



Book of the Short Papers

Editors: Francesco Maria Chelli, Mariateresa Ciommi, Salvatore Ingrassia, Francesca Mariani, Maria Cristina Recchioni



UNIVERSITÀ
POLITECNICA
DELLE MARCHE



LIUC | BUSINESS
ANALYTICS AND
DATA SCIENCE HUB



CHAIRS

Salvatore Ingrassia (Chair of the Program Committee) - *Università degli Studi di Catania*

Maria Cristina Recchioni (Chair of the Local Organizing Committee) - *Università Politecnica delle Marche*

PROGRAM COMMITTEE

Salvatore Ingrassia (Chair), Elena Ambrosetti, Antonio Balzanella, Matilde Bini, Annalisa Busetta, Fabio Centofanti, Francesco M. Chelli, Simone Di Zio, Sabrina Giordano, Rosaria Ignaccolo, Filomena Maggino, Stefania Mignani, Lucia Paci, Monica Palma, Emilia Rocco.

LOCAL ORGANIZING COMMITTEE

Maria Cristina Recchioni (Chair), Chiara Capogrossi, Mariateresa Ciommi, Barbara Ermini, Chiara Gigliarano, Riccardo Lucchetti, Francesca Mariani, Gloria Polinesi, Giuseppe Ricciardo Lamonica, Barbara Zagaglia.

ORGANIZERS OF INVITED SESSIONS

Pierfrancesco Alaimo Di Loro, Laura Anderlucci, Luigi Augugliaro, Iliaria Benedetti, Rossella Berni, Mario Bolzan, Silvia Cagnone, Michela Cameletti, Federico Camerlenghi, Gabriella Campolo, Christian Capezza, Carlo Cavicchia, Mariateresa Ciommi, Guido Consonni, Giuseppe Ricciardo Lamonica, Regina Liu, Daniela Marella, Francesca Mariani, Matteo Mazziotta, Stefano Mazzuco, Raya Muttarak, Livia Elisa Ortensi, Edoardo Otranto, Iliaria Prosdocimi, Pasquale Sarnacchiaro, Manuela Stranges, Claudia Tarantola, Isabella Sulis, Roberta Varriale, Rosanna Verde.

FURTHER PEOPLE OF LOCAL ORGANIZING COMMITTEE

Elisa D'Adamo, Christian Ferretti, Giada Gabbianelli, Elvina Merkaj, Luca Pedini, Alessandro Pionati, Marco Tedeschi, Francesco Valentini, Rostand Arland Yebetchou Tchounkeu

Technical support: Matteo Mercuri, Maila Ragni, Daniele Ripanti

Copyright © 2023

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891935618AAVV

Contents

| | |
|--|-------------|
| Preface | XXII |
| 1 Plenary Sessions | 1 |
| Inequality indices: accurate simulation-based inference Maria-Pia Victoria-Feser | 2 |
| Examples from the Interface of Neural Models and Spatio-Temporal Statistics in Environmental Applications Christopher K. Wikle, Likun Zhang, Myungsoo Yoo and Xiaoyu Ma | 7 |
| Demographic change and sustainability: novel approaches from digital and computational demography Emilio Zagheni | n.a. |
| 2 Invited Sessions | 14 |
| Machine learning in the design, analysis and integration of sample surveys | |
| Causal Discovery for complex survey data Paola Vicard | 15 |
| Data Integration without conditional independence: a Bayesian Networks approach Pier Luigi Conti, Paola Vicard and Vincenzina Vitale | 21 |
| Mass imputation through Machine Learning techniques in presence of multi-source data Fabrizio De Fausti, Marco Di Zio, Romina Filippini and Simona Toti | 27 |
| Machine learning: different uses and perspectives | |
| Evaluation of pollution containment policies in the US and the role of machine learning algorithms Marco Di Cataldo, Margherita Gerolimetto, Stefano Magrini and Alessandro Spiganti | 32 |

| | |
|--|------|
| Machine Learning for Official Statistics: An Application on External Trade | n.a. |
| Mauro Bruno, Maria Serena Causo, Alessio Guandalini, Francesco Ortame and Silvia Russo | |
| Machine learning, data quality and official statistics: challenges and opportunities | n.a. |
| Stefano Menghinello | |

Statistical Machine Learning for environmental applications

| | |
|---|----|
| Gaussian Processes and Deep Neural Networks for Spatial Prediction | 38 |
| Alex Cucco, Luigi Ippoliti, Nicola Pronello, Pasquale Valentini and Carlo Zaccardi | |
| How can we explain Random Forests in a spatial framework? | 42 |
| Natalia Golini, Luca Patelli and Xavier Barber | |
| Recent approaches in coupling deep learning methods with the statistical analysis of spatial point patterns | 48 |
| Jorge Mateu and Abdollah Jalilian | |

Statistical Process Monitoring for Complex Data in Industry 4.0

| | |
|---|----|
| A Kernel-based Nonparametric Multivariate CUSUM for Location Shifts | 53 |
| Konstantinos Bourazas, Konstantinos Fokianos, Christos Panayiotou and Marios Polycarpou | |
| An Approach for Profile Monitoring via Mixture Regression Models | 58 |
| Davide Forcina, Antonio Lepore and Biagio Palumbo | |
| Anomaly Detection in Circular Data | 63 |
| Houyem Demni and Giovanni C. Porzio | |

Advances in Data Science and Statistical Learning [IMS Invited Session]

| | |
|--|------|
| Empirical Bayes approximation of Bayesian learning: understanding a common practice | n.a. |
| Sonia Petrone | |
| Generalized Fiducial Inference on Differentiable Manifolds - a geometric perspective | n.a. |
| Jan Hannig | |
| Model-free bootstrap and conformal prediction in regression | n.a. |
| Dimitris Politis | |

ENBIS Session: System Maintenance, Boosting algorithms for regression, and Research Excellence

| | |
|---|----|
| Boosting Diversity in Regression Ensembles | 69 |
| Mathias Bourel, Jairo Cugliari, Yannig Goude and Jean-Michel Poggi | |
| How ENBIS has contributed to the UK Universities Research Excellence Framework | 71 |
| Shirley Coleman | |
| Maintenance of degrading systems by dynamic programming or reinforcement learning | 75 |
| Antonio Pievatolo | |

Population Dynamics, Climate Change and Sustainability

- Climate change impacts on fertility in low- and middle-income countries: An analysis based on global sub-national data n.a.
Côme Cheritel, Roman Hoffmann and Raya Muttarak
- Environmental Exposures and Under-5 Mortality in India: A Survival Analysis of DHS data 79
Vinod Joseph Kannankeril Joseph
- The impact of temperature on expressed sentiment by migration status: Evidence from geo-located Twitter data 84
Risto Conte Keivabu and Jisu Kim

Statistical Learning for health research and omics data

- An alternative to the Dirichlet-multinomial regression model for microbiome data analysis 95
Roberto Ascari, Sonia Migliorati and Andrea Ongaro
- Modelling ordinal response to treatment in a real-world cohort study 101
Marco Alfò, Maria Francesca Marino and Silvia D'Elia
- On the application of the symmetric graphical lasso for paired data 105
Saverio Ranciati and Alberto Roverato

The Economic behaviour of Sustainability

- Airports performances and sustainable practices. An empirical study on Italian data 110
Riccardo Gianluigi Serio, Maria Michela Dickson, Diego Giuliani and Giuseppe Espa
- Sustainability: still an undefined concept for Italians 116
Raffaele Angelone and Andrea Marletta
- Quasi-experimental evidence on COVID-19 lockdown effects on Italian household food shopping basket composition and its sustainability 122
Beatrice Biondi and Mario Mazzocchi

Advances in statistical methods for complex problems

- Inferring multiple treatment effects from observational studies using confounder importance learning n.a.
Omiros Papaspiliopoulos
- Path analysis in Ising models: an application to cyber-security risk assessment 127
Monia Lupparelli and Giovanni M. Marchetti
- Causal Regularization n.a.
Lucas Kania and Ernst Wit

Explainable machine learning models

- Enhancing Markowitz model: inspection of correlations and tail covariances 133
Gloria Polinesi

| | |
|---|------|
| Objective and subjective dimension of economic well-being: an approach based on statistical matching | 139 |
| Daniela Marella, Vincenzina Vitale and Pierpaolo D'Urso | |
| Sustainable, Accurate, Fair and Explainable Machine Learning Models | n.a. |
| Paolo Giudici and Emanuela Raffinetti | |
| Flexible Learning for Environmental Sustainability | |
| Comparison of traffic flow data sources for air pollution modelling | 145 |
| Theresa Smith and Nick McCullen | |
| Data analysis of photogrammetry-based mapping: the sea cucumbers in the Giglio Island as a case-study | 150 |
| Gianluca Mastrantonio, Daniele Ventura, Edoardo Casoli, Arnold Rakaj, Giovanna Jona Lasinio and Alessio Pollice | |
| Understanding forest damage in Germany: Finding key drivers to help with future forest conversion of climate sensitive | 156 |
| Nicole Augustin, Heike Puhlmann and Simon Trust | |
| Inequalities in higher education outcomes: learning from data | |
| Inequalities in international students mobility | 163 |
| Kristijan Breznik, Giancarlo Ragozini and Marialuisa Restaino | |
| Uncovering the interplay of territorial, socioeconomic, and demographic factors in high school to university transition | 169 |
| Vincenzo Giuseppe Genova, Andrea Priulla and Martina Vittorietti | |
| Statistical Learning of demographic and health dynamics | |
| Estimating the impact of a vaccine mandate: the case of measles in Italy | n.a. |
| Chiara Chiavenna | |
| Leveraging deep neural networks to estimate age-specific mortality from life expectancy at birth | n.a. |
| Andrea Nigri | |
| Nowcasting Daily Population Displacement in Ukraine through Social Media Advertising Data | n.a. |
| Claire Dooley, Ridhi Kashyap, Douglas Leasure and Francesco Rampazzo | |
| Challenges towards Fairness and Transparency for Data Processes, Algorithms and Decision-Support Models | |
| Challenges on Ethics, and Privacy in AI Applications to Fintech | 175 |
| Catarina Silva, Joana Matos Dias and Bernardete Ribeiro | |
| Uncertainty and fairness metrics | 180 |
| Anna Gottard | |

Educational Data mining: methods for complex data in students' assessment

Analysis of University Grades: An IRT Model for Responses and Response Times with Censoring 186
Michela Battauz

Predicting high schools' students performances with registry's data: a machine learning approach 191
Lidia Rossi, Marta Cannistrà and Tommaso Agasisti

Using response times to identify cheaters in CAT: A simulation study 195
Luca Bungaro, Bernard P. Veldkamp and Mariagiulia Matteucci

Spatial and Spatio-Temporal Modeling: Theory and Applications

A geostatistical investigation of the ammonia-livestock relationship in the Po Valley, Italy 200
Paolo Maranzano, Kelly McConville, Philipp Otto and Felicetta Carillo

Bayesian multi-species N-mixture models for large scale spatial data in community ecology 206
Michele Peruzzi

Minimum contrast for point processes' first-order intensity estimation 212
Nicoletta D'Angelo and Giada Adelfio

Statistical Framework for Measuring the Sustainability of Tourism

Data validity and statistical conformity with Benford's Law: the case of tourism in Sicily 217
Roy Cerqueti and Davide Provenzano

Exploring the level of digitalization of the Italian museums through a multilevel ordered logit model 228
Claudia Cappello, Sabrina Maggio and Sandra De Iaco

Functional Partial Least-Squares via Regression Splines. An application on Italian Sustainable Development Goals data 232
Ida Camminatiello, Rosaria Lombardo, Jean-Francois Durand and Leonardo S. Alaimo

Statistical learning for well-being analysis

Assessing multidimensional poverty of the Italian provinces during Covid-19: a small area estimation approach 238
Mariateresa Ciommi, Chiara Gigliarano, Francesca Mariani and Gloria Polinesi

The fuzzy set approach as statistical learning for the analysis of multidimensional well-being 244
Gianni Betti, Federico Crescenzi, Antonella D'Agostino and Laura Neri

What Makes a Satisfying Life? Prediction and Interpretation with Machine-Learning Algorithms n.a.
Conchita D'Ambrosio

Bayesian contributions to Statistical Learning

A Bayesian framework for early cancer screening 249
Sally Paganin and Jeff Miller

Imputing Synthetic Pseudo Data from Aggregate Data: Development and Validation for Precision Medicine n.a.
Cecilia Balocchi

Linear models with assumptions-free residuals: a Bayesian Nonparametric approach 254
Filippo Ascolani and Valentina Ghidini

Data Visualization for Smart Insights and Advanced Predictive Analytics

Applications of data visualization for industry 259
Martina Dossi, Stefano Sangaletti, Marilena Di Bari and Federica Bruschini

Some Notes on the Use of the Circular Boxplot n.a.
Giovanni Camillo Porzio and Davide Buttarazzi

TERRA: a smart visualization tool for international trade in goods statistics 265
Francesco Amato, Mauro Bruno and Maria Serena Causo

Methods for the analysis of distributional data

Clustering of Distributional Data based on LDQ transformation 271
Gianmarco Borrata and Rosanna Verde

Dynamic learning from data streams through the combined use of probability density functions and simplicial functional principal component analysis 276
Francesca Fortuna, Fabrizio Maturo and Tonio Di Battista

Multivariate Parametric Analysis of Distributional Data n.a.
Paula Brito

Migrants and Refugees in Europe: social, economic and health-related issues

Labor Market Return to Refugees' Human Capital Investment: A Natural Experiment in Sweden n.a.
Eleonora Mussino

Social networks and loneliness among older migrants in Italy 282
Viviana Amati, Eralba Cela and Elisa Barbiano di Belgiojoso

The Italian Decree on Security: An Analysis of the Impact on Asylum Applications 287
Giorgio Piccitto

Modelling and Forecasting High-dimensional time series

Adaptive combinations of tail-risk forecasts 293
Alessandra Amendola, Vincenzo Candila, Antonio Naimoli and Giuseppe Storti

Are Monetary Policy Announcements related to Volatility Jumps? 299
Giampiero Gallo, Demetrio Lacava and Edoardo Otranto

| | |
|---|------------|
| Regularized Estimation and Prediction of the El Nino/Southern Oscillation Cycle | n.a. |
| Alessandro Giovannelli and Tommaso Proietti | |
| 3 Contributed Sessions | 305 |
| Bayesian nonparametric methods | |
| Bayesian density estimation for modeling age-at-death distribution | 306 |
| Davide Agnoletto, Tommaso Rigon and Bruno Scarpa | |
| Bayesian mixing distribution estimation in the Gaussian-smoothed 1-Wasserstein distance | 311 |
| Catia Scricciolo | |
| Bayesian nonparametric estimation of heterogeneous intrinsic dimension via product partition models | 316 |
| Francesco Denti, Antonio Di Noia and Antonietta Mira | |
| Bayesian nonparametric multiple change point detection for time series of compositional data | 322 |
| Edoardo Marchionni and Riccardo Corradin | |
| Galton-Watson process: a non parametric prior for the offspring distribution | 328 |
| Massimo Cannas, Michele Guindani and Nicola Piras | |
| Hierarchical processes in survival analysis | 333 |
| Riccardo Cogo, Federico Camerlenghi and Tommaso Rigon | |
| Economics and Statistics | |
| A regression analysis for count data to investigate the effectiveness of incentives on the adoption of 4.0 technologies | 339 |
| Stefano Bonnini and Michela Borghesi | |
| Statistical analysis on SDGs indicators related to environmental sustainability | 344 |
| Najada Firza, Anisa Bakiu and Dante Mazzitelli | |
| Empowering futures adopting a spatial convergence of opinions: a Real-Time Spatial Delphi approach | 349 |
| Yuri Calleo, Simone Di Zio and Francesco Pilla | |
| Stocks price forecasts using Stochastic Differential Equations: an empirical assessment | 355 |
| Dario Frisardi and Matteo Spuri | |
| The Added-Worker Effect within Italian Households | 361 |
| Donata Favaro and Anna Giraldo | |
| Health statistics 1 | |
| A model for the natural history of breast cancer: application to a Norwegian screening dataset | 365 |
| Laura Bondi, Marco Bonetti and Solveig Hofvind | |

| | |
|---|-----|
| Generalized Bayesian Ensemble Survival Trees: an extension to categorical variables to apply it to real data Elena Ballante | 370 |
| Joint modelling of hospitalizations and survival in Heart Failure patients: a discrete non parametric frailty approach Chiara Masci, Marta Spreafico and Francesca Ieva | 375 |
| Mobility trends in Italy during the first wave of Covid-19 pandemic: analysis on Google data Ilaria Bombelli and Daniele De Rocchi | 381 |
| Tracking attitudes towards COVID vaccines: A text mining analysis Leonardo Scarso, Marco Novelli and Francesco Saverio Violante | 387 |
| Treatment effect assessment in observational studies with multi-level treatment and outcome Federica Cugnata, Paola Vicard, Paola M.V. Rancoita, Fulvia Mecatti, Clelia Di Serio and Pier Luigi Conti | 393 |
| Indicators: composition, uses and limitations | |
| Are European consumers willing to pay the true price for sustainable food? Luca Secondi and Mengting Yu | 399 |
| Can the reliability of composite indexes be impacted by uncertainty of individual indicators? Caterina Giusti, Stefano Marchetti and Vincenzo Mauro | 406 |
| Initial Coin Offerings and ESG: allies or enemies? Alessandro Bitetto and Paola Cerchiello | 411 |
| On the impact of intraclass correlation in the ANVUR evaluation of academic departments Giorgio Edoardo Montanari and Marco Doretti | 417 |
| Small area estimation of monetary poverty indicators with poverty lines adjusted using local price indexes Luigi Biggeri, Stefano Marchetti, Caterina Giusti, Monica Pratesi, Francesco Schirripa Spagnolo and Gaia Bertarelli | 422 |
| Smart Composite Indicators Measuring Corporate Sustainability: A Sensitivity Analysis Camilla Salvatore, Annamaria Bianchi and Silvia Biffignandi | 428 |
| Multivariate data analysis 1 | |
| A note on most powerful tests for right censored survival data Maria Veronica Vinattieri and Marco Bonetti | 434 |
| Enhancing Principal Components by a Linear Predictor: an Application to Well-Being Italian Data Laura Marcis, Maria Chiara Pagliarella and Renato Salvatore | 439 |

| | |
|--|-----|
| Proper Bayesian Bootstrap for Bagging tree model in survival analysis with correlated data | 445 |
| Farah Naz and Elena Ballante | |
| ROBOUT: a multi-step methodology for conditional outlier detection | 450 |
| Matteo Farnè and Angelos Vouldis | |
| Robustness of the Efficient Covariate-Adaptive Design for balancing covariates in comparative experiments | 456 |
| Rosamarie Frieri, Alessandro Baldi Antognini, Maroussa Zagoraiou, and Marco Novelli | |
| Separation scores: a new statistical tool for scoring and ranking partially ordered data | 462 |
| Marco Fattore | |
| Statistics in Society 1 | |
| Community detection analysis with robin on hashtag network | 468 |
| Valeria Policastro, Francesco Santelli and Giancarlo Ragozini | |
| Film Tourism Motivation through the lens of Trip Advisor data | 474 |
| Nicolò Biasetton, Marta Disegna, Girish Prayag and Elena Barzizza | |
| Life satisfaction and social activities in later life in Italy: a focus on the Internet use | 480 |
| Claudia Furlan and Silvia Meggiolaro | |
| Social capital endowment's role in the intergenerational transmission of education | 485 |
| Alessandra Trimarchi, Maria Gabriella Campolo and Antonino Di Pino Incognito | |
| Streaming Data from Social Networks to Track Political Trends | 490 |
| Emiliano del Gobbo and Barbara Cafarelli | |
| The scientific production on gender dysphoria: a bibliometric analysis | 495 |
| Maria Gabriella Grassia, Marina Marino, Massimo Aria, Rocco Mazza, Luca D'Aniello and Agostino Stavolo | |
| Assessment and Education | |
| A hierarchical modelling approach to explain differential functioning of mathematics items by student's gender | 500 |
| Clelia Cascella | |
| A latent variable approach to Millennials' knowledge of green finance | 506 |
| Maria Iannario, Alessandra Tanda and Claudia Tarantola | |
| Archetypal analysis and latent Markov models: A step-wise approach | 512 |
| Lucio Palazzo, Rosa Fabbriatore and Francesco Palumbo | |
| From high school to university: academic intentions and enrolment of foreign students in Italy | 518 |
| Francesca Di Patrizio, Eleonora Trappolini and Cristina Giudici | |
| Growth models for the progress test in Italian dentistry degree program | 523 |
| Giulio Biscardi, Leonardo Grilli, Carla Rampichini, Laura Antonucci and Corrado Crocetta | |

| | |
|---|-----|
| The COVID-19 pandemic and academic E-learning: Italian students and instructors' perceptions | 527 |
| Francesco Santelli, Teresa Gentile, Davide Bizjak and Lorenzo Fattori | |
| Working Students and job market outcomes: Insights from the University of Florence | 532 |
| Gabriele Lombardi, Valentina Tocchioni and Alessandra Petrucci | |
| Bayesian methods and applications 1 | |
| Analyzing RNA data with scVelo: identifiability issues and a Bayesian implementation | 538 |
| Elena Sabbioni, Enrico Bibbona, Gianluca Mastrantonio and Guido Sanguinetti | |
| Approximate Bayesian Computation for Probabilistic Damage Identification | 544 |
| Cecilia Viscardi, Silvia Monchetti, Luisa Collodi, Gianni Bartoli, Michele Betti, Michele Boreale and Fabio Corradi | |
| Estimation of scientific productivity with a hierarchical Bayesian model | 550 |
| Maura Mezzetti and Ilia Negri | |
| Heat waves and free-knots splines | 555 |
| Gioia Di Credico and Francesco Pauli | |
| The Hierarchical Beta-Bernoulli Process as Out-of-Scope Query Detector | 560 |
| Marco Dalla Pria and Silvia Montagna | |
| Health and mortality | |
| A novel definition of comorbidity based on the Global Burden of Diseases project weights | 566 |
| Angela Andreella, Lorenzo Monasta and Stefano Campostrini | |
| An Age-Period-Cohort model of gender gap in youth mortality | 572 |
| Giacomo Lanfiuti Baldi and Andrea Nigri | |
| Kinlessness in adult and old age across Europe | 578 |
| Marta Pittavino, Bruno Arpino and Elena Pirani | |
| Parameter orthogonalization for Siler mortality model | 584 |
| Claudia Di Caterina and Lucia Zanotto | |
| Pseudo-observations in survival analysis | 590 |
| Marta Cipriani, Alfonso Piciocchi, Valentina Arena and Marco Alfò | |
| Sex Gap in Cancer-Free Life Expectancy: The Association with Smoking, Obesity and Physical Inactivity | 595 |
| Alessandro Feraldi, Cristina Giudici and Nicolas Brouard | |
| Women's Exposure to HIV in Africa: the Role of Intimate Partner Violence | 599 |
| Micaela Arcaio and Anna Maria Parroco | |

Mixture Models

- An extension of finite mixtures of latent trait analyzers for biclustering bipartite networks 605
Dalila Failli, Maria Francesca Marino and Francesca Martella
- Constrained Mixtures of Generalized Normal Distributions 611
Pierdomenico Dutillo, Alfred Kume and Stefano Antonio Gattone
- Mixture-based clustering with covariates for ordinal responses 617
Kemawadee Preedalikit, Daniel Fernández, Ivy Liuc, Louise McMillan, Marta Nai Ruscone and Roy Costilla
- Partial membership models for soft clustering of multivariate count data 623
Emiliano Seri, Thomas Brendan Murphy and Roberto Rocci
- Regression for mixture models for extremes 629
Viviana Carcaiso, Ilaria Prodocimi and Isadora Antoniano-Villalobos
- Robust matrix-variate mixtures of regressions 635
Salvatore Daniele Tomarchio and Michael P. B. Gallagher

Sampling methods and analysis of survey data

- On the use of auxiliary information to define the sampling design for large-scale geospatial data 641
Chiara Bocci and Emilia Rocco
- Optimal joint inclusion probabilities for spatial sampling n.a.
Giuseppe Arbia, Piero Demetrio Falorsi and Vincenzo Nardelli
- Robustness and Balance of Sampling or Experimental Designs and Mixture of Designs 647
Yves Tillé and Ejub Talovic
- Robustness Bounds for Sampling and Experimental Designs 654
Ejub Talovic and Yves Tillé
- Statistical Matching: Hotdeck or Propensity Score? 661
Elena Dalla Chiara, Marcello D'Orazio and Federico Perali
- The Italian experience on register-based statistics considering measurement, coverage and sampling errors 667
Marco Di Zio, Romina Filippini and Simona Toti

Space-time statistics

- A Hierarchical Spatio-Temporal Model for Time-Frequency Data: An application in bioacoustic analysis 673
Hiu Ching Yip, Gianluca Mastrantonio, Enrico Bibbona, Daria Valente and Marco Gamba
- An approach to cluster time series extremes with spatial constraints 679
Alessia Benevento, Fabrizio Durante and Roberta Pappadà
- An integrated space-time model to evaluate the innovation drivers in Italy 685
Emma Bruno, Rosalia Castellano and Gennaro Punzo

| | |
|--|-----|
| Revealing the dynamic relations between traffic and crowding using big data from mobile phone network | 691 |
| Selene Perazzini, Rodolfo Metulini and Maurizio Carpita | |
| SMaC: Spatial Matrix Completion method | 697 |
| Giulio Grossi, Alessandra Mattei and Georgia Papadogeorgou | |
| The impact of traffic flow and road signs on road accidents: an approach based on spatiotemporal point pattern analysis on linear networks | 702 |
| Andrea Gilardi and Riccardo Borgoni | |
| Clustering and classification 1 | |
| A clustering model for flow data: an application to international student mobility | 708 |
| Cinzia Di Nuzzo and Donatella Vicari | |
| Contingency tables with structural zeros and discrete copulas | 713 |
| Roberto Fontana, Elisa Perrone and Fabio Rapallo | |
| Levels Merging in the Latent Class Model | 719 |
| Christophe Biernacki | |
| Model-based clustering of count processes with multiple change | 725 |
| Shuchismita Sarkar and Xuwen Zhu | |
| Similarity Measures and Internal Evaluation Criteria in Hierarchical Clustering of Categorical Data | 729 |
| Jana Cibulková, Zdeněk Šulc, Hana Řezanková and Jaroslav Horníček | |
| Spectral clustering of mixed data via association-based distance | 735 |
| Alfonso Iodice D'Enza, Francesco Palumbo and Cristina Tortora | |
| Dynamic models and time series | |
| A graph based convolution Neural Network approach for forecast reconciliation | 741 |
| Andrea Marcocchia and Pierpaolo Brutti | |
| A multivariate hidden semi-Markov model for the analysis of multiple air pollutants | 747 |
| Marco Mingione, Pierfrancesco Alaimo Di Loro, Francesco Lagona and Antonello Maruotti | |
| A smooth transition autoregressive model for matrix-variate time series | 753 |
| Andrea Bucci | |
| Dynamic network models with time-varying nodes | 759 |
| Luca Gherardini, Mauro Bernardi and Monia Lupparelli | |
| Time lapse analysis of nuclear calcium spiking in plant cells during symbiotic signaling | 765 |
| Ivan Sciascia, Andrea Crosino and Andrea Genre | |
| Two-stage weighted least squares estimator of multivariate conditional mean observation-driven time series models | 770 |
| Mirko Armillotta | |

Environmental learning and indicators

- Assessing the performance of nuclear norm-based matrix completion methods on CO₂ emissions data 776
Rodolfo Metulini, Francesco Biancalani, Giorgio Gnecco and Massimo Riccaboni
- Deep Learning for smart and sustainable agriculture 782
Amalia Vanacore, Armando Ciardiello, Annalisa Izzo, Pierdomenico Zaffino, Carolina Vecchio, Gennaro Pio Auricchio and Luigi Uccelli
- Do green transition, environmental taxes and renew-able energy promote ecological sustainability in G7 countries? Evidence from panel quantile regression 788
Aamir Javed, Agnese Rapposelli and Asif Javed
- Doubly Robust DID for National Parks evaluation: “just” environmental benefits, or socioeconomics impacts as well? 795
Riccardo D’Alberto, Francesco Pagliacci and Matteo Zavalloni
- On the gap between emitted and absorbed carbon dioxide. Are trees enough to save us? 801
Lorenzo Mori and Maria Rosaria Ferrante
- Small scale analysis of energy vulnerability in the municipality of Palermo 806
Giuliana La Mantia

Health statistics 2

- A test for non-differential misclassification error in database epidemiological studies 812
Giorgio Limoncella, Leonardo Grilli, Emanuela Dreassi, Carla Rampichini, Robert Platt and Rosa Gini
- Is the COVID-19 ‘color code’ of Italian regions subjected to political manipulation? 816
Giovanni Busetta and Fabio Fiorillo
- Modelling multilevel ordinal response under endogeneity: application to DTC patients’ outcome 822
Silvia D’Elia
- Monitoring drugs-based diagnostic therapeutic paths in heart failure patients using state-sequence analysis techniques 827
Nicole Fontana, Laura Savaré and Francesca Ieva
- Optimal two-stage design based on error rates under a Bayesian perspective 833
Susanna Gentile and Valeria Sambucini

Migrants in Italy and return migration

- Comparing migrant and “native” Italian adolescents in risky behaviours from FSS and SHARE Corona surveys n.a.
Daniela Foresta
- EU-Border crisis on Twitter: sentiments and misinformation analysis 839
Elena Ambrosetti, Cecilia Fortunato and Sara Miccoli

| | |
|---|-----|
| Graduates' interregional migration in times of crisis: the Italian case Thaís García-Pereiro, Ivano Dileo and Anna Paterno | 843 |
| Intentions to stay: The experience of return migrants in Albania Maria Carella, Thaís García-Pereiro, Roberta Pace and Anna Paterno | 848 |
| Return migration to home country: a systematic literature review with text mining and topic modelling Cecilia Fortunato, Andrea Iacobucci and Elena Ambrosetti | 853 |
| The allocation of time within native and foreign couples living in Italy Giovanni Busetta, Maria Gabriella Campolo and Antonino Di Pino Incognito | 860 |
| Ειλεΐθυια comes from afar: The foreigners' contribution to fertility by Italian provinces Eleonora Miaci, Cristina Giudici, Eleonora Trappolini, Marina Attili, Cinzia Castagnaro and Antonella Guarneri | 866 |
| Sustainability assessment | |
| ESG, sustainability and stock market risk Michele Costa | 871 |
| Exploring the effect of consumer motivation and perception of sustainability on food choices with a Discrete Choice Experiment Gloria Solano-Hermosilla, Jesus Barreiro-Hurle and Iliaria Amerise | 875 |
| Sustainability explained by ChatGPT artificial intelligence in a HITL perspective: innovative approaches Vito Santarcangelo, Angelo Lamacchia, Emilio Massa, Saverio Gianluca Crisafulli, Massimiliano Giacalone and Vincenzo Basile | 881 |
| Measuring economic and ecological efficiency of urban waste systems in Italy: a comparison of SFA and DEA techniques Massimo Gastaldi, Ginevra Virginia Lombardi, Agnese Rapposelli and Giulia Romano | 887 |
| Profile based latent distance association analysis for sparse tables. Application to the attitude of EU citizens towards sustainable tourism Francesca Bassi, José Fernando Vera and Juan Antonio Marmolejo Martin | 893 |
| Sustainable tourism: a survey on the propensity towards eco-friendly accommodations Claudia Furlan and Giovanni Finocchiaro | 899 |
| Bayesian methods and applications 2 | |
| A comparison of computational approaches for posterior inference in Bayesian Poisson regression Laura D'Angelo | 903 |
| Bias-reduction methods for Poisson regression models Luca Presicce, Tommaso Rigon and Emanuele Aliverti | 908 |
| Finite Mixture Model for Multiple Sample Data Alessandro Colombi, Raffaele Argiento, Federico Camerlenghi and Lucia Paci | 913 |

| | |
|--|-----|
| On Bayesian power analysis in reliability | 918 |
| Fulvio De Santis, Stefania Gubbiotti and Francesco Mariani | |
| Power priors elicitation through Bayes factors | 923 |
| Roberto Macri Demartino, Leonardo Egidi and Nicola Torelli | |
| Predictive Bayes factors | 929 |
| Leonardo Egidi and Ioannis Ntzoufras | |
| Clustering and classification 2 | |
| A Clusterwise Regression Method for Distributional-Valued Data | 935 |
| Antonio Balzanella, Rosanna Verde and Francisco de A.T. de Carvalho | |
| A novel statistical-significance based semi-parametric GLMM for clustering countries standing on their innumeracy levels | 939 |
| Alessandra Ragni, Chiara Masci, Francesca Ieva and Anna Maria Paganoni | |
| Introducing a novel directional distribution depth function for supervised classification | 945 |
| Edoardo Redivo and Cinzia Viroli | |
| Clustering alternatives in the preference-approval context | 950 |
| Alessandro Albano, José Luis Garcia-Lapresta , Mariangela Sciandra and Antonella Plaia | |
| Computational assessment of k-means clustering on a Structural Equation Model based index | 955 |
| Mariaelena Bottazzi Schenone, Elena Grimaccia and Maurizio Vichi | |
| Handling missing data in complex phenomena: an ultrametric model-based approach for clustering | 961 |
| Francesca Greselin and Giorgia Zaccaria | |
| Economics and labour markets | |
| A multivariate ranking analysis on the employability of young adults | 967 |
| Rosa Arboretti, Elena Barzizza, Nicolo Biasetton, Riccardo Ceccato, Monica Fedeli and Concetta Tino | |
| Analysis of the Gender Pay Gap in the Italian Labour Market | 973 |
| Giulia Cappelletti and Daniele Toninelli | |
| Evaluating the effect of home-based working employing causal Bayesian networks and potential outcomes | 979 |
| Lorenzo Giammei | |
| Patterns of flexible employment careers. Does measurement error matter? | 985 |
| Mauricio Garnier-Villarreal, Dimitris Pavlopoulos and Roberta Varriale | |
| Staying or leaving? A nonlinear framework to explore the role of employee well-being on retention | 991 |
| Ulpiani Kocollari, Fabio Demaria and Maddalena Cavicchioli | |
| The CAP instruments impact on GVA and employment: a multivalued treatment approach | 997 |
| Montezuma Dumangane and Marzia Freo | |

| | |
|---|------|
| The determinants of leaving the parental home in Italy: 2012-18 Ilaria Rocco and Gianpiero Dalla Zuanna | 1003 |
| Environmental modeling | |
| A Bayesian weather-driven spatio-temporal model for PM10 in Lombardy Michela Frigeri, Alessandra Guglielmi and Giovanni Lonati | 1109 |
| A preliminary study on shape descriptors for the characterization of microplastics ingested by fish Greta Panunzi, Tommaso Valente, Marco Matiddi and Giovanna Jona Lasinio | 1015 |
| Artificial neural network in predicting odour concentrations: a case study Veronica Distefano and Gideon Mazuruse | 1021 |
| Bayesian analysis of PM10 concentration by spatio-temporal ARIMA and STS models Michela Frigeri and Ilenia Epifani | 1026 |
| Functional ANOVA to monitor yearly Adriatic sea temperature variations Annalina Sarra, Adelia Evangelista, Tonio Di Battista and Nicola Di Deo | 1032 |
| New perspectives in the measurement of biodiversity Linda Altieri, Daniela Cocchi and Massimo Ventrucci | 1038 |
| Multivariate data analysis 2 | |
| Feature Selection via anomaly detection autoencoders in radiogenomics studies Alessia Mapelli, Michela Carlotta Massi, Nicola Rares Franco, Francesca Ieva, Catharine West, Petra Seibold, Jenny Chang-Claude and the REQUITE and RADprecise Consortia | 1044 |
| Further considerations on the Spectral Information Criterion Luca Martino | 1050 |
| How to increase the power of the test in sparse contingency tables: a simulation study Federica Nicolussi and Manuela Cazzaro | 1057 |
| Latent event history models for quasi-reaction systems Matteo Framba, Veronica Vinciotti and Ernst Wit | 1063 |
| Quantile-based graphical models for continuous and discrete variables Luca Merlo, Marco Geraci and Lea Petrella | 1069 |
| The logratio Student t distribution Gianna Monti and Gloria Mateu-Figueras | 1075 |
| Statistics in Society 2 | |
| A decomposition of the changes in tourism demand in Tuscany over the 2019-2021 period Mauro Mussini | 1079 |
| Bayesian networks as a territorial gender impact assessment tool Flaminia Musella, Lorenzo Giammei, Fulvia Mecatti and Paola Vicar | 1084 |

| | |
|--|------|
| Can statistics be helpful in detecting electoral fraud? Massimo Attanasio, Vincenzo G. Genova and Michele Tumminello | 1088 |
| Companies' sustainability disclosure and contrast to hunger: the role of social inclusion Chiara Di Maria and Rodolfo Damiano | 1093 |
| Passing network-based performance indicator in football: evidence from UEFA Champions League 2016-2017 Riccardo Ievoli, Lucio Palazzo and Giancarlo Ragozini | 1099 |
| Topic Modeling for the travel and tourism industry: classical and innovative methods compared Fabrizio Di Mari | 1105 |
| Bayesian methods and applications 3 | |
| An Importance Sampling Algorithm For Bayesian Logistic Regression with Independent Gaussian Scale Mixture Prior Paolo Onorati and Brunero Liseo | 1111 |
| Bayesian analysis of Amazon's best-selling books via finite nested mixture model Laura D'Angelo and Francesco Denti | 1117 |
| Binomial Extended Stochastic Block Model for Brain Networks Valentina Ghidini, Sirio Legramanti and Raffaele Argiento | 1121 |
| Detecting latent spatial patterns in mass spectrometry brain imaging data via Bayesian mixtures Giulia Capitoli, Simone Colombara, Alessia Cotroneo, Francesco De Caro, Riccardo Morandi, Chiara Schembri, Alfredo G. Zapiola and Francesco Denti | 1127 |
| Efficient expectation propagation for high-dimensional probit models Augusto Fasano, Niccolò Anceschi, Beatrice Franzolini and Giovanni Rebaudo | 1133 |
| Model-based clustering of non-stationary time series with common historical change times Riccardo Corradin, Luca Danese, Wasiur KhudaBukhsh and Andrea Ongaro | 1139 |
| Functional Data Analysis | |
| A functional Ground Motion Model for Italy built with a weighted analysis of reconstructed seismic curves Teresa Bortolotti, Riccardo Peli, Giovanni Lanzano, Sara Sgobba and Alessandra Menafoglio | 1145 |
| Conditional Gaussian Graphical Models for Functional Variables with Partial Separable Operators Rita Fici, Gianluca Sottile and Luigi Augugliaro | 1149 |
| Does the Inflation Factor need tuning? Simulation-based adjustment for Outlier Detection via the Functional Boxplot Annachiara Rossi, Andrea Cappozzo and Francesca Ieva | 1155 |
| Functional Graphical Models to map Brexit debate on Twitter Nicola Pronello, Emiliano del Gobbo, Lara Fontanella, Rosaria Ignaccolo, Luigi Ippoliti and Sara Fontanella | 1160 |

| | |
|--|------|
| Measuring Dependence in Multivariate Functional Datasets Francesca Ieva, Michael Ronzulli and Anna Maria Paganoni | 1166 |
| Robust Statistical Process Monitoring of Multivariate Functional Data Christian Capezza, Fabio Centofanti, Antonio Lepore and Biagio Palumbo | 1173 |
| The effects of mobility restrictions on public health: a functional data analysis for Italy over the years 2020 and 2021 Veronica Mazzola, Giovanni Bonaccorsi, Piercesare Secchi and Francesca Ieva | 1179 |
| Machine Learning and text mining | |
| A vocabulary-based approach for risk detection in textual annotations of contracts of public procurement Giulio Giacomo Cantone, Simone Del Sarto and Michela Gnaldi | 1185 |
| Explainable Machine Learning based on Group Equivariant Non-Expansive Operators (GENEOs). Protein pocket detection: a case study Giovanni Bocchi, Alessandra Micheletti, Patrizio Frosini, Alessandro Pedretti, Andrea R. Beccari, Filippo Lunghini, Carmine Talarico and Carmen Gratteri | 1191 |
| Hedging global currency risk with factorial machine learning models Paolo Pagnottoni and Alessandro Spelta | 1197 |
| InstanceSHAP: An instance-based estimation approach for Shapley values Golnoosh Babaei and Paolo Giudici | 1203 |
| Networks & Nature Based Solutions: an application for Milan hydric resources Alessia Forciniti and Emma Zavarrone | 1209 |
| The Roe v. Wade sentence: an analysis of tweets trough Symmetric Non-Negative Matrix Factorization Maria Gabriella Grassia, Marina Marino, Rocco Mazza and Agostino Stavolo | 1215 |
| Multivariate data analysis 3 | |
| A comparison of different techniques for handling missing covariate values in propensity score methods Anna Zanovello, Alessandra R. Brazzale and Omar Paccagnella | 1219 |
| A New Penalized Estimator for Sparse Inference in Gaussian Graphical Models: An Adaptive Non-Convex Approach Daniele Cuntrera, Vito M.R. Muggeo and Luigi Augugliaro | 1224 |
| A tool for assessing weak identifiability of statistical models Antonio Di Noia, Francesco Denti and Antonietta Mira | 1230 |
| Computing Highest Density Regions with Copulae Nina Deliu and Brunero Liseo | 1235 |
| Parameter estimation via Indirect Inference for multivariate Wrapped Normal distributions Francesca Labanca and Anna Gottard | 1241 |

| | |
|--|-------------|
| Sequential marginal likelihood selection for the estimation of sparse correlation matrices | 1246 |
| Claudia Di Caterina and Davide Ferrari | |
| Nonparametric statistical methods | |
| A Comparison of Distribution-Free Control Charts | 1252 |
| Michele Scagliarini | |
| Characterizing Heterogeneity of Causal Effects in Air Pollution in Florida | 1257 |
| Dafne Zorzetto | |
| Comparing three robust procedures for CANDECOMP/PARAFAC estimation | 1262 |
| Valentin Todorov, Violetta Simonacci, Michele Gallo and Nikolay Trendafilov | |
| How active is a genetic pathway? Comparative analysis of post-hoc permutation-based methods | 1268 |
| Anna Vesely and Angela Andreella | |
| Non Parametric Combination methodology: a literature review on recent developments | 1274 |
| Elena Barzizza, Nicolò Biasetton and Riccardo Ceccato | |
| Regression modeling | |
| A Quantile Regression Model to Evaluate the Performance of the Italian Courts of Law | 1280 |
| Carlo Cusatelli, Massimiliano Giacalone and Eugenia Nissi | |
| A variable selection procedure based on predictive ability: a preliminary study on logistic regression | 1285 |
| Rosaria Simone and Mariarosaria Coppola | |
| Comparison of binary regressions with asymmetric link function for imbalanced data | 1291 |
| Michele La Rocca, Marcella Niglio and Marialuisa Restaino | |
| New advances in Regression Forests | 1297 |
| Mila Andreani, Lea Petrella and Nicola Salvati | |
| On the Optimal Non-Convexity of Penalty in Sparse Regression Models | 1303 |
| Daniele Cuntreza, Vito M.R. Muggeo and Luigi Augugliaro | |
| Using expectile regression with latent variables for digital assets | 1309 |
| Beatrice Foroni, Luca Merlo and Lea Petrella | |
| 4 Program | 1315 |

Preface

This book includes the contributions presented at the Intermediate Meeting of the Italian Statistical Society (SIS) "SIS 2023 - Statistical Learning, Sustainability and Impact Evolution" held in Ancona at the Università Politecnica delle Marche, from June 21th to 23th of 2023.

The new challenges of digitalization, innovation and sustainability are showing the crucial role of data-driven approaches in supporting decision-making processes. Methodologies resulting from the integration of different know-how seem to be a reliable way to deal with the increasing need to measure the impact of the policies and to forecast scenarios. This meeting welcomed any attempt to face new challenges.

The conference registered more than 250 presentations, including 3 keynote speakers in 3 plenary sessions and 72 presentations in 24 invited sessions, all dealing with specific themes in methodological and/or applied statistics and demography. Furthermore, more than 180 contributions, with one or more authors, have been spontaneously submitted to the Program Committee and arranged in 30 contributed sessions.

The numerous participation of researchers in the conference shows how the challenges of sustainability, in its broadest sense, are of interest to both methodological and applied statistics.

With the publication of this book, we wish to offer to all members of the Italian Statistical Society, all international academics, researchers, Ph.D. students, and all interested practitioners, a good snapshot of the on-going research in the statistical and demographic fields.

We aim to provide all members of the Italian Statistical Society - as well as international academics, researchers, Ph.D. students, and interested practitioners - with a comprehensive overview of the ongoing research in the fields of statistics and demography.

We extend our heartfelt gratitude to all the contributors for submitting their works to the conference and to the researchers for their outstanding job in serving as referees and discussants with precision and timeliness.

A special appreciation goes to the Scientific and Organizational Committees for their tremendous efforts in managing all the organizational aspects, as well as to the Università Politecnica delle Marche and the Department of Economic and Social Science for making this event possible.

Finally, we wish to express our gratitude to the publisher Pearson Italia for all the support received.

Assessing multidimensional poverty of the Italian provinces during Covid-19: a small area estimation approach

Mariateresa Ciommi^a, Chiara Gigliarano^b, Francesca Mariani^a, and Gloria Polinesi^a

^aUniversità Politecnica delle Marche; m.ciommi@staff.univpm.it, f.mariani@staff.univpm.it, g.polinesi@staff.univpm.it

^bLIUC - Università Carlo Cattaneo; cgigliarano@liuc.it

Abstract

The aim of this paper is to analyse the effect of Covid-19 on multidimensional poverty in Italy and its provinces by measuring changes in individual poverty before and during the pandemic outbreak. To capture the multidimensional nature of poverty, we consider different dimensions: economic well-being, health condition, education, neighborhood quality, subjective well-being.

The empirical application is based on micro-data from the ‘Aspects of daily life’ (AVQ) survey by Istat for the years 2019 (pre-Covid period) and 2020 (Covid period).

Since survey direct estimates are reliable only at regional (NUTS 2) level, the introduction of small area estimation (SAE) techniques becomes of crucial importance to monitor and contrast the phenomenon at a finer geographical level.

Keywords: Multidimensional poverty index, composite indicators, small area estimation, EBLUP estimator.

1. Introduction

In the recent years there has been a considerable agreement that poverty is a multidimensional phenomenon, which cannot be adequately explained by considering only monetary variables. (1) confirms the existence of significant mismatches between monetary poverty and multiple deprivations and, therefore, the need to include many aspects of poverty, such as health, education, housing, satisfaction with life, security.

Few empirical studies have analyzed multidimensional poverty in Italy at sub-national level; see, among others, (5), (4) and (3). In this paper, we propose a deeper attempt in this direction by looking at the provincial level changes in the multidimensional poverty due to Covid-19 pandemic.

The empirical application is based on micro-data from the ‘Aspects of daily life’ (AVQ) Italian survey over the years 2019 (pre-Covid period) and 2020 (Covid period).

Since the AVQ survey is planned to obtain precise estimates at regional level, some alternative solutions should be evaluated to estimate poverty at provincial level. More specifically, two main possible strategies can be employed: i) increasing the sample size of AVQ for the specific province of interest (oversampling) so that direct estimates become reliable and ii) apply small area estimation (SAE) techniques; see, among others, (12) and (9).

We apply the Fay-Herriot (FH) small area estimation model proposed by (6) to estimate dashboard indicators at provincial level. Then, we propose an aggregation technique based on the Adjusted Mazziotta-Pareto Index (see (7) and (8)) for the FH estimates in order to obtain reliable estimates of the multidimensional poverty for the local areas of interest as in (11).

Therefore, the aim is to enhance the knowledge of the spatial distribution of the multidimensional poverty index at local level in Italy, focusing on Italian provinces, in order to help the policy maker to address resources towards the areas where the phenomenon is strongly present.

The empirical analysis reveals that the traditional Italian divide North vs. South becomes less clear when analyzing poverty at provincial level and for the different domains of poverty.

The remainder of the paper is as follows. Sections 2. and 3. describe data and methodologies considered in the analysis. Section 4. is devoted to results of the empirical application. Finally, Section 5. draws some conclusions.

2. Data

Our analysis is based on data from the ‘Aspects of daily life’ (AVQ) survey conducted by Istat, focusing on the years 2019 and 2020. The units of analysis are the Italian households and the original samples include 19536 and 18529 households, for the two years considered respectively.

To assess the multidimensional poverty over 105 Italian provinces¹, we consider five domains (health, education, economic well-being, neighborhood quality and subjective well-being), each composed of different dashboard indicators as described in Table 1.

In order to classify a household as deprived in each of the 13 dashboard indicators, we first identify the deprivation cut-offs following the approach proposed in (5), which we have slightly modified due to the data availability at provincial level. The deprivation cut-offs refers to the head of the household.

Table 1: Multidimensional poverty framework: domains, dashboard indicators, deprivation cut-offs.

| Domains | Deprivation Indicators | Deprivation cut-offs |
|------------------------------|---|--|
| <i>Health</i> | Nutrition | A person is deprived if s/he consumes less than 3 portions of fruits or vegetables a day. |
| <i>Education</i> | Educational deprivation Cultural deprivation | A person is deprived if s/he has not completed higher-secondary school. A person is deprived if in the 12 months before the interview s/he has joined less than 2 among the following activities: 1) at least once to cinema, theatre, exhibitions and museums, archaeological sites, monuments, concerts of classical music, opera, concerts of other kind of music; 2) read the newspaper at least once a week; 3) read at least a book. |
| <i>Economic well-being</i> | Material deprivation Housing deprivation Gas Water Unemployment Financial distress | A person is deprived if s/he possesses less than 4 out of 6 following items: washing machine, color tv, scooter/moto or car, phone, personal computer. A person is deprived if s/he experiences 3 or more among the following deprivations related to the house: overcrowding; distance from basic services (pharmacy, shops, school); overall poor condition of the floors and/or walls; expenses too high; house not owned). A person is deprived if the house is not served by methane gas. A person is deprived if s/he declares to have irregularities in their water supply. A person is deprived if s/he is unemployed. A person is deprived if her/his economic sources are not sufficient to make ends meet. |
| <i>Neighbourhood quality</i> | Noise Crime Pollution | A person is deprived if the area in which s/he lives is declared to be very noisy; at risk of crime; polluted. |
| <i>Subjective well-being</i> | Life satisfaction and future expectations | A person is deprived if experiences 3 or more deprivations related to personal satisfaction (life, economic situation, health, familiar and friends relationship, leisure and future expectations). |

¹The provinces of Vibo-Valentia and Benevento have been excluded from the analysis due the very high percentage of missing data.

3. Methodology

Small area estimation (SAE) combines survey data with auxiliary variables of the population of interest to break down regional estimates into sub-regional ones. These variables are commonly obtained from population censuses or from administrative registers. If auxiliary information are available at unit-level (e.g., individual or household-level) one can consider unit-level SAE models². When, on the contrary, the auxiliary data are available only at area-level (e.g., district, municipality or provincial level), one can use area-level SAE models.

We follow the latter approach and apply the Fay-Herriot area-level small area estimation model introduced by (6), which proposes the empirical best linear unbiased predictor (EBLUP) estimator to obtain estimates of the poverty dashboard indicators at provincial level³.

Consider a finite population U , partitioned into $d = 1, \dots, D$ mutually exclusive and exhaustive areas (in our case, provinces); the Fay-Herriot (FH) model is defined in two stages.

Let $\hat{\delta}_d^{DIR}$ be a direct estimator of δ_d , the parameter of inferential interest for area (province) d . In the first stage, we assume that $\hat{\delta}_d^{DIR}$ is an unbiased estimator of δ_d :

$$\hat{\delta}_d^{DIR} = \delta_d + e_d, \quad e_d \stackrel{ind}{\sim} N(0, \psi_d), \quad (1)$$

where ψ_d is the sampling variance of the direct estimator $\hat{\delta}_d^{DIR}$ given δ_d , assumed to be known for all $d = 1, \dots, D$.

In the second stage, we assume that the area parameters δ_d are linearly related with a p-vector x_d of area-level auxiliary variables as follows:

$$\delta_d = x_d^T \beta + u_d, \quad u_d \stackrel{ind}{\sim} N(0, A). \quad (2)$$

Model (1) is known as *sampling model* because it represents the uncertainty due to the fact that δ_d is unobservable and the direct estimator is based on the sample data, $\hat{\delta}_d^{DIR}$. While, model (2) is called *linking model* because it relates all areas through the common regression coefficients β , allowing us to borrow strength from all areas.

Combining the two model components (1) and (2), we obtain the linear mixed model:

$$\hat{\delta}_d^{DIR} = x_d^T \beta + u_d + e_d, \quad (3)$$

where u_d is independent of e_d .

The EBLUP estimator is the combination of the direct and the regression-synthetic estimators:

$$\hat{\delta}_d^{EBLUP} = \hat{\gamma}_d \hat{\delta}_d^{DIR} + (1 - \hat{\gamma}_d) x_d^T \hat{\beta} \quad (4)$$

where $\hat{\gamma} = \frac{\hat{A}}{\hat{A} + \hat{\psi}_d}$ represents the shrinkage factor.

For each of the 13 dashboard indicators of deprivation j and for each province d we obtain an EBLUP estimate, denoted with $\hat{\delta}_{dj}^{EBLUP}$. For sake of simplicity we will denote estimator $\hat{\delta}_{dj}^{EBLUP}$ with $\hat{\delta}_{dj}$.

Then, in order to obtain a composite indicator of multidimensional poverty, we aggregate the EBLUP provincial estimates $\hat{\delta}_{dj}$, where d indicates the province and j the dashboard indicator, using the Adjusted Mazziotta-Pareto Index (AMPI); see (8).

The AMPI approach considers the normalized provincial estimates r_{dj} defined as:

$$r_{dj} = \frac{\hat{\delta}_{dj} - \text{Min}(\hat{\delta}_{dj})}{\text{Max}(\hat{\delta}_{dj}) - \text{Min}(\hat{\delta}_{dj})} 60 + 70. \quad (5)$$

Denoting with Ref_{δ_j} the reference value for the indicator j , the ‘goalposts’ $\text{Max}(\hat{\delta}_{dj})$ and $\text{Min}(\hat{\delta}_{dj})$ are defined as:

$$\begin{cases} \text{Min}(\hat{\delta}_{dj}) = Ref_{\delta_j} - \Delta \\ \text{Max}(\hat{\delta}_{dj}) = Ref_{\delta_j} + \Delta \end{cases} \quad (6)$$

²The Battese-Harter-Fuller (BHF) model is one of the most used unit-level small area estimation model (2).

³See (10) for notation details.

where $\Delta = \frac{Sup_{\delta_j} - Inf_{\delta_j}}{2}$. Inf_{δ_j} and Sup_{δ_j} are the overall minimum and maximum of the indicator $\hat{\delta}_{dj}$ across all provinces and the two years considered. To facilitate the interpretation of results, we choose as Ref_{δ_j} the mean of the year 2019. The normalized values r_{dj} will fall approximately in the range (70; 130), where 100 represents the reference value.

Denoting with M_d and S_d , respectively, the mean and standard deviation of the normalized indicators r_{dj} of province d , the multidimensional poverty index of province d is computed as:

$$AMPI_d = M_d + S_d \times cv_d \quad (7)$$

where $cv_d = \frac{S_d}{M_d}$ is the coefficient of variation for the local unit d ; see (8).

4. Empirical findings

Figures 1 and 2 illustrate a geographical representations of how the multidimensional poverty indicator as well as the domain-specific deprivation indices are distributed across the Italian provinces in the years 2019 and 2020, respectively. In each panel, value 100 corresponds to the Italian average in 2019, and darker colors indicate higher levels of deprivation.

Comparing Figures 1(d) and 2(d) it emerges how health deprivation strongly worsened during the Covid-19 pandemic especially in Sicily, Sardinia, Lombardy and Piedmont and for the provinces located along the Tyrrhenian coast. On the contrary, Figures 1(e) and 2(e) reveal that life satisfaction and future expectations have not been affected by Covid-19.

Overall, the multidimensional poverty index (Figures 1(a) vs. 2(a)) does not seem to have deteriorated after the first year of the pandemic, although its levels remain higher in the South of Italy. We note that some domains of poverty (economic, education, health) have worsened in many provinces, but their effects have been compensated by the opposite trend registered for subjective well-being and quality of neighborhood. Anyway, the result refers only to the first year of the pandemic. In order to have a complete overview of the effect of Covid-19, one should analyze more recent data.

5. Conclusion

In this paper we have estimated the multidimensional poverty of the Italian households during the period of Covid-19 at provincial level. Since the AVQ survey is planned to obtain precise estimates at regional level, we have applied the FH small area estimation model in order to obtain reliable estimates of the dashboard indicators of deprivation at this finer geographical level. The overall composite indicator of multidimensional poverty has been then obtained using the AMPI approach.

Empirical findings suggest that, among the different domains of poverty considered, health deprivation has strongly worsened during the first year of the pandemic especially for provinces of the Islands (Sicily and Sardinia), Lombardy, Piedmont and the whole region located along the Tyrrhenian coast. Subjective deprivation, on the contrary, has reduced in the period of analysis. Overall, data reveals that the multidimensional poverty has not worsened relevantly, although the provinces of southern Italy are multidimensionally poorer than those of the north.

Further research could be devoted to the use of different techniques of aggregation of the dashboard indicators.

Acknowledgments

This research has been founded by Fondazione Cariplo, under the project *POST-COVID: POverty and vulnerability Scenarios in The era of Covid-19: how the pandemic is affecting the well-being of the Italians* - rif. 2020-4216.

The data used in this paper are from Istat's Aspects of Daily Life ('Aspetti di vita quotidiana') survey. The elaborations were carried out at Istat's ADELE ('Analisi dei Dati ELEMENTARI') laboratory

and in compliance with the legislation on protection of statistical confidentiality and personal data. The responsibility for results and opinions expressed in the paper rests solely with the authors. The analyses have been carried out without sampling weights.

References

- [1] Atkinson, A. B.: Multidimensional deprivation: contrasting social welfare and counting approaches. *The Journal of Economic Inequality* **1** (1), 5–65 (2003)
- [2] Battese, G. E., Harter, R. M., Fuller, W. A.: An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association* **83** (401), 28–36 (1988)
- [3] Betti, G., Cheli, B., Lemmi, A., Verma, V.: The fuzzy set approach to multidimensional poverty: the case of Italy in the 1990s. In: Kakwani, N., Silber, J. (eds). *Quantitative Approaches to Multidimensional Poverty Measurement*. Palgrave Macmillan, London (2008).
- [4] Coromaldi, M., Zoli, M.: Deriving multidimensional poverty indicators: Methodological issues and an empirical analysis for Italy. *Social Indicators Research* **107**, 37–54 (2012)
- [5] De Rosa, D.: Are Italians getting multidimensionally poorer? Evidence on the lack of equitable and sustainable well-being. *Italian Economic Journal* **8** (1), 145–174 (2022)
- [6] Fay, R. E., Herriot, R. A.: Estimation of income from small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association* **74** (366), 269–277 (1979)
- [7] Mazziotta, M., Pareto, A.: On a generalized non-compensatory composite index for measuring socio-economic phenomena. *Social Indicators Research* **127** (3), 983–1003 (2016)
- [8] Mazziotta, M., Pareto, A.: Measuring well-being over time: the adjusted Mazziotta-Pareto index versus other non-compensatory indices. *Social Indicators Research* **136** (3), 967–976 (2018)
- [9] Molina, I., Rao, J. N. K.: Small area estimation of poverty indicators. *The Canadian Journal of Statistics* **38** (3), 369–385 (2010)
- [10] Molina, I., Marhuenda, Y.: Sae: an R package for small area estimation. *The R Journal* **7** (1), 81 (2015)
- [11] Pratesi, M., Quattrocioni, L., Bertarelli, G., Gemignani, A., Giusti, C.: Spatial distribution of multidimensional educational poverty in Italy using small area estimation. *Social Indicators Research* **156**, 563–586 (2021)
- [12] Rao, J. N. K.: Some new developments in small area estimation. *JIRSS* (2003)