahmet Sahiner¹, Shehab A. Ibrahim²

¹Department of Mathematics, Suleyman Demirel University, Isparta, Turkey

² Department of Mathematics, Suleyman Demirel University, Isparta, Turkey

Abstract

Image denoising is an important field in image processing and computer vision studies. If we have any noisy image, the main aim is to recover the original image or at least try to reduce the noise and keep the edge information of this image. Total variation function (TV-function) is one of the most important functions that have been widely used in image denoising because of its feature in maintaining image texture, but this function is non-differentiable and leads us to the search for any approach to make this function (TV-function) differentiable and then use optimization techniques for image denoising easily. In this study, a new smoothing technique is presented to make (TV-function) differentiable and smooth, then the new smoothed function is used for denoising the noisy image with the help of gradient descent method in minimization.

Key Words: Smoothing technique, total variation, image denoising, optimization.

¹ e-mail: ahmetsahiner@sdu.edu.tr

² e-mail: mullaiq@gmail.com

An Exact Approach for A Dynamic Workforce Scheduling Problem

Okan Örsan Özener¹

¹Özyeğin University Industrial Engineering, İstanbul, Turkey

Abstract

Taking into account the global and national rules and regulations, assigning a given number of employees to planned shifts while paying attention to required working hours, rest times and off day/days is one of the most complex planning issue both in literature and real world, and this is the concept of working schedule planning in particular. This issue is deserved to have an intense concern in real world because there are so many considerations that it is not an easy and simple planning issue; rather, one of the most complex problems, namely a subject of NP-Hard Problems.

In this thesis, as a prototype of a possible real world problem in workforce scheduling, we study a company trying to minimize the all the direct and indirect cost related to workforce scheduling. Satisfying the given limited work force and labor-related constraints, our aim is to determine the minimum cost solution. This is why, reviewing the studies in the literature including with examples of some problems, we define our problem along with the main characteristics and assumptions. We propose a solution approach based on an exact solution of the integer programming formulation of the problem and observe that our solution approach generates high quality solutions in acceptable solution time. The optimality gap of the solutions obtained in one-hour computational time limit is only 19.38%.

Key Words: Workforce scheduling, planning, integer programming, cost minimization.

Use of Six Sigma Method as a Process Improvement Techniques

Atakan Gerger¹, Ali Riza Firuzan²

¹Continuous Improvement & Lean, PFW GmbH, Speyer, Germany

²Statistics, Dokuz Eylül University, İzmir, Turkey

Abstract

This study suggested a conceptual Six Sigma/DOE hybrid framework aiming to integrate Taguchi methods and Six Sigma for process improvement in a complex industry environment. The studies employing the Taguchi experiment design usually focus on a single variable and neglect the effects of the adjustments on remaining quality characteristics, in this study a multi-response experiment design was preferred, and all of the quality characteristics are taken into account.

The aim of this study is the implementation of Six Sigma methodology on a company operating within the automotive industry and the statistical evaluation of the methodological effects on the performance of the enterprise. This Six Sigma study was conducted to improve the manufacturing process which caused a customer complaint within the company.

DMAIC phases are used to reduce the nonconformity rate from 23.940 percentage (baseline) to 0.049 percentage. As the result of implementing SS, the sigma level increased from 2.21 (baseline) to 4.80.

Key Words: Six sigma, multi response Taguchi design of experiment, industrial statistics application, industry 4.0 and statistics.

Data Mining in Finance: Comparing Decision Trees and Association Rules for Stock Price Prediction in BIST 100

G.Ataman¹, S.Kahraman²

¹Business Administration, Yasar University, Izmir, Turkey

²Economics, Yasar University, Izmir, Turkey

Abstract

With the increase of financial globalization, there has been a need for methods to predict financial data effectively. Nonetheless, large volume of financial data has far exceeded the simple econometric models for the analysis. Thus, in this context use of data mining techniques becomes unavoidable for researchers to extract useful information. Data mining, which is an emerging technology combining statistics, artificial intelligence and machine learning, are generally used to detect hidden trends and patterns in data sets for building explanatory and predictive models. In finance, not only can data mining help financial institutions to increase efficiency and reduce financial instability, it can also increase the liquidity level (cost and time). In this study, two main data mining techniques, classification and association analysis, are used for modelling potentially successful and risky stocks on the Borsa Istanbul BIST 100 Index. Different variables related to balance sheet and income statement (or such as liquidity ratio, capital ratio, profit level etc) are used as input variables of the models. For classification analysis decision tree based technique namely Classification and Regression Tree (CART) is used and most commonly known algorithm, Apriori, is applied for association analysis. Mined association rules are compared with predictive rules mined with CART.

Key Words: Stock market, data mining, CART, apriori, association, classification.

Energy Consumption and Economic Growth: Panel Simultaneous Equation Models

Pınar Hayaloğlu ¹, Seyfettin Artan ², Selim Koray Demirel ²

¹Doç. Dr., Gümüşhane Üniversitesi, İİBF, İktisat Bölümü, Gümüşhane /Türkiye

²Prof. Dr., Karadeniz Teknik Üniversitesi, İİBF, İktisat Bölümü, Trabzon/Türkiye

³Arş. Gör., Karadeniz Teknik Üniversitesi, SBE, İktisat Bölümü, Trabzon/Türkiye

Abstract

The relationship between energy consumption and economic growth has recently been discussed, especially in the energy literature. However, the results of the empirical studies show differences according to the methods, variables and country groups used in the analysis. In addition, the interaction between energy consumption and economic growth has been largely ignored in the literature. In this study, the relationship between energy consumption and economic growth in OECD countries was investigated for the period 1990-2017 by using panel simultaneous equation systems. The results show that there is positive relationship exists between energy consumption and economic growth in OECD countries. Accordingly, high economic growth requires more energy consumption and more energy consumption needs higher level of economic growth. Thus, countries aiming for sustainable economic growth, will need more energy consumption. Therefore, in order to minimize the potential environmental problems caused by more energy consumption, it is important that policy makers turn towards eco-friendly energy sources.

Key Words: Energy consumption, economic growth, panel simultaneous equation system.

Cumulant Generating Function of Sample Minimum of Order Statistics from Geometric Distribution

Aslıhan Demir¹, Ayşe Metin Karakaş², Sinan Çalık³

¹ Firat University, Department of Statistics, 23000, Elazığ, TURKEY

² Bitlis Eren University, Department of Statistics, 13000, Bitlis, TURKEY

³ Firat University, Department of Statistics, 23000, Elazığ, TURKEY

Abstract

In this study, cumulant generating function of sample minimum from a geometric distribution of statistics based on order statistics are considered. The moments of the sample minimum of order statistics from a geometric distribution are obtained with the help of cumulant generating function. Using these moments, the expected value and variance were obtained as algebraic and numerical.

Key Words: Order statistics, geometric distribution, moment, sample minimum, expected value, variance, cumulant generating function.

¹ e-mail: a.bilici128@gmail.com

² e-mail: aysekarakas5767@gmail.com

³ e-mail: scalik@firat.edu.tr

Volatility Measurement Entropy Methods

Ayşe Metin Karakaş¹, Aslıhan Demir², Sinan Çalık³

¹ Bitlis Eren University, Department of Statistics, 13000, Bitlis, TURKEY

² Firat University, Department of Statistics, 23000, Elazığ, TURKEY

³ Firat University, Department of Statistics, 23000, Elazığ, TURKEY

Abstract

In this paper, we show that the application of different entropy methods for world indices. To do this, we use the world indices such as Istanbul Stock Indices (BIST 30), Brazil Index (Bovespa), Germany Index (DAX), Britain Index (FTSE100), South Korea (KOSPI), Japan Index (Nikkei 225), United States Index (SP 500) and China Index (SHANGAI) that have been investigated over all of 8 years (2010–2018). We obtain Shannon, Tsallis, Rényi and at last the approximate entropy. Consequently, we provide computational results for these entropies for weekly and monthly data.

Key Words: Shannon entropy, Tsallis entropy, Rényi entropy, approximate entropy.

¹ e-mail: aysekarakas5767@gmail.com

² e-mail: a.bilici128@gmail.com

³ e-mail: scalik@firat.edu.tr



Entropy Approach for Wind Energy

Ayşe Metin Karakaş¹, Aslıhan Demir², Sinan Çalık³

¹ Bitlis Eren University, Department of Statistics, 13000, Bitlis, TURKEY

² Firat University, Department of Statistics, 23000, Elazığ, TURKEY

³ Firat University, Department of Statistics, 23000, Elazığ, TURKEY

Abstract

In this study, we give the practice of entropy in wind energy. Firstly, we fit marginal distributions to each of the variables and later demonstrate the notion of entropy to perform a comparison the wind energy data of the stations (Bursa, Elazığ, İstanbul, Muğla, Rize, Tokat, Van and Zonguldak) that have been examined in a period 2015–2018. The results of probability distribution fitting to these wind energy variables show that the wind energy time series of Bursa, Elazığ, İstanbul, Muğla, Rize, Tokat, Van and Zonguldak are best resubmitted by Gamma Burr and Lognormal distributions. Later, we calculate Shannon entropy for several various values, Tsallis entropy, Rényi entropy and the approximate entropy. We form calculation outcomes with these entropies for daily data.

Key Words: Shannon entropy, Tsallis entropy, Rényi entropy, approximate entropy, wind energy.

¹ e-mail: aysekarakas5767@gmail.com

² e-mail: a.bilici128@gmail.com

³ e-mail: scalik@firat.edu.tr



Combination of Forecasting Benchmarks on M4 Competition Data Sets

Tuğçe Ekiz Yılmaz¹

¹Department of Statistics, Dokuz Eylül University, İzmir, Turkey

Abstract

Forecasting competitions have affected the field of forecasting greatly over the years, providing a solid basis for assessing different extrapolation approaches and learning empirically how to advance forecasting theory and practice. For a long time, instead of using an individual model, it has been acknowledged that combinations of individual models enhance considerably the forecasting performance of combinations. There are still two fundamental forecasting methods in the literature: ES (exponential smoothing) and ARIMA. Another method that participated in the M4 competition as an alternative to these major methods is the ATA method. ATA method is purely statistical method which generates more accurate forecasts comparing to other major statistical methods. In this study, improvement of forecasting performance will be analyzed by adapting ATA method to combinations of other major methods.

Key Words: Forecasting, M4-competition, exponential smoothing.

¹ Department of Statistics, Dokuz Eylül University, İzmir, Turkey.

Financial Data Analysis by Exponential Smoothing and ATA Method

Selma Şalk¹, Güçkan Yapar²

¹Department of Statistics, Dokuz Eylül University, İzmir, Turkey

²Department of Statistics, Dokuz Eylül University, İzmir, Turkey

Abstract

Time series occurs by collecting the data in a particular category in a given time period. Accurate analysis of financial data, which is a sort of time series, has a great importance for financial institutions to make predictions for the future. Exponential smoothing method is one of the most used method in time series analysis. Exponential smoothing methods have been used widely for many years due to their simplicity and success in prediction results. The success of the method has proved many times in the famous M-competitions. However, the selection of initial value and smoothing constant according to subjective choices for exponential smoothing method adversely affect the accuracy of this method. The ATA method, which is a new method developed as an alternative to the exponential smoothing method, eliminates these disadvantages of the exponential smoothing method. In this study, the M4 results of exponential smoothing method and ATA method will be compared, especially their performance in financial data will be evaluated.

Key Words: Time series, financial data, exponential smoothing, ATA method, M-competition.

¹ Department of Statistics, Dokuz Eylül University, İzmir, Turkey.

² Department of Statistics, Dokuz Eylül University, İzmir, Turkey.

Exchange Rate Volatility and Stock Returns: A case of China and Turkey*

Tao Wang¹, Özlem Sayılır²

¹Master Student, Anadolu University, Social Sciences Institute, Eskişehir, Turkey

²Anadolu University, Faculty of Business Administration, Eskişehir, Turkey

Abstract

This study investigates the relationship between stock returns and exchange rate volatility in China and Turkey from 1990 to 2016. GARCH (1,1) model is employed to estimate the volatility of exchange rates. The relationship between exchange rate volatility and stock returns is modelled with OLS Regression and Granger Causality methods. The OLS Regression results show no evidence of an impact of exchange rate volatility on stock returns in China or Turkey. On the other hand, OLS Regression results exhibit that GDP has a significant and positive impact on stock returns in Turkey and China. Regarding Granger causality findings, there is evidence of causality from exchange rate volatility to stock returns in Turkey. Moreover, Granger causality from GDP to stock returns in Turkey is found. In China, the Granger causality runs from stock returns to GDP.

Key Words: Exchange rate volatility, stock returns, Granger causality, China, Turkey.

*Extracted from the Master's Thesis of Tao Wang (Anadolu Anadolu University, Social Sciences Institute)

Examining the Financial Success of Stocks by Using Data Mining Techniques

Serpil Altınırmak¹, Ezgi Demir², Batuhan Atasoy³, Çağlar Karamaşa⁴

¹Vocational High School, Anadolu University, Eskişehir, Turkey

²Department of Management Information Systems, Piri Reis University, İstanbul, Turkey

³Department of Mechanical Engineering, Piri Reis University, İstanbul, Turkey

⁴Department of Accounting and Financing, Anadolu University, Eskişehir, Turkey

Abstract

In recently, data mining techniques have been used widely for economy and finance sectors. Because financial data has a large mass, it is important to conclude meaningful associations and relations from these data in terms of using data mining techniques. Decision tree, artificial neural networks, support vector machines and clustering analysis are among the most used techniques in data mining.

In this study, annual data and market financial successes of stocks of 22 companies traded in food sector in BIST have been analyzed through 69 different criteria and then estimated by using data mining techniques.

Key Words: Data mining, financial failure, stocks.

The p-Splines Model in Predicting PM10 Concentrations

Betul Kan-Kilinc^{1*}, Huruy D. Asfha²

¹Science Faculty, Department of Statistics, Eskisehir Technical University, Turkey

²Department of Mathematics and Statistic, Wichita State University, Wichita KS, USA

Abstract

There are numerous examples that can be the sources of air pollution including transportation (vehicles and fuels), increasing numbers of inhabitants, industrial activity, waste reduction of forest pollution, energy (heating, lighting, electricity) and climate. The aim of this study is to predict daily average patterns of PM10 including abnormal observations by regressing on the potential variables when existence of some degree of relationships among them. Increased usage in additive regression models motivated us to examine nonparametric statistical modelling in a class of complications including abnormal observations and multicollinearity. The best fitting model by penalized regression splines and regressors explaining the dependent variable are investigated and compare with linear regression model.

Key Words: Splines, abnormal observations, multicollinearity, additive.

¹ Corresponding Author E-mail: bkan@eskisehir.edu.tr

² E-mail: asfha@math.wichita.edu

Evaluating Negative Binomial Regression and Zero Truncated Negative Binomial Regression for Different Dispersion Parameters

Eda akmak¹, Ergun Karaaaoğlu²

¹Başkent University, Faculty of Health Science, Ankara, Turkey

²Hacettepe University, Faculty of Medicine, Department of Biostatistics, Ankara, Turkey

Abstract

Poisson distribution is the number of events occurring at a definite time interval. Poisson regression method is used in the models as count data, when the dependent variable is considered as the number of events at a certain time interval. Poisson regression analysis is used when the mean of distribution is equal to its variance. However; this condition is not always obtained. When the variance of distribution is greater than the mean, overdispersion in model occurs. In this case, Negative Binomial regression (NB) method is used. NB distribution includes zero counts, but some count data do not include structural zero value. Thus, Zero Truncated Negative Binomial (ZTNB) method is used in these models. The aim of our study is to compare the AIC (Akaike Information Criteria) values of the models created by NB and ZTNB methods for different dispersion parameters (α). As special case with dispersion parameter, when it is zero ($\alpha=0$), Negative Binomial model converges Poisson model. Also when it is one ($\alpha=1$), NB model converges Geometric distribution. In our study, analysis was run in Stata/MP 13 version program. We generated 10000 observations with NB distribution and established a model consisting of three independent variables. In scope of simulation study, we created models for NB and ZTNB distributions for dispersion values ($\alpha=0.05, \alpha=0.10, \alpha=0.25, \alpha=0.50, \alpha=0.75, \alpha=0.90, \alpha=0.99$). Firstly, we estimated the models with NB distribution for dispersion values and then, dropped all zero counts from the data. Therefore, we obtained ZTNB distribution and reestimated the model. Both NB and ZTNB distribution evaluate goodness of fit statistics for dispersion values. We demonstrated that percentage of zero counts change with differing dispersion values. According to our results, when the models created for each alpha value are examined, alpha values increases as AIC values increases as well as zero counts.

Key Words: Negative binomial, zero truncated negative binomial, AIC, dispersion parameter.

References

- [1] Hilbe, J.M. (2011). *Negative Binomial Regression*. Second Edition. New York: Cambridge University Press.
- [2] Hintze, J.L. (2007). *NCSS User's Guide III Regression Curve Fitting*. Kaysville, Utah.
- [3] Long, J.S. (1997). *Regression Model for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

Binary Particle Swarm Optimization as a Detection Tool for Influential Subsets in Linear Regression

Deniz Inan¹, Gokce Deliorman²

¹ Department of Statistics, Marmara University, Istanbul, Turkey

² Department of Mathematics, Marmara University, Istanbul, Turkey

Abstract

An influential point is any point that has a large effect on the slope of a regression line fitting the data. The existence of such observations in the data set reduces the validity and sensitivity of the statistical analyzes. There are many methods used for identifying influential observations in the literature. However, many of these methods require distributional assumptions and are highly influenced by masking and swamping effects. Especially in the presence of influential subsets most of these methods are insufficient.

The aim of this paper is to develop a new detection method, not require any distributional assumptions and also not affected by masking and swamping effects as much. Binary particle swarm optimization is used as a diagnostic tool for this purpose. The performance of the proposed method is analyzed via simulations and real data set applications.

Key Words: Influential subsets, binary particle swarm optimization, heuristic algorithms.

J-Inflated Longitudinal Regression Models for Analyzing Responses with Finite Support

Mojtaba Ganjali¹ and Taban Baghfalaki²

¹Statistics Department, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran.

²Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran.

Abstract

In recent years, zero-inflated Binomial distribution is commonly used for analyzing responses with extra zeros and finite support. In this paper some extensions are presented for analyzing such data: (1) inflation may happen in any point of the support of the distribution such as point J or any two points such as J and K, (2) success probability may be such that the probability of the first success is p and in each successive trial it is dependent on p and the value of the response and so the distribution of the response is the two parameter quasi-binomial, (3) the data may be collected over time and a longitudinal approach that takes into account the correlation between responses of the same individual have to be used. A model considering all of these points is used to analyze a real data set where the support of the response is finite.

Key Words: Transition model, random effects model, binomial distribution, zero-Inflated longitudinal response.

Opposition Based Particle Swarm Optimization for Feature Selection

Merve Özdemir¹, Ümmügülsüm Can¹, Berna Kiraz¹

¹Department of Computer Engineering, FSM Vakif University, Istanbul, Turkey

Abstract

Feature selection is a pre-processing method used to select the most relevant and important features for the problem of interest. Meta-heuristics such as particle swarm optimization, genetic algorithm and differential evolution have been successfully applied to feature selection. In a feature selection approach, the candidate solution is represented as a binary array which shows whether the corresponding feature is included in the feature subset or not. On the other hand, in opposition based learning, a current estimate and its opposite are taken into consideration simultaneously to speed up the performance of the algorithm. In this study, we propose the binary particle swarm optimization (BPSO) with opposition-based learning for feature selection in classification problems. In BPSO, each particle in the swarm has a velocity (continuous value) and a position represented by a binary array. BPSO consists of three steps: (1) initialization of the swarm, (2) evaluation of the swarm and (3) update the components of particles. First, the velocity of each particle is randomly initialized. Then, the position is generated according to a corresponding velocity using sigmoid function. Second, each particle in the swarm is evaluated. Finally, the velocity and the position are updated iteratively until a termination condition is satisfied. As in BPSO, our method has three main steps. In the first step, an opposition-based initialization is used. This method initializes the velocity of each particle randomly and calculates the opposite of the velocity. The opposite point is calculated as mirror reflection utilizing the minimum and maximum values of velocity. The initial swarm and its opposite are merged and the best particles from the merged one are chosen as the initial swarm. To evaluate each particle, we use the support vector machines as classifiers. The fitness of each particle is calculated as the average accuracy on 10-fold cross validation. In the experiments, we consider three data sets, namely breast cancer including 30 features, sonar including 60 features and musk including 166 features. We compare our method to BPSO and sequential forward selection. The results show that our method delivers better performance than the others.

Key Words: Feature selection, binary particle swarm optimization, opposition based learning.

Artificial Bee Colony Algorithm for Estimating Parameters of a Flexible Weibull Distribution

Ilhan Usta¹, Volkan Yesildere¹

¹Department of Statistics, Eskisehir Technical University, Eskisehir, Turkey

Abstract

The two-parameter flexible Weibull distribution introduced by Bebbington, et al [1]. has a hazard function that can be increasing, decreasing or bathtub shape. Maximum likelihood (ML) estimation is a widely used method to estimate the parameters of the flexible Weibull distribution (FWD). However, the ML estimators for FWD cannot be obtained in closed forms. In this study, artificial bee colony (ABC) algorithm, which is a heuristic-based optimization method is used to compute the ML estimates of the parameters of FWD. The ML estimates are also obtained by the numerical method. Furthermore, we conducted an extensive Monte Carlo simulation study to compare the performance of our proposed algorithm with the corresponding numerical method. The results show that the ABC algorithm perform better than the numerical method in terms of mean square error.

Key Words: Flexible Weibull distribution, maximum likelihood estimation, artificial bee colony algorithm.

Reference

[1] Bebbington, M., Lai, C.D., Zitikis, R. Reliability Engineering and System Safety 92:719-726, 2007.

A Study on Loss Probability of a Blocked Queueing System

Erdoğan Yücesoy¹, Vedat Sağlam², Murat Sağır³

¹Department of Mathematics, Ordu University, Ordu, Turkey

²Department of Statistics, Ondokuz Mayıs University, Samsun, Turkey

³Department of Economics, İskenderun Technical University, Hatay, Turkey

Abstract

In this paper the loss probability of a blocked tandem queueing system is investigated. The arrival process to this queueing system is Poisson with parameter λ . There is one service unit at first stage of the system and the service time of this unit is exponentially distributed with μ_1 parameter. There are two parallel service units at second stage and the service time of these service units are exponentially distributed with parameters μ_2 and μ_3 . In this constructed queueing model, the most important measure of performance is the loss probability. The loss probability of the model is theoretically obtained. Furthermore, the mean number of customers is calculated.

Key Words: Stochastic process, queueing model, Markov chains, measure of performance.

Parameter Estimation of the Kies Distribution by Artificial Bee Colony Algorithm

Ilhan Usta¹, Volkan Yesildere¹

¹Department of Statistics, Eskisehir Technical University, Eskisehir, Turkey

Abstract

This paper considers the maximum likelihood (ML) estimates based on artificial bee colony (ABC) algorithm for estimating the four-parameters of Kies distribution, which was introduced by Kies [1] as an alternative to Weibull distribution with much wider applicability. We also calculate the ML estimates of the parameters using the Newton–Raphson method. Monte Carlo simulation studies carried out to evaluate numerically the performance of our proposed algorithm according to bias and mean square error. Simulation results show that the ML estimates via ABC algorithm perform better than the numerical method in terms of resulting in a smaller bias and mean square error in parameter estimation.

Key Words: Kies distribution, maximum likelihood estimation, Newton–Raphson method, artificial bee colony algorithm, mean square error.

Reference

[1] Kies, J.A. The strength of glass performance. Naval Research Lab Report No. 5093, Washington, D.C., 1958.

Data Mining of SILC Data: Turkey Case

Olgun Özdemir¹, İbrahim Demir²,

¹Statistics, Yıldız Technical University, Istanbul, Turkey

²Associate Professor, Statistics, Yıldız Technical University, Istanbul, Turkey

Abstract

Official data produced by the National Statistical Institutes have an essential place in the governmental economic and social decision-making process. Addressing official data with data mining methods rather than traditional statistical approaches is crucial in terms of extracting new information and hidden patterns. However, the usefulness of data mining methods for official statistics remains largely unexplored. In the present study, SILC (Survey of Income and Living Conditions) data for the year 2015 conducted by Turkish Statistical Institute (TurkStat) are examined with data mining methods. Cross-sectional data of 36036 individuals were handled, and the variables affecting the individual income were determined, also the welfare status of the individuals was examined. In order to determine the socio-economic profiles of individuals, latent class analysis (LCA) and k-modes clustering analysis were used. Socio-economic status of individuals was classified using clustering and random forest (RF) algorithm models. In the LCA model with ten classes, it was obtained which probability of a newly selected individual would belong to which class. The latent class profile definitions of the individuals were obtained according to the variable values obtained from the latent classes with the highest probability. Ten clusters obtained as a result of k-modes were defined according to cluster modes, and cluster profile definitions of individuals were obtained, and also their results were compared with LCA results. In the RF model, where individual income is selected as a function of all nine input variables, the importance of the variables was determined. It is observed that education, occupation, and age variables were more important and made the most contribution to the RF model, respectively. In this study in which categorical variables were considered, it was seen that LCA method provided more consistent results than k-modes method. In the SILC data, which is an extensive and detailed data, methods such as LCA and RF seem to be appropriate for the application of data mining and obtaining meaningful results from the data. Similar data mining processes can be used to obtain meaningful results for different official data.

Key Words: Data mining, SILC, cluster analysis, latent class analysis, k-modes, random forests.

¹ Correspondence to: olgunoazdemir35@gmail.com

² idemir@gmail.com

Comparison of Predictive Performances of CART and Logistic Regression Models: Application of Life Satisfaction

Arzu Yavuz ¹, Özgül Vupa Çilengiroğlu ²

Department of Statistics, Dokuz Eylül University, The Graduate School of Natural and Applied Sciences, İzmir, Turkey

Department of Statistics, Dokuz Eylül University, İzmir, Turkey

Abstract

Classification and regression trees; A statistical method that shows the model of meaningful relationships between dependent and independent variables in the form of a tree. This method has been used frequently for easy interpretation, implementation of large data sets and not requiring assumptions. Classification and regression trees take the name "classification tree" if the dependent variable is categorical, and "regression tree" if it is continuous. Logistic regression is one of the classification methods and is used in the case where the dependent variable is on the nominal or ordinal scale. The concept of life satisfaction is a subject that has entered into the interests of many different disciplines nowadays. Life satisfaction is the addressing of how much the individual is enjoying life as a whole. The aim of this study is to make the performance of CART and logistic regression analyses from decision tree methods using the life satisfaction data of the Turkish Statistical Institute (TÜİK) covering the period 2017.

Key Words: Logistic regression, CART algorithm, life satisfaction.

Some Properties of Covariance Matrices of Predictors under Seemingly Unrelated Regression Models

Nesrin Güler

Sakarya University, Statistics Department, Sakarya, Turkey

Abstract

Seemingly unrelated regression (SUR) models are extensions of linear regression models by considering multiple regression equations with correlated errors. Prediction problems of unknown vectors in linear regression models are classical concepts in the statistical inference. Best linear unbiased predictors (BLUPs) and ordinary least squares predictors (OLSPs) are well-known and most popular predictors that are widely considered in statistical literature.

In this study, m -equation SUR models and their joint form corresponding classical linear regression model by making use of block matrices are considered. The main purpose of the study is to obtain some properties between covariance matrices of predictors and also to establish some conditions for performance of BLUPs and OLSPs under considered SUR models. Expressions for covariance matrices of BLUPs and OLSPs consist lots of complicated matrix expressions with Moore-Penrose generalized inverses of matrices. In order to establish variety of equalities and inequalities on the comparisons of covariance matrices of predictors under SUR models and to simplify matrix expressions, some rank formulas of matrices are used.

Key Words: BLUP, covariance matrix, OLSP, rank, SUR model.

Bayesian Inference for Tobit Model and an Application using with WinBUGS Program

Samet Kaya¹, Esin Köksal Babacan²
^{1,2}Ankara University, Ankara, Turkey

Abstract

In this study, Tobit model, which refers to dependent variable regression models with censored from below or above, has been studied. The aim of the study is to compare the results of the estimate obtained for Tobit regression and Bayesian Tobit regression. Bayesian Tobit regression estimates a linear regression model with a censored dependent variable using a Gibbs sampler.

In the study, first we have generated data from simulation using with a particular Tobit model. Then, we have compared the estimate of model parameters obtained with using both classic Tobit and Bayesian Tobit. We have concluded that both estimation results are similar. Finally we have performed a real data application for both classic and Bayesian Tobit. We have used the data of the 3 most preferred automobile models in the USA. For classic regression inference we have used Matlab program and for Bayesian tobit regression inference we have used Winbugs program

Key Word: Tobit model, WinBUGS, MCMC (Markov Chain Monte Carlo), Bayesian inference.

Wage Dynamics According to Income Group in Turkey: 2014-2017 Period

Şenay Üçdoğruk Birecikli¹, Nilgün Acar Balaylar², Hamdi Emeç³

¹Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü İzmir, Türkiye

² Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İktisat Bölümü İzmir, Türkiye

³ Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü İzmir, Türkiye

Abstract

Factors revealing wage differences were analyzed in economic theory by Becker (1964; 1965) who is known for his work on human capital, Chiswick (1971) and especially Mincer (1958, 1962, 1974). According to the Human Capital Theory, investment in human capital is the most important determinant of economic growth and wage differences. At the core of Human Capital Theory, there is the education level of individuals. The increase in the level of education rises the level of income of individuals as well as rises economic growth. The increase in productivity caused by education contributes to production. At the same time, both the consumption expenditures and savings of the individuals with increased income increase. While demand from increased spending encourages production, increased savings fund-raising for production growth. Consequently, the high wage level is both the cause and the result of the increase in economic welfare.

The aim of this study is to measure the effects determining the real wages of individuals who are in employment in the labor market in Turkey. For this, pooled data obtained by micro data set of 2014-2017 Household Labor Force Survey prepared by Turkey Statistical Institute (TÜİK) is used. The number of individuals between the ages of 15-65 and employed is 303,633. The method used in the research is the Least Squares and Slice Regression method. Since the extreme values in the data lead to varying variance, the differences between the low and high slices of the individual's real wage with the method used made a difference to the study.

In the model, Mincer's basic human capital and extended human capital model will be used first for estimating the real wages obtained by the individuals employed. Then the factors influencing the determination of real wages in Turkey will be analyzed by using the variables of difference between the consumer price index and the producer price index, the real exchange rate, labor productivity and the cost of minimum wages cost to employers.

Key Words: Human capital, quantile regression, real wages.

¹ Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü İzmir, Türkiye

² Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İktisat Bölümü İzmir, Türkiye

³ Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü İzmir, Türkiye

Modelling the Mediator Effect of Demographic Features on Clients' Deposit via Structural Equation Modelling

Sinan Saraçlı¹, Cengiz Gazeloğlu²

¹Department of Statistics, Afyon Kocatepe University, Afyonkarahisar, Turkey

²Department of Statistics, Süleyman Demirel University, Isparta, Turkey

Abstract

The Purpose of this study is to determine the mediator effect of demographics factors on bank clients' a term deposit and this relation via Structural Equation Modeling.

With this purpose a real data collected from a Portuguese retail bank, from May 2008 to June 2013 is used. The data set which Moro *et al.*, (2014) used in their study is downloaded from the web site of UC Irvine Machine Learning Repository. The data set includes 41188 observations. LISREL Software is used to analyze the data set.

The results of the study indicate that the latent variable about the demographic features of the Clients, which include age, job and marital status has a significant effect on their current deposit situations. The detailed results are also given in the related tables and figures.

Key Words: Bank clients, term deposit, structural equation modelling.

¹Corresponding author, E-mail: ssaracli@aku.edu.tr

² E-mail: cengizgazeloglu@sdu.edu.tr

The Impacts of Oil Price Fluctuations on the Turkish Stock Market: Correlations, Volatility Spillover and Hedging

Vasif Abioglu¹

Aksaray University, FEAS, Department of Economics, 68000 Aksaray, Turkey.

Abstract

The purpose of this paper is to investigate the conditional correlations and volatility transmission between oil price and stock market in Turkey. For this purpose, VAR-GARCH approach are used to analyze conditional correlations, volatility spillovers as well as hedging effectiveness between oil prices and the sectoral stock returns. We conduct our study at both the aggregate as well as 23 sector levels using weekly data for the period 2002-2018. Four different multivariate GARCH models (BEKK-GARCH, VARMA-GARCH, CCC-GARCH, and DCC-GARCH) are compared and contrasted. In addition, we analyze the optimal weights and hedge ratios for oil-stock portfolio holdings based on the results of multivariate GARCH models. Better understanding of conditional correlations and volatility spillover is crucial for portfolio management in the presence of oil price risk.

¹e-mail: vabiyev@aksaray.edu.tr, vasifabiyev@gmail.com, Tel: 0382 288 2448.

Linear and Nonlinear Unit Root Tests Application: Turkish Electricity Consumption Per Person

Fatih Çemrek¹, Büşra Yıldız²

¹Department of Statistics, Eskişehir Osmangazi University, Eskişehir, Turkey

²Department of Statistics, Eskişehir Osmangazi University, Eskişehir, Turkey

Abstract

The aim of this study was to investigate the per capita consumption of electricity in Turkey with linear and nonlinear unit root tests. The related series covers the first month of 2016 and the sixth month of 2019. First, conventional linear unit root tests were applied to the series. According to the Augmented Dickey Fuller and Phillips Perron unit root tests, it was concluded that the series was not stationary in the fixed term model and the fixed term and trend model, whereas the stationary term and non-trend model were not stationary. According to the results of Zivot Andrews structural fracture unit root test applied to the series, it was found that the series did not contain structural fracture unit root in A, B and C models. Harvey and Leybourne (2007) and Harvey et al. (2008) linearity tests of the series will be determined using linearity tests. Then non-linear unit root tests Kapetanios et al. (2003) and the development of this test will be investigated using Kruse (2011) and Guris (2018) tests.

Key Words: Nonlinear unit root, linear unit root, structural fracture, linearity, traditional unit root.

¹ fcemrek@gmail.com.

² busrayildiz1996@hotmail.com.

Noise Reduction in Chaotic Time Series for Improved Feature Selection, Abnormal Process Detection, and Process Modeling

Puneet Choudhary¹, Shibu Meher², Brahma Deo³, Manish Goyal⁴

¹Indian Institute of Technology Bhubaneswar, Odisha, India

²Indian Institute of Technology Bhubaneswar, Odisha, India

³Indian Institute of Technology Bhubaneswar, Odisha, India

⁴All India Institute of Medical Science, Odisha, India

Abstract

Noise removal in time series data of multidimensional non-stationary chaotic systems poses special problems. Data preprocessing involves removal of artifacts and reduction of noise followed by usual methods of treatment chaotic time series to find out time delay, embedding dimension, correlation dimension, and Lyapunov exponent. The underlying process dynamics is understood through the nature and regions of attractor present in the system. Artificial neural nets and extreme learning machine are used to model the system on the basis of obtained correlation dimension and time delay from the time series data. As examples of application we present treatment of non-stationary time series pertaining to altogether two different chaotic systems including (a) a high temperature thermal and chemical reaction system in engineering and (b) chaotic time series of EEG data for detection of epilepsy. Both Kalman filter and a non-linear polynomial filters are applied for noise reduction followed by modeling of the underlying dynamics and also abnormal process detection.

Key Words: Noise reduction, time series, chaos, modeling, filter.

Detection of Mode-Mixing Problem in EMD of Time Series Signal Data

Atacan Erdiş¹, M. Akif Bakır², Muhammed I Jaiteh³

¹Gazi University, Faculty of Sciences, Department of Statistics, Ankara, Turkey

²Gazi University, Faculty of Sciences, Department of Statistics, Ankara, Turkey

³Gazi University, Faculty of Sciences, Department of Statistics, Ankara, Turkey

Abstract

Signals are things that convey information about the status of a physical phenomenon in successive time intervals and can be defined mathematically in functional form. The signals may be kind of audio, video, speech, image, communication and musical etc. The waveform of the characteristic structure allows the signal to be explained by concepts such as amplitude, frequency and period. While the signals are conceptually divided into analog and digital, they can also be grouped as simple and compound signals (multi-component signals) in terms of their physical structure. Signals consisting of a mixture of two or more different signals may be referred to as compound signals. It is well known that the time series data generated from signals can be analyzed in frequency domain as well or instead of time domain analysis. Although there are various decomposition techniques in the literature to decompose compound signals into sub-signals that make up it, the Fourier transform is the most well known. However, the Fourier transformation requires that the signal data be linear and stationary. On the other hand, data driven Empirical Mode Decomposition technique (s) is proposed to decompose non-linear and non-stationary signals into single oscillatory modes called intrinsic mode functions (IMFs), sub-signals. The success criterion of separation is that all IMFs that make up the signal can be decomposed so that independent, that is, to reveal all the physical characteristics of the signal. However, the mode mixing is a common problem in EMD method. The mode mixing can occur in the form that a portion of the IMF may be quite similar to the adjacent IMFs. When mode mixing occur, the IMFs can lose their physical meaning hindering the interpretation of the results of the analysis. Although there exist some algorithm such as EEMD, CEEMDAN, which are extension of classical EMD, to overcome the mode mixing problem, in the literature, there is limited number of proposal to detect the existence of mode-mixing in IMFs. In this study, a new method is proposed for detection of mode-mixing in IMFs. Applying the proposed method through simulated data results in successfully detection of existency of mode-mixing in IMFs.

Key Words: Intrinsic mode function, empirical mode decomposition, mode mixing problem.

¹ Research Assistant, atacanerdis@gazi.edu.tr

² Prof. Dr., mabakir@gazi.edu.tr

³ Student, jaitehmuhammedi@yahoo.com

Investigation of the Relationship Between Turkey and Economic Growth Industrial Production Index

Fatih Çemrek¹, Büşra Yıldız²

¹Department of Statistics, Eskişehir Osmangazi University, Eskişehir, Turkey

²Department of Statistics, Eskişehir Osmangazi University, Eskişehir, Turkey

Abstract

In this study, Turkey's third quarter 2018 with the fourth quarter of 2006, including the base year 2015 = 100 total industrial production index and total gross domestic product (thousand) data is used. The aim of the study is to determine the cointegration relationship between two series which are not stationary and have the same level of seasonality. For this purpose, firstly the graphs of the two series were examined and it was checked whether the series had seasonality. Then, Augmented Dickey Fuller unit root test was applied to both series separately and it was determined that both series were stationary in I (2). Hylleberg-Engle-Granger-Yoo seasonal unit root test was applied to determine whether the series contain seasonal unit root at the same frequency. Both series contain seasonal unit root at zero frequency, half annual frequency and quarter annual frequency. According to Engle-Granger, if two non-stationary series can be stationary from the same order, the cointegration relationship between two series is examined since cointegration analysis can be performed. Cointegration regression equation was established between two series. Then, analysis was performed for $t = 0$ and $t = 4$ delay using Engle-Granger method. It is concluded that there is a cointegration relationship between industrial production index and gross domestic product series. Although they are not stationary in the two series level values, it is not inconvenient to include them in the regression equation as non-stationary since they move in parallel in the long run.

Key Words: Seasonal cointegration, seasonal unit root, stability, seasonality, unit root test.

¹ fcemrek@gmail.com.

² busrayildiz1996@hotmail.com.

Estimating the Output GAP for Lebanon: A Production Function Methodology

Latifa Ghalayini¹

¹Professor of Economics, Lebanese University

Abstract

This paper estimates the output Gap to analyze the economic policy and to judge the stance of the Lebanese economy. Therefore, a Cobb-Douglas production function is estimated for the period Q₁1998 to Q₄2015 and potential output is calculated by substituting for potential levels of the factors in the estimated production function. However, since the potential labor estimations require the calculation of the NAIRU, this paper calculates three types of NAIRU. The results of output gap calculations show that the Lebanese production inputs are hitting their limits. Furthermore, findings confirm the trade-off between unemployment and inflation for the Lebanese Economy and show that the labor market is characterized by high levels of NAIRU which restricted potential output growth. Therefore, any policy aiming to increase economic growth, while neglecting structural reforms will prove to be unsustainable.

Key Words: Production function, factor productivity, unemployment, potential output, monetary policy.

¹ lghalayini@hotmail.com

Sunk Costs and Exporting Behavior: A Sectoral Analysis

Kurmaş Akdoğan¹, Laura M. Werner²

¹ Central Bank of the Republic of Turkey, Structural Economic Research Department, Ulus, 06050, Ankara, Turkey.

² FernUniversität in Hagen, Faculty of Business Administration and Economics, 58084 Hagen, Germany.

Abstract

This article examines the hysteresis behavior due to sunk costs in exports of the Turkish manufacturing sector. The results of the analysis using the Preisach method for 2006Q1 to 2018Q2 reveal hysteresis for only one sector: The manufacturing of wearing apparel, dressing and dyeing of fur (*clothing*). To shed more light on this result we provide detailed information on the multi-layered production structure of the clothing sector. We argue that the sub-contracting capacity of intermediaries with their previous export experience and established connections, low importance of plant size in the entry decision, easier financing conditions and price advantage due to a real exchange rate depreciation are the main determinants of relatively lower sunk costs in this sector.

Key Words: Nonlinearity, path-dependency, exports.

JEL: C19, F14, L60

¹ kurmas@gmail.com

² laura.werner@fernuni-hagen.de

Complexly Interwoven Stochastic and Chaotic Tendencies in the Profit-Interest Ratios at a Recent Crisis Juncture in Turkey: Possibilities for Crisis-Management

Ahmet Kara¹

¹Department of Economics, Istanbul Commerce University, Istanbul, Turkey

Abstract

This paper presents a simple model of complexly interwoven stochastic and chaotic tendencies in the profit-interest ratios at a crisis juncture in Turkey between the years 2013 and 2016. During the period in question, Turkey experienced a crisis of intricately interlinked social, political and economic dimensions, which resulted in an attempted coup in July 2016. The strategic gridlock at the crisis juncture influenced the interactive dynamics of participatory (participation) banks and conventional banks in the Turkish banking system. In this paper, we developed a simple discrete-time aggregated model of the system so as to analyze the trajectory of one of the key variables of this dynamics, namely profit-interest ratio (relative profit rate) over time. During the period in question, we demonstrated the existence of two equilibria, one of which was stable and the other was unstable. We explored and exemplified the possibilities of joint stochastic and chaotic tendencies in the profit-interest ratios. We pointed out the possibility of crisis management policies that could get the relative profit rates over time out of the problem-prone instability zone and prevent them from sliding into chaotic regions with aperiodic fluctuations.

Key Words: Relative profit ratio, stability, instability, chaos.

¹ kara2010@ticaret.edu.tr

Practical Tests of Financial Bubbles: Better Critical Values for Detecting Exuberant Behavior

Ufuk Demiroğlu¹, Aytül Ganioglu²

¹Structural Economic Research Department, Central Bank of the Republic of Turkey, Ankara, Turkey

²Structural Economic Research Department, Central Bank of the Republic of Turkey, Ankara, Turkey

Abstract

A recent econometric methodology has become increasingly popular in policy work as a practical test to identify and date financial bubbles in real time, developed by Phillips et al. [Phillips, P. C., S. Shi and J. Yu (2015), "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500," *International Economic Review*, 56 (4), 1043-1077]. This methodology is employed to detect both expansions and collapses of financial asset bubbles. However, in many (if not most) applications the interest is in detecting only expansionary bubbles. We modify the Phillips et. al. methodology in a way that focuses only on expansionary episodes and provide new critical values for practical applications that aims only to identify such episodes.

Key Words: Tests of financial bubbles, explosive behavior in asset prices.

Inference of Economic Performance using Global Trade Data

Orhun Sevinç¹²

¹Structural Economic Research, Central Bank of the Republic of Turkey, Ankara, Turkey

²Centre for Macroeconomics, London, United Kingdom

Abstract

From Adam Smith to David Ricardo economists relate trade performance of nations to their economic performance. We review our latest methodology on how to use global trade data to infer economic growth and development. We use global trade data at year-country-product level and basic statistical techniques to develop productivity and misallocation measures that are also consistent with the economic theory.

The information contained in the exports of a country compared to others has been considered to reveal comparative advantages since Balassa (1965), and still attracts attention of academic research. Existing approaches range from building better indexes than Balassa's original contribution, to building theoretically consistent ways to estimate comparative advantage to understand their impact on economic performance. While most papers in this literature are concerned with the level of comparative advantages, the current paper explores the value of information embedded in the cross-country differences in the variance of comparative advantages within sectors. Our analysis reveals interesting patterns on countries' development paths. Notably, a large and historical allocative efficiency gap exists across the West and the East of Europe, while the latter shows a remarkable progress in catching up. A similar temporal pattern is also observed between North and Central America. On the other hand, the substantial gap between the North and South America is highly persistent and slightly increasing over time. There is evidence of divergence in the East of Asia. South-Eastern Asian countries on average exhibit increasing misallocation over time, contrary to those of East Asia, which gradually increase efficiency in resource allocation. We then investigate the negative association of misallocation to economic performance by employing a multivariate regression analysis. The index of misallocation inversely and robustly predicts cross-country productivity differences using the available data from the Penn World Table. Since the theoretical channel that misallocation affect economic performance is through aggregate TFP, I show that misallocation index contains valuable information in cross-country GDP per capita regressions containing human and physical capital.

Given the sectoral availability of inferred measures of economic performance we introduce a method for selecting target industries that could identify the key areas for promising economic development. The idea is based on targeting incentives to those that are the most sophisticated, having the highest comparative advantage, and subject to highest distortions. An application for the Turkish economy is presented.

Key Words: Panel fixed effect models, latent factor based indexes, economic growth and development.

Change Point Detection in ARMA(p,q) Process and Application on Headway Data

Almila Hacıoğlu¹, Aylin Alın², Süheyla Pelin Çalışkanelli³, Ufuk Beyaztaş⁴

¹Statistics, Dokuz Eylul University, Izmir, Turkey

²Statistics, Dokuz Eylul University, Izmir, Turkey

³Civil Engineering, Dokuz Eylul University, Izmir, Turkey

⁴Statistics, Bartın University, Bartın, Turkey

Abstract

Traffic flow is essentially a random process. Therefore, various characteristics of traffic flow can be described by the techniques of probability and statistics. Vehicle Time Headway is a fundamental measure in both traffic flow theory and transportation applications. The headway is the time interval between two vehicles. Structural stability is important for time series data analysis like the headway data. The common way to evaluate the structural stability is to test the model parameters for a possible change at an unknown time point that is generally called change point. Detecting change points is crucial for determining the headway characteristics. One of the problems related to detecting changes in the data is the outliers. If the test statistic used for the estimation is not robust to these outliers, the existing outliers can be falsely estimated as the change point.

Headway data generally have the structure of ARMA(p,q) time series process. We propose a robust test statistic based on weighted likelihood approach for change point detection in ARMA(p,q) process. We would like to determine correctly the change points in the headway data using proposed method, and to contribute to the improvement of traffic flow by making adjustments in the traffic lights by minimizing the change points.

Key Words: Change point, ARMA process, weighted likelihood estimation, vehicle time headway.

A New Robust Unit Root Test with a Simulation Study

Hakan Türkay¹, Özge Gündoğdu²

¹Econometrics, Cumhuriyet University, Sivas, Turkey

²Econometrics, Suleyman Demirel University, Isparta, Turkey

Abstract

Stationarity condition is one of the important terms in the time series analysis. Accurate detection of stationarity in the variables to be used in the analysis is important in term of accuracy in analysis results and validity of interpretation. Dickey-Fuller (DF) test that first of the unit root tests are used widely in determining of the stationary. DF test is based on Ordinary Least Square (OLS) and OLS is very sensitive to outliers in the data set and against to deviations from assumptions. It is clear that, If OLS give misleading results, DF test will lead to making wrong decisions about the unit root. Robust regression methods that given more robust results according to OLS in the presence of outlier have been developed. These methods are not affected by the existence of outlier values. In this study, it aims to reduce the sensitivity against outliers in the test by DF regression equation estimating with S, M and MM estimators which are robust regression methods. A simulation study was performed by adding outliers of various numbers and sizes to the data set for this purpose. Obtained results from OLS and other robust methods were compared. It is concluded that it is a more appropriate approach to predict DF regression equation by means of robust estimators in terms of deciding stationarity accurately according to results.

Key Words: Unit root test, robust regression methods, OLS method, Dickey-Fuller Test, simulation.

¹ Econometrics, Cumhuriyet University, Sivas, Turkey.

² Econometrics, Suleyman Demirel University, Isparta, Turkey.

A Meta Analysis Approach on Work – Family Balance and Job Satisfaction

Ebru Özgür Güler¹, Didem Keskin²

¹Associate Professor Dr., Çukurova University, Department of Econometrics, Adana, Türkiye

²Graduate Student, Çukurova University, Department of Econometrics, Adana, Türkiye

Abstract

Nowadays, the number of scientific studies is increasing rapidly. It is noteworthy that in independent studies conducted on a specific subject, different results are often reached. In order to interpret this mass of information and lead to new studies, comprehensive and reliable studies are needed. For this reason, rapidly increasing researches with different findings have increased the need for meta-analysis. Meta-analysis is a quantitative method used to combine the results of multiple studies into a single result. It provides researchers with the opportunity to combine and analyse quantitative findings of a large number of studies, and at the same time draw conclusions about the overall impact between studies by combining quantitative results from many studies.

In this study, the relationship between Job - Family Balance and Job Satisfaction, which has become an important research area in recent years and which has been applied several times in the field of social sciences and which contains different findings, will be examined by using Meta-analysis method. In the conclusion part of the study, the results of similar studies related to the related topic will be combined and the research findings and analysis results that will reveal the relationship between work-family balance and job satisfaction will be combined and a synthesis will be formed and the results will be interpreted and suggestions will be made.

Key Words: Meta-analysis, work-family balance, job satisfaction.

POSTERS

The Effects of Time Interval on Test-Retest Reliability

Ilker Ercan¹, Fatma Ezgi Can^{2,3}, Guven Ozkaya¹, Mehmet Onur Kaya⁴

¹Department of Biostatistics, Faculty of Medicine, Bursa Uludag University, Bursa, Turkey

²Department of Biostatistics, Faculty of Medicine, Izmir Katip Celebi University, Izmir, Turkey

³Department of Biostatistics, Institute of Health Sciences, Bursa Uludag University, Bursa, Turkey

⁴Department of Biostatistics, Faculty of Medicine, Firat University, Elazig, Turkey

Abstract

Reliable scale is the scale that measure precisely. Since it is not possible to have an accurate measurement, increasing the reliability of the measurement will be possible by minimizing the error. Reliability of the scale can be examined different ways such as applying a scale once, applying a scale twice or applying two equivalent scales once.

The aim of the study is to determine the appropriate time interval between the applications of the scales by applying the scales more than once with the test-retest method.

In our study, two different scales, Self-Esteem Scale and Self-Decision Scale were considered. Each scale was reapplied immediately, 1 hour, 1 week, 2 weeks and 1 month after the first application. The reliability of each scale was determined by Cronbach alpha coefficient. In addition, the correlation coefficient between the first and last application scale total scores was investigated for each application time.

The reliability of the self-esteem scale was obtained as 0.902 (n = 101) when applied “immediately”, 0.946 (n = 119) when applied after “1 hour”, 0.829 (n = 71) after “1 week”, 0.871 (n=72) after “2 weeks” and 0.897 (n = 59) after “1 month”.

The reliability of the self-decision scale was obtained as 0.865 (n = 103) when applied “immediately”, 0.909 (n = 105) when applied after “1 hour”, 0.903 (n = 99) after “1 week”, 0.889 (n=81) after “2 weeks” and 0.897 (n = 59) after “1 month”.

The definition of the scales can be expressed as considerably reliable in the applying two scales at different times. There were no differences in defining scale reliability levels when applying different scales at different times.

Key Words: Test-retest, reliability, measurement.

Penalized Power Properties of Computation Methods for the Generalized F-test

Berna Yazıcı¹, Mustafa Çavuş¹

¹Department of Statistics, Eskisehir Technical University, Eskisehir, Turkey

Abstract

Weerahandi (1995)¹ proposed Generalized F (GF) test based on the generalized p-value method for testing equality of group means as alternative of Classical F (CF) test in case of unequal variances. Gamage and Weerahandi (1998)² conducted a Monte-Carlo simulation study to compare the performance of GF test with the counterparts and CF test. They concluded that the GF test is superior than the CF test in terms of Type I error probability and power. Thus, GF test is used by many researchers in their applications. Also, it is modified by Cavus et al. (2017)³ for non-normality caused by outliers and Cavus et al. (2018)⁴ showed its usefulness in the real data applications.

Weerahandi (2004)⁵ used two computation methods for the GF test after proposing it. These methods based on chi-squared and beta random samples. It is mentioned that the number of treatments being compared is very large, the results of these methods are approximated by Monte-Carlo simulations. However, the using of chi-squared based method is suggested in Weerahandi (2004). In the literature, Hartung et al. (2008)⁶ and Krishnamoorthy et al. (2007)⁷ used the chi-squared based computation method. In this study, the difference between the results of these computation methods are investigated by a Monte-Carlo simulation study. The properties of the computation methods for the GF test are obtained in terms of penalized power.

Key Words: Generalized F-test, penalized power, computation methods, Monte-Carlo.

¹ Weerahandi, S. (1995). ANOVA under unequal variances. *Biometrics*, 51, 589–599. doi.org/10.2307/252947.

AUTHOR INDICES

A

Abdulhamid · 216
Abdulqader · 153
Abiad · 49, 50
Abioglu · 243
Acar · 139
Acaravcı · 196
Adalı · 182, 184
Aftab · 62
Ahmadian · 159
Ahmed · 108
Akalın · 192
Akdoğan · 249
Akdur · 102
Akgüller · 66, 124, 125, 127
Akhanlı · 61
Akin · 120
Akkaya · 44, 79
Akkoyunlu · 42
Akkuş · 23, 36, 155
Akpolat · 186, 214
Aksu · 197
Alın · 253
Alpaslan · 88, 90
Alpkoçak · 156
Al-Sharadqah · 178
Altan · 201
Altınırnak · 228
Altıntaş · 69
Altinisik · 51
Apaydın · 69, 213
Ariöz · 75
Arslan · 37, 77, 130, 171
Artan · 196, 221
Asfha · 229
Aslam · 43
Aşnaz · 93
Ataman · 220
Atasoy · 173, 228
Atukalp · 45
Ayan · 70
Aydın · 22

Ayhan · 190

B

Babacan · 240
Baghestani · 55
Baghfalaki · 232
Bahçecitapar · 207
Bakar · 114
Bakır · 168, 246
Bal · 177
Balaylar · 241
Balcı · 66, 124, 125, 127
Baltacı · 42
Barber · 60
Barbiero · 180
Bari · 31
Barlas · 183
Bas · 26, 27, 28, 86
Başarslan · 47
Başdoğan · 125
Başer · 69, 213
Başoğlu Kabran · 176
Baycan · 32
Bayrak · 44
Bekiroğlu · 152
Bekiryazici · 100
Beyaztaş · 253
Biçer · 158, 160
Bilen · 98
Birecikli · 241
Bohdalová · 103
Bozkus · 60

C

Cagcag Yolcu · 26, 27, 85
Cakmak · 64
Can · 142, 159, 233, 257
Canbolat · 82, 89
Canga · 120, 123
Cankaya · 51, 77



Celik · 81, 120, 123, 126
Ceritoğlu · 33
Cetin · 116
Chaudhry · 52
Chen · 6, 206
Cheng · 30
Chien · 30, 63
Choudhary · 245
Comic · 169, 189
Coskun · 64
Coşkun · 58

Ç

Çakmak · 122, 230
Çalık · 222, 223, 224
Çalışkan · 73
Çalışkanelli · 253
Çankaya · 185
Çavdar · 54
Çavuş · 258
Çelikoklar · 72
Çemrek · 244, 247
Çiçekli · 91
Çiftçi · 25
Çilengiroğlu · 238
Çoban · 80
Çolak · 92

D

Dalar · 87
Dalkilic · 150
Dedeakayoğulları · 115, 197
Degirmen · 175
Değirmen · 135, 194
Değirmenci · 70
Deliorman · 231
Demir · 173, 222, 223, 224, 228, 237
Demirbüken · 168
Demircioğlu · 179
Demirel · 166, 221
Demirhan · 104
Demiroğlu · 251
Deo · 245
Depren · 157

Derya · 87
Devrani · 136
Djokovic · 189
Dmitrovic · 209
Doğan · 200, 201
Doğaner · 199
Doğru · 37, 122
Dumlupınar · 98
Durak · 115, 171, 197
Durmaz · 137
Duru · 152

E

Efendioğlu · 137
Egrioglu · 26, 27, 28, 86
Eker · 90
Ekici · 185
Ekinci · 68
Eliyi · 58, 72, 170, 172
Emeç · 241
Epstein · 205
Erar · 177
Ercan · 159, 192, 257
Erdaş · 162, 164, 198
Erdem · 130, 131
Erdemir · 147
Erdiş · 246
Erdoğan · 196
Erginel · 139
Ergül · 146, 161
Ergün · 93, 162
Erol · 4
Ervural · 188
Erzengin · 40
Esemen · 144
Eskin · 167
Evkaya · 106
Evren · 67, 78
Eyduran · 120, 123

F

Fatima · 108
Fayyazı · 157
Firat · 35, 117



Firuzan · 219

Fokoue · 24

Inan · 187, 231

Inostroza-Quezada · 205

Ionescu · 49, 50

G

Gacki · 141

Ganioğlu · 195, 251

Ganjali · 232

Gazeloğlu · 211, 242

Gerger · 219

Gever · 112

Ghalayini · 248

Goyal · 245

Göktaş · 23, 36

Gözen · 200

Greguš · 103

Guler · 111

Güdücü · 166

Gülay · 152, 166

Güler · 38, 210, 239, 255

Gültekin · 167, 191

Gündoğdu · 254

Gündüz · 22, 24

Güney · 162

Güngör · 92, 95

Güray · 136

Gürdal · 96, 98

Gürler · 144

Güzel · 127

H

Habibullah · 8

Hacioğlu · 253

Hamurkaroğlu · 145

Hanaloğlu · 112, 200

Hasançebi · 71

Hayaloğlu · 221

Hosseini-Baharanchi · 55

Hou · 206

I

Ibrahim · 217

Ikizoglu · 173

i

İçen · 114

İçli · 91

İdin · 82, 89

İnan · 39

İyit · 119

J

Jaiteh · 246

Josaphat · 65

K

Kaçıranlar · 38, 56, 99, 129

Kaçmaz · 151

Kahraman · 220

Kalaycı · 196

Kalaycıoğlu · 61

Kamal · 52

Kan-Kilinc · 229

Kapucu · 186

Kara · 166, 250

Karaağaoğlu · 230

Karabulut · 68

Karaçoban · 194

Karaibrahimoglu · 59, 190

Karakaş · 222, 223, 224

Karakoca · 126

Karakoç · 132

Karamaşa · 76, 228

Kaya · 66, 77, 240, 257

Kayaalp · 47

Kayacan · 113

Kesemen · 100, 149

Kesin · 159

Keskin · 91, 92, 95, 255

Khaniyev · 100, 110

Kılınçkaya · 96

Kırkağaç · 147



Kızılkaya · 158, 160

Kızılkum · 161

Kiraz · 233

Kirisci · 85

Koçak · 28

Kol · 124

Konar · 208

Korucu · 154

Krasachat · 118

Kulawik · 141

Kuran · 181

Kurnaz · 46, 106

Kurniawati · 30

Kutlu · 69

Kuvat · 23

Küçük · 71

Küçükbay · 174

L

Lee · 28, 30

Lesina · 209

Lyu · 206

M

Mangal · 74

Månsson · 101

Maral · 200

Meher · 245

Mehmood · 43

Merdan · 100

Mermi · 36

Miçooğulları · 135

Mirezi · 38

O

O'Brien · 9

Okidoi · 138

Ozkaya · 257

Ozkok · 81

Ozonur · 121

Ozturk · 59

Ö

Ölçer · 164, 198

Özbay · 53, 99, 101, 129

Özcan · 170, 172

Özçil · 182, 184

Özdemir · 171, 233, 237

Özen · 156, 193

Özener · 185, 218

Özkan · 129

Öztaş · 182, 184

P

Pak · 149

Pehlivan · 94

Peker · 140

Pekmezci · 83

Plazola · 109

Q

Qasim · 56

R

Radojicic · 169

Raharjanti · 41

Riaz · 52, 108

Rizvi · 74

Rulli · 178

S

Sadiq · 43

Sağır · 235

Sağlam · 235

Sahin · 78

Sahiner · 215, 216, 217

Saltık · 194

Saltik · 175

Sancar · 187

Sandoval · 109

Saraçlı · 211, 242

Saraçoğlu · 54



Satıcı · 155
Sayılır · 227
Selamlar · 212
Selçuk-Kestel · 204
Selec · 209
Senvar · 64
Setyowati · 63
Sevgenler · 123
Sevinc · 33, 110
Sevinç · 23, 106, 144, 252
Sezer · 126, 143
Shaik · 107
Sırgancı · 133, 134
Silahtaroglu · 82, 89
Silva · 138
Sohail · 62
Sokołowski · 2
Soyu · 175
Subat · 151
Suleman · 108
Syuhada · 65

Ş

Şahin · 94, 168, 193
Şalk · 226
Şener · 128

T

Tak · 39
Tanaka · 57, 67
Tank · 213
Tatlídil · 34
Taylan · 212
Temiz · 64
Terzi · 71
Tiryaki · 149
Toker · 53, 56, 99, 101
Toktaş · 142, 143, 163
Toluk · 192
Toptancı · 139
Tuna · 67
Tunca · 155
Tuncer · 79
Turhan · 174

Tuş · 182, 184
Türkay · 254
Türkşen · 132, 203

U

Ulu · 202
Ulucan · 25
Usta · 234, 236
Uyumaz · 133, 134

Ü

Üçüş · 131
Ünlü · 176
Ünvan · 29, 105
Üstündağ Şiray · 48, 56, 99

V

Vahaplar · 166
Vallim · 138
Vallim-Filho · 138
Veysikarani · 210
Vukmirovic · 169, 189

W

Wang · 227
Werner · 249

Y

Yabacı · 159
Yalaz · 165
Yaman · 150
Yandı · 133, 134
Yanık · 34
Yapar · 3, 226
Yardibi · 35
Yavuz · 116, 146, 161, 238
Yazıcı · 258
Yazıcıoğlu · 191
Yesildere · 234, 236



Yıldırım · 86

Yıldırım Külekci · 204

Yıldız · 117, 244, 247

Yıldızbaşı · 75

Yılmaz · 84, 225

Yilmaztürk · 82, 89

Yiğit · 97

Yiğiter · 145

Yildirim · 88, 90

Yılmaz · 215, 216

Yolcu · 26, 27, 84

Yonar · 119

Yücesoy · 235

