UNIVERSITÀ DI PISA

Sant'Anna
Scuola Universitaria Superiore Pisa

Consiglio Nazionale delle Ricerche

Università degli Studi
di Cagliari

Book of Short Papers SIS 2021

SIS2021
PISA, 21-25 GIUGNO 2021

SIS
Società
Italiana di
Statistica

Editors: Cira Perna, Nicola Salvati and Francesco Schirripa Spagnolo

BANCA DI PISA
E FORNACETTE
CREDITO COOPERATIVO
Banca aderente al Gruppo Bancario Cooperativo Iccrea

TStat
Distribuzione Software | Formazione Professionale
Statistica | Economia | Finanza | Biostatistica | Epidemiologia
Sanità Pubblica | Scienze Sociali
www.tstat.it | www.tstattraining.eu

Copyright © 2021

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891927361

Contents

Preface **XIX**

1 Plenary Sessions **1**

1.1 Citizen data, and citizen science: a challenge for official statistics. 2
Monica Pratesi

2 Specialized Sessions **8**

2.1 A glimpse of new data and methods for analysing a rapidly changing population 9

2.1.1 The diffusion of new family patterns in Italy: An update. 10
Arnstein Aassve, Letizia Mencarini, Elena Pirani and Daniele Vignoli

2.1.2 Causes of death patterns and life expectancy: looking for warning signals. 16
Stefano Mazzuco, Emanuele Aliverti, Daniele Durante and Stefano Campostrini

2.2 Advances in ecological modelling 22

2.2.1 A Bayesian joint model for exploring zero-inflated bivariate marine litter data. 23
Sara Martino, Crescenza Calculli and Porzia Maiorano

2.3 Advances in environmental statistics 29

2.3.1 Bayesian small area models for investigating spatial heterogeneity and factors affecting the amount of solid waste in Italy. 30
Crescenza Calculli and Serena Arima

2.3.2 A spatial regression model for for predicting abundance of lichen functional groups. 36
Pasquale Valentini, Francesca Fortuna, Tonio Di Battista and Paolo Giordani

2.4 Advances in preference and ordinal data theoretical improvements and applications 42

2.4.1 Boosting for ranking data: an extension to item weighting. 43
Alessandro Albano, Mariangela Sciandra and Antonella Plaia

2.4.2 An Extended Bradley-Terry Model For The Analysis Of Financial Data. 49
Alessio Baldassarre, Elise Dusseldorp and Mark De Rooij

2.5 Business system innovation, competitiveness, productivity and internationalization 55

2.5.1 An analysis of the dynamics of the competitiveness for some European Countries. 56
Andrea Marletta, Mauro Mussini and Mariangela Zenga

2.5.2 National innovation system and economic performance in EU. An analysis using composite indicators. 62
Alessandro Zeli

2.6 Challenges for observational studies in modern biomedicine 68

2.6.1 Data integration: a Statistical view. 69
Pier Luigi Conti

2.6.2 Exploring patients’ profile from COVID-19 case series data: beyond standard statistical approaches. 75
Chiara Brombin, Federica Cugnata, Pietro E. Cippà, Alessandro Ceschi, Paolo Ferrari and Clelia di Serio

2.6.3 On the statistics for some pivotal anti-COVID-19 vaccine trials. 81
Mauro Gasparini

2.7 Data Science for Industry 4.0 (ENBIS) 87

2.7.1 Sample selection from a given dataset to validate machine learning models. 88
Bertrand Iooss

2.7.2 Reliable data-drive modelling and optimisation of a batch reactor using bootstrap aggregated deep belief networks. 94
Changhao Zhu and Jie Zhang

2.8 Integration of survey with alternative sources of data 100

2.8.1 A parametric empirical likelihood approach to data matching under nonignorable sampling and nonresponse. 101
Daniela Marella and Danny Pfeffermann

2.8.2 Survey data integration for regression analysis using model calibration. 107

2.8.3 Latent Mixed Markov Models for the Production of Population Census Data on Employment. 112
Danila Filipponi, Ugo Guarnera and Roberta Varriale

2.9 Media, social media and demographic behaviours 118

2.9.1 Monitoring the Numbers of European Migrants in the United Kingdom using Facebook Data. 119
Francesco Rampazzo, Jakub Bijak, Agnese Vitali, Ingmar Weber and Emilio Zagheni

2.10 New developments in ensemble methods for classification 125

2.10.1 An alternative approach for nowcasting economic activity during COVID-19 times. 126
Alessandro Spelta and Paolo Pagnottoni

2.10.2 Assessing the number of groups in consensus clustering by pivotal methods. 132
Roberta Pappadà, Francesco Pauli and Nicola Torelli

2.10.3 Clustering of data recorded by Distributed Acoustic Sensors to identify vehicle passage and typology. 138
Antonio Balzanella and Stefania Nacchia

2.11 New developments in latent variable models 144

2.11.1 A Hidden Markov Model for Variable Selection with Missing Values. 145
Fulvia Pennoni, Francesco Bartolucci, and Silvia Pandolfi

2.11.2 Comparison between Different Likelihood Based Estimation Methods in Latent Variable Models for Categorical Data. 151
Silvia Bianconcini and Silvia Cagnone

2.11.3 A Comparison of Estimation Methods for the Rasch Model. 157
Alexander Robitzsch

2.12 New issues on multivariate and univariate quantile regression 163

2.12.1 Directional M-quantile regression for multivariate dependent outcomes. 164
Luca Merlo, Lea Petrella and Nikos Tzavidis

2.13 Semi-parametric and non-parametric latent class analysis 170

2.13.1 Stepwise Estimation of Multilevel Latent Class Models. 171
Zsuzsa Bakk, Roberto di Mari, Jennifer Oser and Jouni Kuha

2.13.2 Distance learning, stress and career-related anxiety during the Covid-19 pandemic: a students perspective analysis. 177
Alfonso Iodice D'Enza, Maria Iannario, Rosaria Romano

2.13.3 A Tempered Expectation-Maximization Algorithm for Latent Class Model Estimation. 183
Luca Brusa, Francesco Bartolucci and Fulvia Pennoni

2.14 Statistics for finance high frequency data, large dimension and networks 189

2.14.1 The Italian debt not-so-flash crash. 190
Maria Flora and Roberto Reno'

3 Solicited Sessions 197

3.1 Advances in social indicators research and latent variables modelling in social sciences 198

3.1.1 A composite indicator to measure frailty using administrative healthcare data. 199
Margherita Silan, Rachele Brocco and Giovanna Boccuzzo

3.1.2 Clusters of contracting authorities over time: an analysis of their behaviour based on procurement red flags. 205
Simone Del Sarto, Paolo Coppola and Matteo Troìa

3.1.3 An Application of Temporal Poset on Human Development Index Data. 211
Leonardo Salvatore Alaimo, Filomena Maggino and Emiliano Seri

3.1.4 The SDGs System: a longitudinal analysis through PLS-PM. 217
Rosanna Cataldo, Maria Gabriella Grassia and Laura Antonucci

3.2 Changes in the life course and social inequality 223

3.2.1 Heterogeneous Income Dynamics: Unemployment Consequences in Germany and the US. 224
Raffaele Grotti

3.2.2 In-work poverty in Germany and in the US: The role of parity progression. 230
Emanuela Struffolino and Zachary Van Winkle Z.

3.2.3 Parenthood, education and social stratification. An analysis of female occupational careers in Italy. 236
Gabriele Ballarino and Stefano Cantalini

3.3 Composition in the Data Science Era 242

3.3.1 Can we Ignore the Compositional Nature of Compositional Data by using Deep Learning Aproaches? 243
Matthias Templ

3.3.2 Principal balances for three-way compositions. 249
Violetta Simonacci

3.3.3 Robust Regression for Compositional Data and its Application in the Context of SDG. 255
Valentin Todorov and Fatemah Alqallaf

3.4 Evaluation of undercoverage for censuses and administrative data 261

3.4.1 Spatially balanced indirect sampling to estimate the coverage of the agricultural census. 262
Federica Piersimoni, Francesco Pantalone and Roberto Benedetti

3.4.2 Next Census in Israel: Strategy, Estimation and Evaluation. 268
Danny Pfeffermann

3.4.3 Administrative data for population counts estimations in Italian Population Census. 274
Antonella Bernadini, Angela Chieppa, Nicola Cibella and Fabrizio Solari

3.4.4 LFS non response indicators for population register overcoverage estimation. 279
Lorella Di Consiglio, Stefano Falorsi

3.5 Excesses and rare events in complex systems 285

3.5.1 Space-time extreme rainfall simulation under a geostatistical approach. 286
Gianmarco Callegher, Carlo Gaetan, Noemie Le Carrer and Ilaria Prosdocimi

3.6 Hierarchical forecasting and forecast combination 292

3.6.1 Density calibration with consistent scoring functions. 293
Roberto Casarin and and Francesco Ravazzolo

3.6.2 Forecasting combination of hierarchical time series: a novel method with an application to CoVid-19. 298
Livio Fenga

3.7 Household surveys for policy analysis 304

3.7.1 Did the policy responses to COVID-19 protect Italian households' incomes? Evidence from survey and administrative data. 305
Maria Teresa Monteduro, Dalila De Rosa and Chiara Subrizi

3.8 Learning analytics methods and applications 311

3.8.1 Open-Source Automated Test Assembly: the Challenges of Large-Sized Models. 312
Giada Spaccapanico Proietti

3.8.2 How Much Tutoring Activities May Improve Academic Careers of At-Risk Students? An Evaluation Study. 318
Marta Cannistra, Tommaso Agasisti, Anna Maria Paganoni and Chiara Masci

3.8.3 Composite–based Segmentation Trees to Model Learners' performance. 324
Cristina Davino and Giuseppe Lamberti

3.8.4 Test-taking Effort in INVALSI Assessments. 330
Chiara Sacco

3.9 Light methods for hard problems 336

3.9.1 Fast Divide-and-Conquer Strategies to Solve Spatial Big Data Problems. 337
Michele Peruzzi

3.9.2 Application of hierarchical matrices in spatial statistics. 343
Anastasiia Gorshechnikova and Carlo Gaetan

3.10 Management and statistics in search for a common ground (AIDEA) 349

3.10.1 Customer Segmentation: it's time to make a change. 350
Fabrizio Laurini, Beatrice Luceri and Sabrina Latusi

3.10.2 Multivariate prediction models: Altman's ZScore and CNDCEC's sectoral indicators. 356
Alessandro Danovi, Alberto Falini and Massimo Postiglione

3.10.3 Comparing Entrepreneurship and Perceived Quality of Life in the European Smart Cities: a "Posetic" Approach. 362
Lara Penco, Enrico Ivaldi and Andrea Ciacci

3.10.4 The Relationship between Business Economics and Statistics: Taking Stock and Ways Forward. 368
Amedeo Pugliese

3.11 Mathematical methods and tools for finance and insurance (AMASES) 373

3.11.1 On the valuation of the initiation option in a GLWB variable annuity. 374
Anna Rita Bacinello and Pietro Millossovich

3.11.2 Modern design of life annuities in view of longevity and pandemics. 380
Annamaria Olivieri

3.11.3 Risk Management from Finance to Production Planning: An Assembly-to-Order Case Study. 386
Paolo Brandimarte, Edoardo Fadda and Alberto Gennaro

3.11.4 Some probability distortion functions in behavioral portfolio selection. 392
Diana Barro, Marco Corazza and Martina Nardonthors

3.12 Multiple system estimation 398

3.12.1 Multiple Systems Estimation in the Presence of Censored Cells. 399
Ruth King, Oscar Rodriguez de Rivera Ortega and Rachel McCrea

3.12.2 Bayesian population size estimation by repeated identifications of units. A semi-parametric mixture model approach. 405
Tiziana Tuoto, Davide Di Cecco and Andrea Tancredi

3.13 Network sampling and estimation 411

3.13.1 Targeted random walk sampling. 412
Li-Chun Zhang

3.13.2 Estimation of poverty measures in Respondent-driven sampling. 418
María del Mar Rueda, Ismael Sànchez-Borrego and Héctor Mullo

3.13.3 Sampling Networked Data for Semi-Supervised Learning Algorithms. 423
Simone Di Zio, Lara Fontanella, Francesco Pantalone and Federica Piersimoni

3.13.4 A sequential adaptive sampling scheme for rare populations with a network structure. 429
Emilia Rocco

3.14 New perspectives on multidimensional child poverty 435

3.14.1 Estimating uncertainty for child poverty indicators: The Case of Mediterranean Countries. 436
Ilaria Benedetti, Federico Crescenzi and Riccardo De Santis

3.14.2 Child poverty and government social spending in the European Union during the economic crisis. 442
Angeles Sánchez and María Navarro

3.14.3 The Children’s Worlds Study: New perspectives on children’s deprivation research. 448
Caterina Giusti and Antoanneta Potsi

3.14.4 The impact of different definition of “households with children” on deprivation measures: the case of Italy. 454
Laura Neri and Francesca Gagliardi

3.15 Perspectives in social network analysis applications 460

3.15.1 A comparison of student mobility flows in Eramus and Erasmus+ among countries. 461
Kristijan Breznik, Giancarlo Ragozini and Marialuisa Restaino

3.15.2 Network-based approach for the analysis of LexisNexis news database. 467
Carla Galluccio and Alessandra Petrucci

3.15.3 A multiplex network approach to study Italian Students’ Mobility. 473
Ilaria Primerano, Francesco Santelli and Cristian Usala

3.15.4 Ego-centered Support Networks:a Cross-national European Comparison. 479
Emanuela Furfaro, Elvira Pelle, Giulia Rivellini and Susanna Zaccarin

3.16 Statistical analysis of energy data 485
3.16.1 Machine learning models for electricity price forecasting. 486
Silvia Golia, Luigi Grossi, Matteo Pelagatti
3.16.2 The impact of hydroelectric storage in the Italian power market. 492
Filippo Beltrami
3.16.3 Jumps and cojumps in electricity price forecasting. 498
Peru Muniain, Aitor Ciarreta and Ainhoa Zarraga

3.17 Statistical methods and models for the analysis of sports data 507
3.17.1 Football analytics: a Higher-Order PLS-SEM approach to evaluate players' performance. 508
Mattia Cefis and Maurizio Carpita
3.17.2 Bayesian regularized regression of football tracking data through structured factor models. 514
Lorenzo Schiavon and Antonio Canale
3.17.3 A dynamic matrix-variate model for clustering time series with multiple sources of variation. 520
Mattia Stival
3.17.4 Evaluating football players' performances using on-the-ball data. 526
David Dandolo

3.18 The social and demographic consequences of international migration in Western societies 532
3.18.1 Employment and job satisfaction of immigrants: the case of Campania (Italy). 533
Alessio Buonomo, Stefania Capecchi, Francesca Di Iorio and Salvatore Strozza
3.18.2 Social stratification of migrants in Italy: class reproduction and social mobility from origin to destination. 539
Giorgio Piccitto, Maurizio Avola and Nazareno Panichella

3.19 Well-being, healthcare, integration measurements and indicators (SIEDS) 545
3.19.1 A Composite Index of Economic Well-being for the European Union Countries. 546
Andrea Cutillo, Matteo Mazziotta and Adriano Pareto
3.19.2 Poverty orderings and TIP curves: an application to the Italian regions. 552
Francesco M. Chelli, Mariateresa Ciommi and Chiara Gigliarano

4 Contributed Sessions 558

4.1 Advances in clinical trials 559
4.1.1 Quantitative depth-based [18F]FMCH-avid lesion profiling in prostate cancer treatment. 560
Lara Cavinato, Alessandra Ragni, Francesca Ieva, Martina Sollini, Francesco Bartoli and Paola A. Erba
4.1.2 Modelling longitudinal latent toxicity profiles evolution in osteosarcoma patients. 566
Marta Spreafico, Francesca Ieva and Marta Fiocco
4.1.3 Information borrowing in phase II basket trials: a comparison of different designs. 572
Marco Novelli
4.1.4 Q-learning Estimation Techniques for Dynamic Treatment Regime. 578
Simone Bogni, Debora Slanzi and Matteo Borrotti
4.1.5 Sample Size Computation for Competing Risks Survival Data in GS-Design. 584
Mohammad Anamul Haque and Giuliana Cortese

4.2 Advances in neural networks 590

4.2.1 Linear models vs Neural Network: predicting Italian SMEs default. 591
Lisa Crosato, Caterina Liberati and Marco Repetto

4.2.2 Network estimation via elastic net penalty for heavy-tailed data. 596
Davide Bernardini, Sandra Paterlini and Emanuele Taufer

4.2.3 Neural Network for statistical process control of a multiple stream process with an application to HVAC systems in passenger rail vehicles. 602
Gianluca Sposito, Antonio Lepore, Biagio Palumbo and Giuseppe Giannini

4.2.4 Forecasting air quality by using ANNs. 608
Annalina Sarra, Adelia Evangelista, Tonio Di Battista and Francesco Bucci

4.3 Advances in statistical methods 614

4.3.1 Robustness of Fractional Factorial Designs through Circuits. 615
Roberto Fontana and Fabio Rapallo

4.3.2 Multi-objective optimal allocations for experimental studies with binary outcome. 621
Alessandro Baldi Antognini, Rosamarie Frieri, Marco Novelli and Maroussa Zagoraiou

4.3.3 Analysis of three-way data: an extension of the STATIS method. 627
Laura Bocci and Donatella Vicari

4.3.4 KL-optimum designs to discriminate models with different variance function. 633
Alessandro Lanteri, Samantha Leorato and Chiara Tommasi

4.3.5 Riemannian optimization on the space of covariance matrices. 639
Jacopo Schiavon, Mauro Bernardi and Antonio Canale

4.4 Advances in statistical methods and inference 645

4.4.1 Estimation of Dirichlet Distribution Parameters with Modified Score Functions. 646
Vincenzo Gioia and Euloge Clovis Kenne Pagui

4.4.2 Confidence distributions for predictive tail probabilities. 652
Giovanni Fonseca, Federica Giummolè and Paolo Vidoni

4.4.3 Impact of sample size on stochastic ordering tests: a simulation study. 658
Rosa Arboretti, Riccardo Ceccato, Luca Pegoraro and Luigi Salmaso

4.4.4 On testing the significance of a mode. 664
Federico Ferraccioli and Giovanna Menardi

4.4.5 Hommel BH: an adaptive Benjamini-Hochberg procedure using Hommel's estimator for the number of true hypotheses. 670
Chiara G. Magnani and Aldo Solari

4.5 Advances in statistical models 676

4.5.1 Specification Curve Analysis: Visualising the risk of model misspecification in COVID-19 data. 677
Venera Tomaselli, Giulio Giacomo Cantone and Vincenzo Miracula

4.5.2 Semiparametric Variational Inference for Bayesian Quantile Regression. 683
Cristian Castiglione and Mauro Bernardi

4.5.3 Searching for a source of difference in undirected graphical models for count data – an empirical study. 689
Federico Agostinis, Monica Chiogna, Vera Djordjilović, Luna Pianesi and Chiara Romualdi

4.5.4 Snipped robust inference in mixed linear models. 695
Antonio Lucadamo, Luca Greco, Pietro Amenta and Anna Crisci

4.6 Advances in time series 701

4.6.1 A spatio-temporal model for events on road networks: an application to ambulance interventions in Milan. 702
Andrea Gilardi and Riccardo Borgoni and Jorge Mateu

4.6.2 Forecasting electricity demand of individual customers via additive stacking. 708
Christian Capezza, Biagio Palumbo, Yannig Goude, Simon N. Wood and Matteo Fasiolo

4.6.3 Hierarchical Forecast Reconciliation on Italian Covid-19 data. 714
Andrea Marcocchia, Serena Arima and Pierpaolo Brutti

4.6.4 Link between Threshold ARMA and tdARMA models. 720
Guy Mélard and Marcella Niglio

4.7 Bayesian nonparametrics 726

4.7.1 Bayesian nonparametric prediction: from species to features. 727
Lorenzo Masoero, Federico Camerlenghi, Stefano Favaro and Tamara Broderick

4.7.2 A framework for filtering in hidden Markov models with normalized random measures. 733
Filippo Ascolani, Antonio Lijoi, Igor Prünster and Matteo Ruggiero

4.7.3 On the convex combination of a Dirichlet process with a diffuse probability measure. 739
Federico Camerlenghi, Riccardo Corradin and Andrea Ongaro

4.7.4 Detection of neural activity in calcium imaging data via Bayesian mixture models. 745
Laura D'Angelo, Antonio Canale, Zhaoxia Yu and Michele Guindani

4.8 Clustering for complex data 751

4.8.1 Clustering categorical data via Hamming distance. 752
Edoardo Filippi-Mazzola, Raffaele Argiento and Lucia Paci

4.8.2 Penalized model-based clustering for three-way data structures. 758
Andrea Cappozzo, Alessandro Casa, and Michael Fop

4.8.3 Does Milan have a smart mobility? A clustering analysis approach. 764
Nicola Cornali, Matteo Seminati, Paolo Maranzano and Paola M. Chiodini

4.8.4 A Fuzzy clustering approach for textual data. 770
Irene Cozzolino, Maria Brigida Ferraro and Peter Winker

4.8.5 Valid Double-Dipping via Permutation-Based Closed Testing. 776
Anna Vesely, Livio Finos, Jelle J. Goeman and Angela Andreella

4.9 Data science for complex data 782

4.9.1 Text mining on large corpora using Taltac4: An explorative analysis of the USPTO patents database. 783
Pasquale Pavone, Arianna Martinelli and Federico Tamagni

4.9.2 Emotion pattern detection on facial videos using functional statistics. 789
Rongjiao Ji, Alessandra Micheletti, Natasa Krklec Jerinkic and Zoranka Desnica

4.9.3 The spread of contagion on Twitter: identification of communities analysing data from the first wave of the COVID-19 epidemic. 795
Gianni Andreozzi, Salvatore Pirri, Giuseppe Turchetti and Valentina Lorenzoni

4.9.4 Composition-on-Function Regression Model for the Remote Analysis of Near-Earth Asteroids. 801
Mara S. Bernardi, Matteo Fontana, Alessandra Menafoglio, Alessandro Pisello, Massimiliano Porreca, Diego Perugini and Simone Vantini

4.9.5 Determinants of football coach dismissal in Italian League Serie A. 805
Francesco Porro, Marialuisa Restaino, Juan Eloy Ruiz-Castro and Mariangela Zenga

4.10 Data science for unstructured data 810

4.10.1 Identification and modeling of stop activities at the destination from GPS tracking data. 811
Nicoletta D'Angelo, Giada Adelfio, Antonino Abbruzzo and Mauro Ferrante

4.10.2 A generalization of derangement. 817
Maurizio Maravalle and Ciro Marziliano

4.10.3 Analysis of clickstream data with mixture hidden markov models. 823
Furio Urso, Antonino Abbruzzo and Maria Francesca Cracolici

4.10.4 Using Google Scholar to measure the credibility of preprints in the COVID-19 Open Research Dataset (CORD-19). 829
Manlio Migliorati, Maurizio Carpita, Eugenio Brentari

4.10.5 Mobile phone use while driving: a Structural Equation Model to analyze the Behavior behind the wheel. 835
Carlo Cavicchia and Pasquale Sarnacchiaro

4.11 Demographic analysis 841

4.11.1 Life expectancy in the districts of Taranto. 842
Stefano Cervellera, Carlo Cusatelli and Massimiliano Giacalone

4.11.2 Family size and Human Capital in Italy: a micro-territorial analysis. 848
Gabriele Ruiu, Marco Breschi and Alessio Fornasin

4.11.3 Estimate age-specific fertility rates from summary demographic measures. An Indirect Model Levering on Deep Neural Network. 854
Andrea Nigri

4.11.4 Patterns in the relation between causes of death and gross domestic product. 860
Andrea Nigri and Federico Crescenzi

4.11.5 Locally sparse functional regression with an application to mortality data. 866
Mauro Bernardi, Antonio Canale, Marco Stefanucci

4.12 Environmental statistics 871

4.12.1 A Distribution-Free Approach for Detecting Radioxenon Anomalous Concentrations. 872
Michele Scagliarini, Rosanna Gualdi, Giuseppe Ottaviano, Antonietta Rizzo and Franca Padoani

4.12.2 Ecosud Car, a novel approach for the predictive control of the territory. 878
Giacomo Iula, Massimo Dimo, Saverio Gianluca Crisafulli, Marco Vito Calciano, Vito Santarcangelo and Massimiliano Giacalone

4.12.3 Effect of ties on the empirical copula methods for weather forecasting. 884
Elisa Perrone, Fabrizio Durante and Irene Schicker

4.12.4 Spatio-temporal regression with differential penalization for the reconstruction of partially observed signals. 890
Eleonora Arnone and Laura M. Sangalli

4.12.5 Sea Surface Temperature Effects on the Mediterranean Marine Ecosystem: a Semiparametric Model Approach. 895
Claudio Rubino, Giacomo Milisenda, Antonino Abbruzzo, Giada Adelfio, Mar Bosch-Belmar, Francesco Colloca, Manfredi Di Lorenzo and Vita Gancitano

4.13 Functional data analysis 901

4.13.1 Remote Analysis of Chapas Stops in Maputo from GPS data: a Functional Data Analysis Approach. 902
Agostino Torti, Davide Ranieri and Simone Vantini

4.13.2 A Conformal approach for functional data prediction. 907
Jacopo Diquigiovanni, Matteo Fontana and Simone Vantini

4.13.3 Block testing in covariance and precision matrices for functional data analysis. 911
Marie Morvan, Alessia Pini, Madison Giacofci and Valerie Monbet

4.13.4 Analysing contributions of ages and causes of death to gender gap in life expectancy using functional data analysis. 917
Alessandro Feraldi, Virginia Zarulli, Stefano Mazzuco and Cristina Giudici

4.13.5 Supervised classification of ECG curves via a combined use of functional data analysis and random forest to identify patients affected by heart disease. 923
Fabrizio Maturo and Rosanna Verde

4.14 Mixture models 929

4.14.1 Alternative parameterizations for regression models with constrained multivariate responses. 930
Roberto Ascari, Agnese Maria Di Brisco, Sonia Migliorati and Andrea Ongaro

4.14.2 Spatially dependent mixture models with a random number of components. 936
Matteo Gianella, Mario Beraha and Alessandra Guglielmi

4.14.3 Finite mixtures of regression models for longitudinal data. 942
Marco Alfò and Roberto Rocci

4.14.4 Mixtures of regressions for size estimation of heterogeneous populations. 948
Gianmarco Caruso

4.14.5 Finite mixtures of regressions with random covariates using multivariate skewed distributions. 954
Salvatore D. Tomarchio, Michael P.B. Gallaugher, Antonio Punzo and Paul D. McNicholas

4.15 New applications of regression models 960

4.15.1 The Shapley-Lorenz decomposition approach to mitigate cyber risks. 960
Paolo Giudici and Emanuela Raffinetti

4.15.2 A spatially adaptive estimator for the function-on-function linear regression model with application to the Swedish Mortality dataset. 967
Fabio Centofanti, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini

4.15.3 POSetR: a new computationally efficient R package for partially ordered data. 972
Alberto Arcagni, Alessandro Avellone and Marco Fattore

4.15.4 Multi Split Conformal Prediction. 978
Aldo Solari and Vera Djordjilović

4.15.5 Changes in the consumption of fruits and vegetables among university students during master courses: an analysis of data automatically collected from cashier transactions. 984
Valentina Lorenzoni, Giuseppe Turchetti and Lucio Masserini

4.16 New challenges in clustering and classification techniques 990

4.16.1 A Dynamic Stochastic Block Model with infinite communities. 991
Roberto Casarin and Ovielt Baltodano Lòpez

4.16.2 Cross-Subject EEG Channel Selection for the Detection of Predisposition to Alcoholism. 997
Michela Carlotta Massi and Francesca Ieva

4.16.3 Some Issues on the Parameter Selection in the Spectral Methods for Clustering. 1003
Cinzia Di Nuzzo and Salvatore Ingrassia

4.16.4 The link-match tale: new microdata from unit level association. 1009
Riccardo D'Alberto, Meri Raggi and Daniela Cocchi

4.17 New developments in Bayesian methods 1015

4.17.1 Spatio-temporal analysis of the Covid-19 spread in Italy by Bayesian hierarchical models. 1016
Nicoletta D'Angelo, Giada Adelfio and Antonino Abbruzzo

4.17.2 Modelling of accumulation curves through Weibull survival functions. 1021
Alessandro Zito, Tommaso Rigon and David B. Dunson

4.17.3 Model fitting and Bayesian inference via power expectation propagation. 1026
Emanuele Degani, Luca Maestrini and Mauro Bernardi

4.17.4 Bayesian quantile estimation in deconvolution. 1032
Catia Scricciolo

4.17.5 Bayesian inference for discretely observed non-homogeneous Markov processes. 1038
Rosario Barone and Andrea Tancredi

4.18 New developments in composite indicators applications 1044
4.18.1 Building composite indicators in the functional domain: a suggestion for an evolutionary HDI. 1045
Francesca Fortuna, Alessia Naccarato and Silvia Terzi
4.18.2 Small Area Estimation of Inequality Measures via Simplex Regression. 1051
Silvia De Nicolò, Maria Rosaria Ferrante and Silvia Pacei
4.18.3 Relational Well-Being and Poverty in Italy Benessere relazionale e povertà in Italia. 1057
Elena Dalla Chiara and Federico Perali
4.18.4 A composite indicator to assess sustainability of agriculture in European Union countries. 1063
Alessandro Magrini and Francesca Giambona
4.18.5 Interval-Based Composite Indicators with a Triplex Representation: A Measure of the Potential Demand for the "Ristori" Decree in Italy. 1069
Carlo Drago

4.19 New developments in GLM theory and applications 1075
4.19.1 Variational inference for the smoothing distribution in dynamic probit models. 1076
Augusto Fasano and Giovanni Rebaudo
4.19.2 Interpretability and interaction learning for logistic regression models. 1082
Nicola Rares Franco, Michela Carlotta Massi, Francesca Ieva and Anna Maria Paganoni
4.19.3 Entropy estimation for binary data with dependence structures. 1088
Linda Altieri and Daniela Cocchi
4.19.4 A Comparison of Some Estimation Methods for the Three-Parameter Logistic Model. 1094
Michela Battauz and Ruggero Bellio
4.19.5 A statistical model to identify the price determinations: the case of Airbnb. 1100
Giulia Contu, Luca Frigau, Gian Paolo Zammarchi and Francesco Mola

4.20 New developments in social statistics analysis 1106
4.20.1 Data-based Evaluation of Political Agents Against Goals Scheduling. 1107
Giulio D'Epifanio
4.20.2 Local heterogeneities in population growth and decline. A spatial analysis for Italian municipalities. 1113
Federico Benassi, Annalisa Busetta, Gerardo Gallo and Manuela Stranges
4.20.3 The assessment of environmental and income inequalities. 1119
Michele Costa
4.20.4 Household financial fragility across Europe. 1125
Marianna Brunetti, Elena Giarda and Costanza Torricelli
4.20.5 Refugees' perception of their new life in Germany. 1131
Daria Mendola and Anna Maria Parroco

4.21 New perspectives in clinical trials 1137
4.21.1 Improved maximum likelihood estimator in relative risk regression. 1138
Euloge C. Kenne Pagui, Francesco Pozza and Alessandra Salvan
4.21.2 Development and validation of a clinical risk score to predict the risk of SARS-CoV-2 infection. 1144
Laura Savaré, Valentina Orlando and Giovanni Corrao
4.21.3 Functional representation of potassium trajectories for dynamic monitoring of Heart Failure patients. 1150
Caterina Gregorio, Giulia Barbati1 and Francesca Ieva
4.21.4 Effect of lung transplantation on the survival of patients with cystic fibrosis: IMaCh contribution to registry data. 1156
Cristina Giudici, Nicolas Brouard and Gil Bellis
4.21.5 Categories and Clusters to investigate Similarities in Diabetic Kidney Disease Patients. 1162
Veronica Distefano, Maria Mannone, Claudio Silvestri and Irene Poli

4.22 New perspectives in models for multivariate dependency 1168

4.22.1 Parsimonious modelling of spectroscopy data via a Bayesian latent variables approach. 1169
Alessandro Casa, Tom F. O'Callaghan and Thomas Brendan Mur

4.22.2 Bias reduction in the equicorrelated multivariate normal. 1175
Elena Bortolato and Euloge Clovis Kenne Pagui

4.22.3 Some results on identifiable parameters that cannot be identified from data. 1181
Christian Hennig

4.23 Novel approaches for official statistics 1187

4.23.1 Web data collection: profiles of respondents to the Italian Population Census. 1188
Elena Grimaccia, Gerardo Gallo, Alessia Naccarato, Novella Cecconi and Alessandro Fratoni

4.23.2 Trusted Smart Surveys: architectural and methodological challenges at a glance. 1194
Mauro Bruno, Francesca Inglese and Giuseppina Ruocco

4.23.3 On bias correction in small area estimation: An M-quantile approach. 1200
Gaia Bertarelli, Francesco Schirripa Spagnolo, Raymond Chambers and David Haziza

4.23.4 The address component of the Statistical Base Register of Territorial Entities. 1206
Davide Fardelli, Enrico Orsini and Andrea Pagano

4.23.5 A well-being municipal indicator using census data: first results. 1212
Massimo Esposito

4.24 Prior distribution for Bayesian analysis 1218

4.24.1 On the dependence structure in Bayesian nonparametric priors. 1219
Filippo Ascolani, Beatrice Franzolini, Antonio Lijoi, and Igor Prünster

4.24.2 Anisotropic determinantal point processes and their application in Bayesian mixtures. 1226
Lorenzo Ghilotti, Mario Beraha and Alessandra Guglielmi

4.24.3 Bayesian Screening of Covariates in Linear Regression Models Using Correlation Thresholds. 1232
Ioannis Ntzoufras and Roberta Paroli

4.25 Recent advances in clustering methods 1238

4.25.1 Biclustering longitudinal trajectories through a model-based approach. 1239
Francesca Martella, Marco Alfò and Maria Francesca Marino

4.25.2 Monitoring tools for robust estimation of Cluster Weighted models. 1245
Andrea Cappozzo and Francesca Greselin

4.25.3 Co-clustering Models for Spatial Transciptomics: Analysis of a Human Brain Tissue Sample. 1251
Andrea Sottosanti and Davide Risso

4.25.4 Graph nodes clustering: a comparison between algorithms. 1257
Ilaria Bombelli

4.26 Social demography 1263

4.26.1 Childcare among migrants: a comparison between Italy and France. 1264
Eleonora Trappolini, Elisa Barbiano di Belgiojoso, Stefania Maria Lorenza Rimoldi and Laura Terzera

4.26.2 Employment Uncertainty and Fertility in Italy: The Role of Union Formation. 1270
Giammarco Alderotti, Valentina Tocchioni and Alessandra De Rose

4.26.3 Determinants of union dissolution in Italy: Do children matter? 1276
Valentina Tocchioni, Daniele Vignoli, Eleonora Meli and Bruno Arpino

4.26.4 Working schedules and fathers' time with children: A Sequence Analysis. 1282
Annalisa Donno and Maria Letizia Tanturri

4.26.5 Correlates of the non-use of contraception among female university students in Italy. 1288
Annalisa Busetta, Alessandra De Rose and Daniele Vignoli

4.27 Social indicators applications and methods 1294

4.27.1 A logistic regression model for predicting child language performance. 1295
Andrea Briglia, Massimo Mucciardi and Giovanni Pirrotta

4.27.2 Subject-specific measures of interrater agreement for ordinal scales. 1301
Giuseppe Bove

4.27.3 A Tucker3 method application on adjusted-PMRs for the study of work-related mortality. 1307
Vittoria Carolina Malpassuti, Vittoria La Serra and Stefania Massari

4.27.4 Two case-mix adjusted indices for nursing home performance evaluation. 1313
Giorgio E. Montanari and Marco Doretti

4.27.5 The ultrametric covariance model for modelling teachers' job satisfaction. 1319
Carlo Cavicchia, Maurizio Vichi and Giorgia Zaccaria

4.28 Some recent developments in compositional data analysis 1325

4.28.1 A Robust Approach to Microbiome-Based Classification Problems. 1326
Gianna Serafina Monti and Peter Filzmoser

4.28.2 What is a convex set in compositional data analysis? 1332
Jordi Saperas i Riera, Josep Antoni Martín Fernández

4.28.3 Compositional Analysis on the Functional Distribution of Extended Income. 1338
Elena Dalla Chiara and Federico Perali

4.28.4 Evaluating seasonal-induced changes in river chemistry using Principal Balances. 1344
Caterina Gozzi and Antonella Buccianti

4.28.5 Compositional Data Techniques for the Analysis of the Ragweed Allergy. 1350
Gianna S. Monti, Maira Bonini, Valentina Ceriotti, Matteo Pelagatti and Claudio M. Ortolani

4.29 Spatial data analysis 1356

4.29.1 Spatial multilevel mixed effects modeling for earthquake insurance losses in New Zealand. 1357
F. Marta L. Di Lascio and Selene Perazzini

4.29.2 Weighted distances for spatially dependent functional data. 1363
Andrea Diana, Elvira Romano, Claire Miller and Ruth O'Donnell

4.29.3 Spatial modeling of childcare services in Lombardia. 1369
Emanuele Aliverti, Stefano Campostrini, Federico Caldura and Lucia Zanotto

4.29.4 On the use of a composite attractiveness index for the development of sustainable tourist routes. 1375
Claudia Cappello, Sandra De Iaco, Sabrina Maggio and Monica Palma

4.30 Statistical applications in education 1381

4.30.1 Does self-efficacy influence academic results? A separable-effect mediation analysis. 1382
Chiara Di Maria

4.30.2 Statistics Knowledge assessment: an archetypal analysis approach. 1388
Bruno Adabbo, Rosa Fabbricatore, Alfonso Iodice D'Enza and Francesco Palumbo

4.30.3 Exploring drivers for Italian university students' mobility: first evidence from AlmaLaurea data. 1394
Giovanni Boscaino and Vincenzo Giuseppe Genova

4.30.4 Can Grading Policies influence the competition among Universities of different sizes? 1400
Gabriele Lombardi and Antonio Pio Distaso

4.30.5 The class A journals and the Italian academic research outcomes in Statistical Sciences. 1406
Maria Maddalena Barbieri, Francesca Bassi, Antonio Irpino and Rosanna Verde

4.31 Statistical methods for finance 1412

4.31.1 Hypotheses testing in mixed–frequency volatility models: a bootstrap approach. 1413
Vincenzo Candila and Lea Petrella

4.31.2 Quantile Regression Forest with mixed frequency Data. 1419
Mila Andreani, Vincenzo Candila and Lea Petrella

4.31.3 Higher order moments in Capital Asset Pricing Model betas. 1425
Giuseppe Arbia, Riccardo Bramante and Silvia Facchinetti

4.31.4 When Does Sentiment Matter in Predicting Cryptocurrency Bubbles? 1431
Arianna Agosto and Paolo Pagnottoni

4.32 Statistical methods for high dimensional data 1437

4.32.1 Virtual biopsy in action: a radiomic-based model for CALI prediction. 1438
Francesca Ieva, Giulia Baroni, Lara Cavinato, Chiara Masci, Guido Costa, Francesco Fiz, Arturo Chiti and Luca Viganò

4.32.2 Functional alignment by the "light" approach of the von Mises-Fisher-Procrustes model. 1444
Angela Andreella and Livio Finos

4.32.3 A screening procedure for high-dimensional autologistic models. 1450
Rodolfo Metulini and Francesco Giordano

4.32.4 Covariate adjusted censored gaussian lasso estimator. 1456
Luigi Augugliaro, Gianluca Sottile and Veronica Vinciotti

4.32.5 Ranking-Based Variable Selection for ultra-high dimensional data in GLM framework. 1462
Francesco Giordano, Marcella Niglio and Marialuisa Restaino

4.33 Statistical methods in higher education 1468

4.33.1 Effects of remote teaching on students' motivation and engagement: the case of the University of Modena & Reggio Emilia. 1469
Isabella Morlini and Laura Sartori

4.33.2 A random effects model for the impact of remote teaching on university students' performance. 1475
Silvia Bacci, Bruno Bertaccini, Simone Del Sarto, Leonardo Grilli and Carla Rampichini

4.33.3 Multinomial semiparametric mixed-effects model for profiling engineering university students. 1481
Chiara Masci, Francesca Ieva and Anna Maria Paganoni

4.33.4 Evaluating Italian universities: ANVUR periodic accreditation judgment versus international rankings. 1487
Angela Maria D'Uggento, Nunziata Ribecco and Vito Ricci

4.33.5 Women's career discrimination in the Italian Academia in the last 20. 1493
Daniele Cuntrera, Vincenzo Falco and Massimo Attanasio

4.34 Statistical methods with Bayesian networks 1499

4.34.1 Statistical Micro Matching Using Bayesian Networks. 1500
Pier Luigi Conti, Daniela Marella, Paola Vicard and Vincenzina Vitale

4.34.2 Modeling school managers challenges in the pandemic era with Bayesian networks. 1506
Maria Chiara De Angelis and Flaminia Musella and Paola Vicard

4.34.3 Structural learning of mixed directed acyclic graphs: a copula-based approach. 1512
Federico Castelletti

4.34.4 Inference on Markov chains parameters via Large Deviations ABC. 1518
Cecilia Viscardi, Fabio Corradi, Michele Boreale and Antonietta Mira

4.34.5 A propensity score approach for treatment evaluation based on Bayesian Networks. 1524
Federica Cugnata, Paola M.V. Rancoita, Pier Luigi Conti, Alberto Briganti, Clelia Di Serio, Fulvia Mecatti and Paola Vicard

4.35 Statistical modelling for the analysis of contemporary societies 1530

4.35.1 Social Network Analysis to analyse the relationship between 'victim-author' and 'motivation' of violence against women in Italy. 1531
Alessia Forciniti

4.35.2 Satisfaction and sustainability propensity among elderly bike-sharing users. 1537
Paolo Maranzano, Roberto Ascari, Paola Maddalena Chiodini and Giancarlo Manzi

4.35.3 Media and Investors’ Attention. Estimating analysts’ ratings and sentiment of a financial column to predict abnormal returns. 1543
Riccardo Ferretti and Andrea Sciandra

4.35.4 Predictions of regional HCE: spatial and time patterns in an ageing population framework. 1549
Laura Rizzi, Luca Grassetti, Divya Brundavanam, Alvisa Palese and Alessio Fornasin

4.36 Surveillance methods and statistical models in the Covid-19 crisis 1555

4.36.1 The Italian Social Mood on Economy Index during the Covid-19 Crisis. 1556
Alessandra Righi and Diego Zardetto

4.36.2 Modeling the first wave of the COVID-19 pandemic in the Lombardy region, Italy, by using the daily number of swabs. 1562
Claudia Furlan and Cinzia Mortarino

4.36.3 Analysing the Covid-19 pandemic in Italy with the SIPRO model. 1568
Martina Amongero, Enrico Bibbona and Gianluca Mastrantonio

4.36.4 Intentions of union formation and dissolution during the COVID-19 pandemic. 1574
Bruno Arpino and Daniela Bellani

4.37 Time series methods 1580

4.37.1 Bootstrap-based score test for INAR effect. 1581
Riccardo Ievoli and Lucio Palazzo

4.37.2 Evaluating the performance of a new picking algorithm based on the variance piecewise constant models. 1587
Nicoletta D’Angelo, Giada Adelfio, Antonino D’Alessandro and Marcello Chiodi

4.37.3 Conditional moments based time series cluster analysis. 1593
Raffaele Mattera and Germana Scepi

4.37.4 On the asymptotic mean-squared prediction error for multivariate time series. 1599
Gery Andrés Díaz Rubio, Simone Giannerini, and Greta Goracci

4.37.5 Spherical autoregressive change-point detection with applications. 1605
Federica Spoto, Alessia Caponera and Pierpaolo Brutti

5 Posters 1611

5.1 A method for incorporating historical information in non-inferiority trials. 1612
Fulvio De Santis and Stefania Gubbiotti

5.2 Optimal credible intervals under alternative loss functions. 1618
Fulvio De Santis and Stefania Gubbiotti

5.3 Statistical learning for credit risk modelling. 1624
Veronica Bacino, Alessio Zoccarato, Caterina Liberati and Matteo Borrotti

5.4 Evaluating heterogeneity of agreement with strong prior information. 1630
Federico M. Stefanini

5.5 Analysis of the spatial interdependence of the size of endoreduplicated nuclei observed in confocal microscopy. 1636
Ivan Sciascia, Andrea Crosino, Gennaro Carotenuto and Andrea Genre

5.6 A Density-Peak Approach to Clustering Graph-Structured Data. 1642
Riccardo Giubilei

5.7 The employment situation of people with disabilites in Tuscany, A Survey on the workplace. 1648
Paolo Addis, Alessandra Coli and Gianfranco Francese

5.8 Robustness of statistical methods for modeling paired count data using bivariate discrete distributions with general dependence structures. 1654
Marta Nai Ruscone and Dimitris Karlis

6 Satellite events 1660

6.1 Measuring uncertainty in key official economic statistics 1661

6.1.1 Uncertainty in production and communication of statistics: challenges in the new data ecosystem. 1662
Giorgio Alleva and Piero Demetrio Falorsi

6.1.2 Uncertainty and variance estimation techniques for poverty and inequality measures from complex surveys: a simulation study. 1668
Riccardo De Santis, Lucio Barabesi and Gianni Betti

6.1.3 Pandemics and uncertainty in business cycle analysis. 1674
Jacques Anas, Monica Billio, Leonardo Carati, Gian Luigi Mazzi and Hionia Vlachou

6.2 Covid-19: the urgent call for a unified statistical and demographic challenge 1680

6.2.1 Environmental epidemiology and the Covid-19 pandemics 1681

6.2.1.1 The Covid-19 outbreaks and their environment: The Valencian human behaviour. 1682
Xavier Barber, Elisa Espín, Lucia Guevara, Aurora Mula, Kristina Polotskaya and Alejandro Rabasa

6.2.2 Estimation of Covid 19 prevalence 1686

6.2.2.1 Optimal spatial sampling for estimating the SARS-Cov-2 crucial parameters. 1687
Piero Demetrio Falorsi and Vincenzo Nardelli

6.2.2.2 Survey aimed to estimate the seroprevalence of SARS-CoV-2 infection in Italian population at national and regional level. 1693
Stefano Falorsi, Michele D'Alò, Claudia De Vitiis, Andrea Fasulo, Danila Filipponi, Alessio Guandalini, Francesca Inglese, Orietta Luzi, Enrico Orsini and Roberta Radini

6.2.3 Measuring and modeling inequalities following the Covid-19 crisis 1699

6.2.3.1 COVID-19 impacts on young people's life courses: first results. 1700
Antonietta Bisceglia, Concetta Scolorato and Giancarlo Ragozini

6.2.3.2 Exploring Students'Profile and Performance Before and After Covid-19 Lock-down. 1705
Cristina Davino and Marco Gherghi

6.2.4 Nowcasting the Covid-19 outbreaks methods and applications 1711

6.2.4.1 Modeling subsequent waves of COVID-19 outbreak: A change point growth model. 1712
Luca Greco, Paolo Girardi and Laura Ventura

6.2.4.2 The second wave of SARS-CoV-2 epidemic in Italy through a SIRD model. 1718
Michela Baccini and Giulia Cereda

6.2.5 The impact of Covid-19 on survey methods 1724

6.2.5.1 Collecting cross-national survey data during the COVID-19 pandemic: Challenges and implications of data collection for the 50+ population in the Survey of Health, Ageing and Retirement in Europe (SHARE). 1725
Michael Bergmann, Arne Bethmann, Yuri Pettinicchi and Borsch-Supan

6.2.5.2 Adapting a Long-Term Panel Survey to Pandemic Conditions. 1731
Peter Lynn

6.2.6 Young contributions in Covid-19 statistical modelling 1737

6.2.6.1 Statistical communication of COVID-19 epidemic using widely accessible interactive tools. 1738
Marco Mingione and Pierfrancesco Alaimo Di Loro

6.2.6.2 Modelling COVID-19 evolution in Italy with an augmented SIRD model using open data. 1744
Vincenzo Nardelli, Giuseppe Arbia, Andrea Palladino and Luigi Giuseppe Atzeni

Directional M-quantile regression for multivariate dependent outcomes

Regressione M-quantile direzionale per dati multivariati dipendenti

Merlo Luca, Petrella Lea and Tzavidis Nikos

Abstract In the present work we generalize the univariate M-quantile regression to the analysis of multivariate dependent outcomes. Extending the notion of directional quantiles, we introduce directional M-quantiles which are obtained as projections of the original data on a specified unit norm direction. In order to take into consideration the correlation within grouped measurements and to increase efficiency, we develop a marginal M-Quantile regression model extending the well-known generalized estimating equations approach. We build M-quantile regions and contours which allow us to investigate the effect of the covariates on the location, spread and shape of the distribution of the responses. To identify potential outliers and provide a simple visual representation of the variability of the M-quantile contours estimator, we construct confidence envelope via nonparametric bootstrap. The validity of our method is analyzed through the study of the wages data from the National Longitudinal Survey of Youth.

Abstract In questo lavoro si estende la regressione M-quantilica univariata per l'analisi di dati multivariati dipendenti introducendo la definizione di M-quantile direzionale associato a variabili risposta vettoriali. Al fine di incorporare la struttura di correlazione dei dati nella procedura di stima e determinare stimatori più efficienti, si considera un modello marginale M-quantile estendendo l'approccio delle equazioni di stima generalizzate. Inoltre, proponiamo di utilizzare i contorni M-quantile per investigare l'effetto delle covariate sulla distribuzione delle variabili risposta e, per esaminare la loro variabilità, costruiamo degli insiemi di confidenza attraverso l'approccio bootstrap. L'analisi empirica si concentra sulle retribuzioni salariali di giovani americani ottenute dal National Longitudinal Survey of Youth.

Merlo Luca
Department of Statistical Sciences, Sapienza University of Rome, e-mail: luca.merlo@uniroma1.it

Petrella Lea
MEMOTEF Department, Sapienza University of Rome, e-mail: lea.petrella@uniroma1.it

Tzavidis Nikos
Department of Social Statistics and Demography, Southampton Statistical Sciences Research Institute, University of Southampton, e-mail: N.TZAVIDIS@soton.ac.uk

Key words: Correlated data, GMQEE, Marginal approach, M-quantile contours

1 Introduction

In the univariate setting, the quantile regression approach proposed by [7] has attracted considerable interest in many applications because it provides a way to model the conditional quantiles of a response as a function of explanatory variables in order to have a more complete picture of the entire conditional distribution compared to the classical mean regression. For a detailed review and list of references see [8]. Within the quantile regression framework, a possible alternative is represented by the M-quantile regression approach proposed by [2]. This method provides a "quantile-like" generalization of mean regression based on influence functions combining in a common framework the robustness and efficiency properties of quantiles and expectiles [12], respectively. Although M-quantiles have a less intuitive interpretation than standard quantiles, with respect to the latter, they are very versatile. Specifically, they allow for robust estimation in the presence of influential observations, they can trade robustness for efficiency, ensure uniqueness of the Maximum Likelihood solutions and offer greater stability as a wide range of continuous influence functions can be employed. Unfortunately, M-quantiles have remained relegated to univariate problems due to the lack of a natural ordering in a p-dimensional space, $p > 1$, which preclude the laying down of pertinent concepts of multivariate M-quantiles, ranks and signs. Yet, an extension to higher dimensions could prove to be very useful in many fields of applied statistics when the problem being studied involves the characterization of the distribution of a multivariate response. In the literature some proposals for defining the multivariate M-quantile have been put forward by [2], [9] and [1], for example.

In the present paper we generalize the univariate M-quantile regression to the multivariate setting for the analysis of dependent data by extending the notion of directional quantiles in [10]. More in detail, we introduce directional M-quantiles which are obtained as projections of the original data on a specified unit norm direction. In real world scenarios, observations are often correlated with each other across time, space, or other dimensions, like groups, and their analysis deserves specific instruments which have received enormous attention over the years [3, 4]. In order to take into consideration the correlation within grouped measurements and to increase efficiency, we develop a Marginal M-Quantile (MMQ) regression model. The marginal approach refers to a general class of statistical methods that are used to model dependent data where observations within a cluster are correlated with each other ([11, 5, 3, 4]). A popular estimation procedure for estimating the marginal model parameters is the Generalized Estimating Equations (GEE) approach of [11]. Because the true correlation structure of the data is unknown, the GEE formulates a "working covariance matrix" to capture dependence between observations and incorporate that structure into the model. To estimate the model parameters, we extend the well-known GEE approach of [11] and present the Generalized M-Quantile Estimating Equations (GMQEE). For a fixed direction, we derive the asymptotic properties for the proposed estimator and establish consistency and asymptotic normality. We also investigate M-quantile regions and contours for a given quantile level and we propose to use M-quantile contour lines to investigate the effect of the covariates on the response variables. In order to visualize the sample variability of the M-quantile contours estimator, we construct confidence envelopes via nonparametric bootstrap. From an empirical point of view, we exploit the proposed MMQ regression model to track the labor-market experiences of male high school dropouts collected by the National Longitudinal Survey of Youth (NLSY).

2 Methodology

Let $\mathbf{Y}_{ij} = (Y_{ij}^{(1)}, \dots, Y_{ij}^{(p)})'$ and $\mathbf{X}_{ij} = (X_{ij}^{(1)}, \dots, X_{ij}^{(k)})$ denote a continuous p-variate response variable and a k-dimensional vector of explanatory variables for the i-th statistical unit in the j-th cluster of size n_j, for $j = 1, ..., d$ and $i = 1, ..., n_j$ with $n = \sum_{j=1}^{d} n_j$, respectively. We define $\mathbf{u}$ a unit norm direction vector ranging over $\mathscr{S}^{p-1} = \{\mathbf{z} \in \mathbb{R}^p : ||\mathbf{z}|| = 1\}$. To simplify the notation, we stack up the projected responses on $\mathbf{u}$ to the n_j dimensional vector $\tilde{\mathbf{Y}}_j = (\mathbf{u}'\mathbf{Y}_{1j}, \dots, \mathbf{u}'\mathbf{Y}_{n_j j})'$, while $\mathbf{X}_j = (\mathbf{X}_{ij}, \dots, \mathbf{X}_{n_j j})$ is a $n_j \times k$ matrix collecting the covariates for group j. Extending [10], we define the directional M-quantile for multivariate distributions as follows.

Definition 1. Let $\mathbf{Y}$ be a p-dimensional random vector with absolutely continuous distribution function. For any $\tau \in (0,1)$ and direction $\mathbf{u} \in \mathscr{S}^{p-1}$, let $\psi_\tau(u) = | \tau - \mathbf{1}_{(u<0)} | \psi(u)$ denote the asymmetric Huber influence function with $\psi(u) = u\mathbf{1}_{(|u| \leq c)} + c\,\mathrm{sign}(u)\mathbf{1}_{(|u|>c)}$, where c denotes a tuning constant bounded away from zero. Then, the directional M-quantile of order τ in the direction $\mathbf{u}$, $\theta_{\mathbf{u}}(\tau)$, is the τ-th M-quantile of the corresponding projection of the distribution of $\mathbf{Y}$, namely:

$$\int \psi_\tau(\mathbf{u}'\mathbf{y} - \theta_{\mathbf{u}}(\tau)) \mathrm{d}F_{\mathbf{u}'\mathbf{Y}} = 0. \tag{1}$$

The proposed directional M-quantile is real-valued and it corresponds to the univariate τ-th M-quantile of the distribution of $\mathbf{u}'\mathbf{Y}$ where the direction $\mathbf{u}$ can be interpreted as a weight vector for each marginal distribution of $\mathbf{Y}$ involved in the regression problem. In addition, directional M-quantiles inherit the computational advantages, robustness and efficiency properties of standard univariate M-quantiles. Specifically, by varying the tuning constant c in $\psi_\tau(\cdot)$, directional M-quantiles reduce to directional quantiles of [10] when $c \to 0$ and reduce to directional expectiles for c large. Clearly, Definition 1 includes the traditional univariate one when $p = 1$. In the regression context, the proposed definition can be easily extended to conditional distributions when covariates are available. For a given $\tau \in (0,1)$ and $\mathbf{u} \in \mathscr{S}^{p-1}$, the directional M-quantile model is defined as:

$$\theta_{\mathbf{u},\mathbf{X}}(\tau) = \mathbf{X}_{ij}'\beta(\tau), \qquad i = 1, ..., n_j \text{ and } j = 1, ..., d, \tag{2}$$

where $\beta(\tau)$ is the k-dimensional vector of regression coefficients.

To account for the dependence structure of the data we consider the so called marginal modeling approach and estimate the parameters using the GEE. By introducing a suitable correlation matrix $\mathbf{C}_j(\mathbf{r}_j)$ of size n_j indexed by the s_j-dimensional vector $\mathbf{r}_j$ which characterizes the correlation within groups, $j = 1,...,d$, we are able to capture within group dependence and enhance the efficiency of the regression coefficients estimator [11]. Following [13] and [11], for a given τ and direction $\mathbf{u}$, we define the estimator $\widehat{\beta}_{MMQ}(\tau)$ as the solution of the following Generalized M-quantile Estimating Equations (GMQEE):

$$\mathbf{U}(\beta(\tau)) = \sum_{j=1}^{d} \mathbf{U}_j(\beta(\tau)) = \sum_{j=1}^{d} \mathbf{X}_j'\boldsymbol{\Sigma}_j^{-1}(\mathbf{r}_j)\mathbf{V}_j^{\frac{1}{2}}\psi_\tau(\mathbf{z}_j) = \mathbf{0}, \tag{3}$$

where $\mathbf{z}_j = \mathbf{V}_j^{-\frac{1}{2}}(\tilde{\mathbf{Y}}_j - \mathbf{X}_j\beta(\tau))$ denotes the n_j-dimensional vector of standardized residuals, $\mathbf{V}_j$ is the diagonal matrix of size n_j which contains the scale parameter σ_τ^2 for the residuals' distribution $\tilde{\mathbf{Y}}_j - \mathbf{X}_j\beta(\tau)$ and $\boldsymbol{\Sigma}_j(\mathbf{r}_j) = \phi\mathbf{V}_j^{\frac{1}{2}}\mathbf{C}_j(\mathbf{r}_j)\mathbf{V}_j^{\frac{1}{2}}$ is the "working" covariance matrix with ϕ being a positive nuisance parameter. It is worth noticing that, when $\mathbf{C}_j(\mathbf{r}_j) = \mathbf{I}_{n_j}$, with $\mathbf{I}_{n_j}$ being the identity matrix of size n_j, independence between clustered observations is assumed. Several choices for $\mathbf{C}_j(\mathbf{r}_j)$ have been proposed in the related literature, such as the exchangeable correlation structure, the AR(1) structure, or the totally unspecified structure. Their specification and parameters interpretation generally depend on the application under investigation. For fixed τ and $\mathbf{u}$, under mild regularity conditions the estimator $\widehat{\beta}_{MMQ}(\tau)$ is consistent and asymptotically normally distributed. In addition, an estimate of the model parameters $(\beta_{MMQ}(\tau), \sigma_\tau, \phi, \mathbf{r}_j)$ can be obtained using a Newton-Raphson algorithm to solve the GMQEE in (3).

To provide a full description of the dependence of the responses $\mathbf{Y}$ on the regressors $\mathbf{X}$, we investigate how directional M-quantiles can provide a summary when, theoretically, all directions over $\mathscr{S}^{p-1}$ are investigated simultaneously, for fixed τ. Let $\mathbf{y}$ denote the realization of the random vector $\mathbf{Y}$. For a given $\tau \in (0,1)$ and $\mathbf{u} \in \mathscr{S}^{p-1}$, we first define the τ-th directional M-quantile regression hyperplane $\pi_{\mathbf{u},\mathbf{x}}(\tau) = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{u}'\mathbf{y} = \theta_{\mathbf{u},\mathbf{x}}(\tau)\}$. Each hyperplane $\pi_{\mathbf{u},\mathbf{x}}(\tau)$ characterizes a lower (open) and an upper (closed) M-quantile regression halfspace $H^-_{\mathbf{u},\mathbf{x}}(\tau) = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{u}'\mathbf{y} < \theta_{\mathbf{u},\mathbf{x}}(\tau)\}$ and $H^+_{\mathbf{u},\mathbf{x}}(\tau) = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{u}'\mathbf{y} \geq \theta_{\mathbf{u},\mathbf{x}}(\tau)\}$, respectively. For $\tau \in (0, \frac{1}{2}]$, the τ-th M-quantile region conditional on $\mathbf{X} = \mathbf{x}$, $R_{\mathbf{x}}(\tau) \subset \mathbb{R}^p$, is defined as:

$$R_{\mathbf{x}}(\tau) = \bigcap_{\mathbf{u} \in \mathscr{S}^{p-1}} H^+_{\mathbf{u},\mathbf{x}}(\tau). \tag{4}$$

The region defined in (4) is convex, compact and bounded, and the corresponding conditional M-quantile contour of order τ is defined as the boundary $\partial R_{\mathbf{x}}(\tau)$ of $R_{\mathbf{x}}(\tau)$. Such quantities are of crucial interest as they are able to detect covariate-dependent features of the distribution of the responses given $\mathbf{X}$, while ensuring robustness to outlying data. Specifically, for fixed τ, when $c \to 0$, M-quantile contours reduce to directional quantile envelopes illustrated in [10]; on the other hand, when

$c \to \infty$ they generate expectile contours. Meanwhile, for a given c, as $\tau \to 0$ the M-quantile contour of order τ approaches the convex hull of the sample data providing valuable information about the extent of extremality of points.

3 Application

The proposed methodology has been applied to the NLSY data (NLS79.txt). The NLSY is a longitudinal study that follows the lives of a sample of American youth born between 1980-84. The considered data contains measurements on hourly log-wages (Lnw), years of experience (Exper) in the workforce, unemployment rates in the local geographic region (Uerate) and race (White (baseline), Black) of male high-school dropouts, aged between 14 and 17 years when first measured. The considered sample consists of $d = 500$ men for a total of $n = 3749$ observations. The aim of this analysis is to investigate how the local area unemployment rate and men race affect differently hourly earning and the workers' experience of disadvantages young Americans (low quantiles) and high earners (high quantiles). To handle dependence between repeated measurements and account for stronger dependence between adjacent measurements than for distant ones, we assume an AR(1) structure, $[\mathbf{C}_j(\mathbf{r}_j)]_{ik} = r^{|i-k|}$, working correlation structure and, the tuning constant c in Definition 1 has been set to 1.345 which gives reasonably efficiency under normality and protects against outliers (see [6]).

Table 1 shows point estimates of the regression coefficients and of the correlation parameter for the MMQ model at $\tau = (0.1, 0.25, 0.5, 0.75, 0.9)$ and for two directions, $\mathbf{u}_1 = (1,0)$ and $\mathbf{u}_2 = (0,1)$, which reduces the multidimensional problem to two MMQ regressions on each component of the bivariate response. In addition, the results of a Marginal Mean (MM) model fitted with the standard GEE approach are reported. We observe that the MM and MMQ models produce comparable estimates at the center of the distribution (the MM model cannot be applied to estimate the covariates' effects in the tails of the distribution). The results show that there is evidence of a negative association between the considered covariates and either log wage and working experience. In particular, the effect is statistically significant at the investigated quantile levels and it is more pronounced at the high quantiles. By looking at the estimated correlation parameters, as expected, there is a high within-subject correlation which is consistent with the repeated measures design.

To provide a graphical representation of the effects of the included covariates in the tails of the responses distribution, we fit the MMQ model at $\tau = (0.01, 0.25)$ for 200 equispaced directions in $\mathscr{S}^{p-1}$ and construct M-quantile regression contours using (4). Figure 1 illustrates the estimated $\partial R_{\mathbf{x}}(\tau)$ conditional on race, white (red curves) and black (blue curves), at the 0.50-th (left), 0.75-th (center) and 0.99-th (right) empirical quantiles of the unemployment rate which correspond to an unemployment rate of 6.9%, 12.2% and 22.9%. The shaded areas represent 95% confidence envelopes for M-quantile contours obtained using nonparametric block bootstrap. The contours for smaller τ capture the effects for more extreme workers e.g., men with low levels of income and experience and those with exceptionally high values of income and experience. The elongated curves indicate that there is

u	Variable	MM	MMQ				
			0.1	0.25	0.5	0.75	0.9
$\mathbf{u}_1$	Intercept	5.204^{***}	2.035^{***}	2.759^{***}	4.352^{***}	6.382^{***}	7.808^{***}
	Black	-0.681^{***}	-0.270^{***}	-0.358^{***}	-0.569^{***}	-0.826^{***}	-0.876^{***}
	Uerate	-0.156^{***}	-0.070^{***}	-0.078^{***}	-0.105^{***}	-0.117^{***}	-0.103^{***}
	r	0.817^{***}	0.637^{***}	0.778^{***}	0.853^{***}	0.824^{***}	0.741^{***}
$\mathbf{u}_2$	Intercept	2.031^{***}	1.702^{***}	1.829^{***}	2.017^{***}	2.245^{***}	2.473^{***}
	Black	-0.089^{***}	-0.071^{***}	-0.077^{***}	-0.094^{***}	-0.094^{***}	-0.074
	Uerate	-0.017^{***}	-0.017^{***}	-0.016^{***}	-0.018^{***}	-0.022^{***}	-0.024^{***}
	r	0.656^{***}	0.480^{***}	0.528^{***}	0.620^{***}	0.539^{***}	0.433^{***}

Table 1 MM and MMQ model parameters estimates at the investigated quantile levels. ***, ** and * denote statistical significance at the $0.01, 0.05$ and 0.1 levels, respectively.

more variability in years of working experience and it can also be easily seen that blue curves always lie below and to the left of the red ones, suggesting the existence of a significant racial wage gap that disadvantages young African-American males, especially at $\tau = 0.25$. Finally, as the unemployment rate increases the M-contours rapidly descend downward from right to left and become cone-shaped which exert downward pressure on both wages and years of working experience.

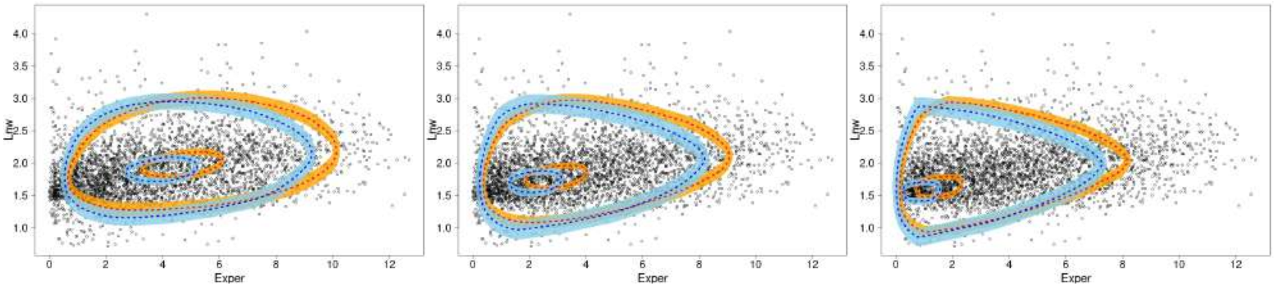

Fig. 1 Estimated M-quantile contours at $\tau = (0.01, 0.25)$ conditional on race, white (red curves) and black (blue curves), at the 0.50-th (left), 0.75-th (center) and 0.99-th (right) empirical quantiles of the unemployment rate. The shaded surfaces represent 95% confidence envelopes for M-quantile contours obtained using nonparametric block bootstrap.

References

[1] Alfò, M., Marino, M. F., Ranalli, M. G., Salvati, N. and Tzavidis, N. [2021], 'M-quantile regression for multivariate longitudinal data with an application to the Millennium Cohort Study', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **70**(1), 122–146.
[2] Breckling, J. and Chambers, R. [1988], 'M-quantiles', *Biometrika* **75**(4), 761–771.
[3] Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S. et al. [2002], *Analysis of longitudinal data*, Oxford University Press.
[4] Goldstein, H. [2011], *Multilevel statistical models*, Vol. 922, John Wiley & Sons.
[5] Heagerty, P. J., Zeger, S. L. et al. [2000], 'Marginalized multilevel models and likelihood inference', *Statistical Science* **15**(1), 1–26.
[6] Huber, P. and Ronchetti, E. [2009], *Robust Statistics*, Wiley.
[7] Koenker, R. and Bassett Jr, G. [1978], 'Regression quantiles', *Econometrica: Journal of the Econometric Society* pp. 33–50.
[8] Koenker, R., Chernozhukov, V., He, X. and Peng, L. [2017], *Handbook of Quantile Regression*, CRC press.
[9] Kokic, P., Breckling, J. and Lübke, O. [2002], A new definition of multivariate M-quantiles, *in* 'Statistical data analysis based on the L_1-norm and related methods', Springer, pp. 15–24.
[10] Kong, L. and Mizera, I. [2012], 'Quantile tomography: using quantiles with multivariate data', *Statistica Sinica* pp. 1589–1610.
[11] Liang, K.-Y. and Zeger, S. L. [1986], 'Longitudinal data analysis using generalized linear models', *Biometrika* **73**(1), 13–22.
[12] Newey, W. K. and Powell, J. L. [1987], 'Asymmetric least squares estimation and testing', *Econometrica* pp. 819–847.
[13] Sinha, S. K. and Rao, J. [2009], 'Robust small area estimation', *Canadian Journal of Statistics* **37**(3), 381–399.