



SIS | 2022

51st Scientific Meeting
of the Italian Statistical Society

Caserta, 22-24 June

V: Università
degli Studi
della Campania
Luigi Vanvitelli

SIS
Società
Italiana di
Statistica



www.unicampania.it



Book of the Short Papers

**Editors: Antonio Balzanella, Matilde Bini,
Carlo Cavicchia, Rosanna Verde**



1222-2022
800
ANNI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI SCIENZE
STATISTICHE



sas

UNIVERSITÀ
DEGLI STUDI
DEL
SANNIO
Benevento

P
Pearson

Matilde Bini (Chair of the Program Committee) - *Università Europea di Roma*

Rosanna Verde (Chair of the Local Organizing Committee) - *Università della Campania "Luigi Vanvitelli"*

PROGRAM COMMITTEE

Matilde Bini (Chair), Giovanna Boccuzzo, Antonio Canale, Maurizio Carpita, Carlo Cavicchia, Claudio Conversano, Fabio Crescenzi, Domenico De Stefano, Lara Fontanella, Ornella Giambalvo, Gabriella Grassia - Università degli Studi di Napoli Federico II, Tiziana Laureti, Caterina Liberati, Lucio Masserini, Cira Perna, Pier Francesco Perri, Elena Pirani, Gennaro Punzo, Emanuele Raffinetti, Matteo Ruggiero, Salvatore Strozza, Rosanna Verde, Donatella Vicari.

LOCAL ORGANIZING COMMITTEE

Rosanna Verde (Chair), Antonio Balzanella, Ida Camminatiello, Lelio Campanile, Stefania Capecchi, Andrea Diana, Michele Gallo, Giuseppe Giordano, Ferdinando Grillo, Mauro Iacono, Antonio Irpino, Rosaria Lombardo, Michele Mastroianni, Fabrizio Maturo, Fiammetta Marulli, Paolo Mazzocchi, Marco Menale, Giuseppe Pandolfi, Antonella Rocca, Elvira Romano, Biagio Simonetti.

ORGANIZERS OF SPECIALIZED, SOLICITED, AND GUEST SESSIONS

Arianna Agosto, Raffaele Argiento, Massimo Aria, Rossella Berni, Rosalia Castellano, Marta Catalano, Paola Cerchiello, Francesco Maria Chelli, Enrico Ciavolino, Pier Luigi Conti, Lisa Crosato, Marusca De Castris, Giovanni De Luca, Enrico Di Bella, Daniele Durante, Maria Rosaria Ferrante, Francesca Fortuna, Giuseppe Gabrielli, Stefania Galimberti, Francesca Giambona, Francesca Greselin, Elena Grimaccia, Raffaele Guetto, Rosalba Ignaccolo, Giovanna Jona Lasinio, Eugenio Lippiello, Rosaria Lombardo, Marica Manisera, Daniela Marella, Michelangelo Misuraca, Alessia Naccarato, Alessio Pollice, Giancarlo Ragozini, Giuseppe Luca Romagnoli, Alessandra Righi, Cecilia Tomassini, Arjuna Tuzzi, Simone Vantini, Agnese Vitali, Giorgia Zaccaria.

ADDITIONAL COLLABORATORS TO THE REVIEWING ACTIVITIES

Ilaria Lucrezia Amerise, Ilaria Benedetti, Andrea Bucci, Annalisa Busetta, Francesca Condino, Anthony Corsari, Paolo Carmelo Cozzucoli, Simone Di Zio, Paolo Giudici, Antonio Irpino, Fabrizio Maturo, Elvira Romano, Annalina Sarra, Alessandro Spelta, Manuela Stranges, Pasquale Valentini, Giorgia Zaccaria.

Copyright © 2022

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891932310

Contents

| | |
|--|--------------|
| Preface | XXVII |
| 1 Plenary Sessions | 1 |
| Causal inference in air pollution epidemiology <i>Francesca Dominici</i> | 2 |
| Clustering of Attribute Data and Network <i>Anuška Ferligoj</i> | 11 |
| Bayesian approaches for capturing the heterogeneity of neuroimaging experiments <i>Francesco Denti, Laura D'Angelo and Michele Guindani</i> | 17 |
| 2 Specialized Sessions | 30 |
| Advances in Bayesian nonparametric methodology | 31 |
| Repulsive mixture models for high-dimensional data <i>Lorenzo Ghilotti, Mario Beraha and Alessandra Guglielmi</i> | 32 |
| Bayesian nonparametric mixtures of directed acyclic graph models <i>Federico Castelletti and Guido Consonni</i> | 37 |
| Bayesian Clustering of Brain Regions via Extended Stochastic Block Models <i>Sirio Legramanti, Tommaso Rigon and Daniele Durante</i> | 45 |
| Data Science skills for next generation statisticians | 52 |
| Cluster based oversampling for imbalanced learning <i>Giola Di Credico and Nicola Torelli</i> | 53 |
| Estimating the effect of remote teaching for university students through generalised linear mixed models <i>Silvia Bacci, Bruno Bertaccini, Simone Del Sarto, Leonardo Grilli and Carla Rampichini</i> | 65 |
| Perceived stress across EU countries: does working from home impact? <i>Stefania Capecchi, Francesca Di Iorio and Nunzia Nappo</i> | 71 |

| | |
|--|------------|
| Investigating effects of air pollution on health: a challenge for statisticians | 77 |
| Investigating effect of air pollution on health via Spatial-Resolution Varying Coefficient Models <i>Garritt L. Page and Massimo Ventrucchi</i> | 78 |
| A statistical framework for evaluating health effect of PM sources <i>Monica Pirani, Georges Bucyibaruta, Gary Fuller, David Green, Anja Tremper, Christina Mitsakou and Marta Blangiardo</i> | 84 |
| Adjusting for unmeasured spatial confounding through shrinkage methods <i>Pasquale Valentini, Alexandra M. Schmidt, Carlo Zaccardi and Luigi Ippoliti</i> | 91 |
| Explainable Artificial Intelligence methods | 98 |
| Multidimensional Time Series Analysis via Bayesian Matrix Auto Regression <i>Alessandro Celani and Paolo Pagnottoni</i> | 99 |
| Advances in Classification and Data Analysis | 109 |
| Optimizing time slots in scientific meetings: a Latent Dirichlet allocation approach <i>Luca Frigau</i> | 110 |
| Clustering artists based on the energy distributions of their songs on Spotify via the Common Atoms Model <i>Francesco Denti, Federico Camerlenghi, Michele Guindani and Antonietta Mira</i> | 121 |
| Hidden markov models for four-way data <i>Salvatore D. Tomarchio, Antonio Punzo and Antonello Maruotti</i> | 127 |
| Family demography | 133 |
| Does family of origin make the difference in occupational outcomes? <i>Annalisa Busetta, Elena Fabrizi, Isabella Sulis and Giancarlo Ragozini</i> | 134 |
| Is there a cultural driver pushing Italian low fertility? <i>Francesca Luppi, Alessandro Rosina and Maria Rita Testa</i> | 144 |
| Unpaid family work and the subjective well-being of Italian women during lockdown <i>Marina Zannella, Erica Aloé, Marcella Corsi and Alessandra de Rose</i> | 155 |
| New Frontiers in the theory of composite indicators | 164 |
| Methodological PLS-PM Framework for Model Based Composite Indicators <i>Rosanna Cataldo</i> | 165 |
| Open issues in composite indicators construction <i>Leonardo Salvatore Alaimo</i> | 176 |
| The posetic approach to the construction of socio-economic indicators: open issues and research opportunities <i>Marco Fattore</i> | 186 |

| | |
|--|------------|
| Advances in complex sampling strategies | 197 |
| Random forest model-assisted estimation for finite population totals <i>Mehdi Dagdoug, Camelia Goga and David Haziza</i> | 198 |
| Design-based consistency of the Horvitz-Thompson estimator in spatial sampling <i>Lorenzo Fattorini</i> | 208 |
| The responsive-adaptive survey design approach for planning the permanent census of population and housing <i>Claudia De Vitiis, Stefano Falorsi, Alessio Guandalini, Francesca Inglese, Paolo Righi and Marco D. Terribili</i> | 216 |
| Socio-demographic aspects of aging in Italy | 228 |
| Socio-economic and spatial stratification of frailty in the older population <i>Margherita Silan</i> | 229 |
| Time allocation and wellbeing in later life: the case of Italy <i>Annalisa Donno and Maria Letizia Tanturi</i> | 241 |
| The role played by migration and fertility on Italy's demographic aging trends: a provincial-level analysis <i>Thais García-Pereiro and Anna Paterno</i> | 250 |
| New challenges in the labour market | 260 |
| Detecting changes and evolution in specialized professional figures: an application on the Italian IT & Digital sector <i>Andrea Marletta</i> | 261 |
| How did the COVID-19 pandemic affect the genderpay gap in EU countries? <i>Antonella Rocca, Paolo Mazzocchi, Giovanni De Luca, Rosalia Castellano and Claudio Quintano</i> | 272 |
| Skill Similarities and Dissimilarities in Online Job Vacancy Data across Italian Regions <i>Adham Kahlawi, Lucia Buzzigoli, Laura Grassini and Cristina Martelli</i> | 284 |
| Small area estimation methods with socioeconomic applications | 292 |
| Exploring Small Area Estimation techniques to address uncertainty in Spatial Price Indexes <i>Ilaria Benedetti and Federico Crescenzi</i> | 293 |
| Small Area Estimation of Relative Inequality Indices using Mixture of Beta <i>Silvia De Nicolò and Silvia Pacei</i> | 301 |
| Inference for big data assisted by small area methods: an application to OBEC (on-line based enterprise characteristics) <i>Monica Pratesi, Francesco Schirripa Spagnolo, Gaia Bertarelli, Stefano Marchetti, Monica Scannapieco, Nicola Salvati and Donato Summa</i> | 305 |

| | |
|--|------------|
| Statistical methods and models for Sports Analytics | 312 |
| The 'hot shoe' in soccer penalty shootouts <i>Andreas Groll and Marius Otting</i> | 313 |
| G-RAPM: revisiting player contributions in regularized adjusted plus-minus models for basketball analytics <i>Luca Grassetti</i> | 319 |
| Formative vs Reflective constructs: a CTA-PLS approach on a goalkeepers' performance model <i>Mattia Cefis and Eugenio Brentari</i> | 323 |
| Integrating available Data Sources for Official Statistics | 329 |
| The Use of Administrative Data for the Estimation of Italian Usually Resident Population <i>Marco Caputi, Giampaolo De Matteis, Gerardo Gallo and Donatella Zindato</i> | 330 |
| New frontiers for the analysis of the territorial economic phenomena | 339 |
| An empirical tool to classify industries by regional concentration and spatial polarization <i>Diego Giuliani, Maria Michela Dickson, Flavio Santi and Giuseppe Espa</i> | 340 |
| Comparing non-compensatory composite indicators: a case study based on SDG for Mediterranean countries <i>Francesca Mariani, Mariateresa Ciommi, Maria Cristina Recchioni, Giuseppe Ricciardo Lamonica and Francesco Maria Chelli</i> | 346 |
| Evaluating the determinants of innovation from a spatio-temporal perspective. The GWPR approach <i>Gaetano Musella, Giorgia Riveccio and Emma Bruno</i> | 354 |
| Dimension Reduction for complex data | 366 |
| Discrimination and clustering via principal components <i>Nikolay Trendafilov and Violetta Simonacci</i> | 367 |
| Exploratory graph analysis for configural invariance assessment <i>Sara Fontanella, Alex Cucco and Nicola Pronello</i> | 373 |
| Penalized likelihood factor analysis <i>Kei Hirose</i> | 379 |

| | |
|---|------------|
| 3 Solicited Sessions | 385 |
| Bayesian nonparametric modelling and learning | 386 |
| A regularized-entropy estimator to enhance cluster interpretability in Bayesian nonparametrics <i>Beatrice Franzolini and Giovanni Rebaudo</i> | 387 |
| Exact confidence sets from credible sets with finite amounts of data <i>Bas J. K. Kleijn</i> | 399 |
| Empirical Bayesian analysis of componentwise maxima in multivariate samples <i>Simone A. Padoan and Stefano Rizzelli</i> | 411 |
| Processing of textual data in large corpora | 420 |
| Predictive performance comparisons of different feature extraction methods in a financial column corpus <i>Andrea Sciandra and Riccardo Ferretti</i> | 421 |
| Topics and trends in the End-of-Year addresses of the Presidents of the Italian Republic (1949-2021) <i>Matilde Trevisani and Arjuna Tuzzi</i> | 428 |
| Thematic analysis on online education issues during COVID-19 <i>Valerio Basile, Michelangelo Misuraca and Maria Spano</i> | 437 |
| What do we learn by applying multiple methods in topic detection? A comparative analysis on a large online dataset about mobility electrification <i>Fabrizio Alboni, Margherita Russo and Pasquale Pavone</i> | 446 |
| Businesses in industry: new challenges in sustainability, innovation, performance and competitiveness | 454 |
| Multidimensional assessment of Eco-Innovation and its link with Marketing Innovations <i>Ida D'Attoma and Marco Ieva</i> | 455 |
| Circular Economy practices in the European SMEs: company-level and country-level drivers <i>Francesca Bassi, José G. Dias and Nunzio Tritto</i> | 462 |
| The employment effects of Italian Jobs Act. An ex-post impact evaluation <i>Alessandro Zeli and Leopoldo Nascia</i> | 474 |
| Statistics for finance: new models, new data | 482 |
| The News-Jumps Relationship in the Cryptocurrency Market <i>Ahmet Faruk Aysan, Massimiliano Caporin, Oguzhan Cepni, and Francesco Poli</i> | 483 |
| A weighted quantile approach to Expected Shortfall forecasting <i>Giuseppe Storti and Chao Wang</i> | 489 |

| | |
|--|------------|
| Smooth and abrupt dynamics in financial volatility: the MS-MEM-MIDAS <i>Giampiero M. Gallo, Edoardo Otranto and Luca Scaffidi Domianello</i> | 492 |
| The tail index and related quantities for volatility models <i>Fabrizio Laurini</i> | 501 |
| Bayesian inference for complex random structures | 507 |
| Bayesian nonparametric modeling of mortality curves via functional Dirichlet processes <i>Emanuele Aliverti and Bruno Scarpa</i> | 508 |
| Bayesian nonparametric clustering of spatially-referenced spike train data <i>Laura D'Angelo</i> | 514 |
| Bayesian Analysis of Mortality in Iceland via Locally Adaptive Splines <i>Federico Pavone and Sirio Legramanti</i> | 520 |
| Advances in clustering | 526 |
| A Two-step Latent Class Approach with Measurement Equivalence Testing <i>Zsuzsa Bakk, Roberto Di Mari, Jennifer Oser and Marc Hooghe</i> | 527 |
| Group-wise penalized estimation schemes in model-based clustering <i>Alessandro Casa, Andrea Cappozzo and Michael Fop</i> | 534 |
| Extending finite mixtures of latent trait analyzers for bipartite networks <i>Dalila Failli, Maria Francesca Marino and Francesca Martella</i> | 540 |
| A Fast Majorization-Minimization Algorithm for Convex Clustering <i>Daniel J.W. Touw, Patrick J.F. Groenen and Yoshikazu Terada</i> | 551 |
| Statistical Methods for Complex Evolutionary Data | 558 |
| A FANOVA model with repeated measures for detecting patterns in biomechanical data <i>Ana M. Aguilera, Christian Acal and Manuel Escabias</i> | 559 |
| Modes of variation for Lorenz curves <i>Enea G. Bongiorno and Aldo Goia</i> | 565 |
| Analyzing textual data through Word Embedding: experiences in Istat <i>Mauro Bruno, Elena Catanese, Massimo De Cubellis, Fabrizio De Faustis, Francesco Pugliese, Monica Scannapieco and Luca Valentino</i> | 571 |
| Functional Horvitz-Thompson estimator for convex curves <i>Adella Evangelista, Francesca Fortuna, Stefano Antonio Gattone and Tonio Di Battista</i> | 584 |

| | |
|---|------------|
| Children, parents, grandparents: a look on changing relationships | 590 |
| Changes in social relationships of Italian older people. Evidence from FSS and SHARE Corona surveys <i>Elvira Pelle, Giulia Rivellini and Susanna Zaccarin</i> | 591 |
| Internet use and contacts with children among older Europeans <i>Bruno Arpino</i> | 600 |
| A time-based comparative approach to study the changing demography of grandparenthood in Italy <i>***Elisa Cisotto, Eleonora Meli and Giulia Cavrini</i> | 607 |
| Carry that weight: Parental separation and children's Body Mass Index from childhood to young adulthood <i>Marco Tosi</i> | 616 |
| Living conditions, well-being and poverty | 622 |
| Analyzing the impact of COVID-19 pandemic on elderly population well-being <i>Gloria Polinesi, Mariateresa Ciommi and Chiara Gigliarano</i> | 623 |
| Exploring sustainable food purchasing behaviour using Italian scanner data <i>Ilaria Benedetti, Alessandro Brunetti, Federico Crescenzi and Luigi Palumbo</i> | 629 |
| The evaluation of heat vulnerability in Friuli-Venezia Giulia <i>Laura Pagani, Maria Chiara Zanarotti and Anja Habus</i> | 635 |
| Data Science for Functional and Complex Data | 641 |
| A parsimonious approach to representing functional <i>Enea G. Bongiorno and Aldo Goia</i> | 642 |
| Mixed-effects high-dimensional multivariate regression via group-lasso regularization <i>Francesca Ieva, Andrea Cappozzo, and Giovanni Fiorito</i> | 648 |
| The integration of immigrants in Italy: a multidimensional perspective | 654 |
| Albanian, Romanian and Italian women's fertility intentions: a comparative perspective among migrants, stayers and natives <i>Thaís García-Pereiro and Anna Paterno</i> | 655 |
| Does self-employment in the origin-country affect self-employment after migration? Evidence from Italy and Spain <i>Floriane Bolazzi and Ivana Fellini</i> | 662 |
| The impact of integration on immigrants' health behaviours in Italy <i>Giovanni Minchio, Raffaella Rusciani and Teresa Spadea</i> | 675 |
| Migration, gender, and the distribution of paid and unpaid labour. Preliminary perspectives on foreign couples in Italy <i>Rocco Molinari, Agnese Vitali and Ester Gallo</i> | 687 |

| | |
|--|------------|
| Sampling techniques for big data analysis | 695 |
| Non-probability samples and big data: how to use them? <i>Pier Luigi Conti</i> | 696 |
| Combining Big Data with probability survey data: a comparison of methodologies for estimation from non-probability surveys <i>Maria del Mar Rueda, Ramn Ferri-Garcia and Luis Castro-Martin</i> | 707 |
| A Bayesian approach for combining probability and non-probability samples surveys <i>Camilla Salvatore, Silvia Biffignandi, Joseph Sakshaug, Bella Struminskaya and Arkadiusz Wisniowski</i> | 717 |
| Big data and Official Statistics: some evidences <i>Paolo Righi, Natalia Golini and Gianpiero Bianchi</i> | 723 |
| The analysis of students performance and behaviour based on large databases | 735 |
| Students enrolled in STEM discipline in Italy: patterns of retention, dropout and switch <i>Valentina Tocchioni, Carla Galluccio, Maria Francesca Morabito and Alessandra Petrucci</i> | 736 |
| The routes of Southern Italy University students: an explorative analysis <i>Gabriele Ruiu and Vincenzo Giuseppe Genova</i> | 747 |
| A new bipartite matching approach for record linkage: the case of two big Italian databases <i>Martina Vittorietti, Andrea Priulla, Vincenzo Giuseppe Genova, Giovanni Boscaino and Ornella Giambalvo</i> | 754 |
| Statistical Methods for Science Mapping | 761 |
| A word embedding strategy to study the thematic evolution of ageing and healthcare expenditure growth literature <i>Milena Lopreite, Michelangelo Misuraca and Michelangelo Puliga</i> | 762 |
| An automatic approach for bibliographical co-words networks labelling <i>Manuel J. Cobo and Maria Spano</i> | 773 |
| Characterising research areas in the field of AI <i>Alessandra Belfiore, Angelo Salatino and Francesco Osborne</i> | 780 |
| Mapping evolutionary paths of a society: the longitudinal analysis of the Italian Economia Aziendale <i>Corrado Cuccurullo, Luca D'Aniello and Michele Pizzo</i> | 786 |
| Modelling complex structures in ecological data | 793 |
| New insights on the ecology and conservation of Mediterranean sharks through the development of Citizen Science networks and new modeling approaches <i>Stefano Moro, Francesco Ferretti, Francesco Colloca</i> | 794 |

| | |
|--|------------|
| An overdispersed Poisson model for forest fires occurrences in Southern Italian municipalities <i>Crescenza Calcuttoli and Serena Arima</i> | 798 |
| Assessment of the impact of anthropic pressures on the Giglio island meadow of <i>Posidonia oceanica</i> <i>Gianluca Mastrantonio, Daniele Ventura, Gianluca Mancini and Giandomenico Arlizzone</i> | 804 |
| Accounting for observation processes in spatio-temporal ecological data <i>Janine Illian</i> | 811 |
| Statistics and indicators for the recovery and resilience plan | 815 |
| The prominence of statistical information for the monitoring and effective implementation of the NRRP <i>Andrea Petrella</i> | 816 |
| Big Data Analytics in mobile cellular networks as enabler for innovative statistics to evaluate the effects of Recovery and Resilience Plan actions <i>Andrea Zaramella, Dario Di Sorte, Denis Cappellari and Bruno Zamengo</i> | 819 |
| Measuring the digital transition within the PA: proposals comparison <i>Susanna Traversa and Enrico Ivaldi</i> | 823 |
| Guest Session - European Network for Business and Industrial Statistics (ENBIS) | 828 |
| Interpretability in functional clustering with an application to resistance spot welding process in the automotive industry <i>Christian Capezza, Fabio Centofanti, Antonio Lepore and Biagio Palumbo</i> | 829 |
| Statistical process monitoring of thermal images in additive manufacturing: a nonparametric solution for in-situ monitoring <i>Panagiotis Tsiamirtzis, Marco Luigi Giuseppe Grasso and Bianca Maria Colosimo</i> | 835 |
| Guest Session - International Biometric Society (IBS) - Italian region | 837 |
| Multiple arrows in the Bayesian quiver: Bayesian learning of partially directed structures from heterogeneous data <i>Luca La Rocca, Federico Castelletti, Stefano Peluso, Francesco Claudio Stingo and Guido Consonni</i> | 838 |

| | |
|---|------------|
| 4 Contributed Sessions | 844 |
| Applications in Machine Learning | 845 |
| A neural network approach to survival analysis with time-dependent covariates for modelling time to cardiovascular diseases in HIV patients <i>Federica Corso, Agostino Lurani Cernuschi, Laura Galli, Chiara Masci, Camilla Muccini, Anna Maria Paganoni and Francesca Ieva</i> | 846 |
| Analyzing the Correlation Structure of Financial Markets Using a Quantile Graphical Model <i>Beatrice Foroni, Luca Merlo and Lea Petrella</i> | 852 |
| Neural Network for statistical process control of a multiple stream binomial process with an application to HVAC systems in passenger rail vehicles <i>Gianluca Sposito, Antonio Lepore, Biagio Palumbo and Giuseppe Giannini</i> | 858 |
| Sparse signal extraction via variational SVM <i>Cristian Castiglione and Mauro Bernardi</i> | 864 |
| Bayesian modelling and inference 1 | 870 |
| Bayesian Inference for the Multinomial Probit Model under Gaussian Prior Distribution <i>Augusto Fasano, Giovanni Rebaudo and Niccolò Anceschi</i> | 871 |
| Mapping Indicators on the Unit Interval: the tipsae Shiny App <i>Silvia De Nicolò and Aldo Gardini</i> | 877 |
| A Bayesian spatio-temporal model of PM10 pollutant in the Po Valley <i>Matteo Gianella, Alessandra Guglielmi and Giovanni Lonati</i> | 883 |
| Construction of a proper prior for a Bayesian envelope model <i>Andrea Mascaretti</i> | 889 |
| Hilbert principal component regression for bimodal bounded responses <i>Enea G. Bongiorno, Agnese M. Di Brisco, Aldo Gola, and Sonia Migliorati</i> | 895 |
| Methods of causal inference | 901 |
| Bayesian causal mediation analysis through linear mixed-effect models <i>Chiara Di Maria, Antonino Abbruzzo and Gianfranco Lovison</i> | 902 |
| Bootstrap-aggregated adjustment set selection <i>Lorenzo Giannini</i> | 908 |
| Exploiting partial knowledge to evaluate the average causal effect via an ABC perspective <i>Giulia Cereda, Fabio Corradi and Cecilia Viscardi</i> | 914 |

| | |
|--|------------|
| Intertemporal propensity score matching for casual inference: an application to covid-19 lockdowns and air pollution in Northern Italy | 920 |
| <i>Daniele Bondonio and Paolo Chirico</i> | |
| Methods for Spatio-temporal data | 926 |
| Local Spatio-Temporal Log-Gaussian Cox Processes for seismic data analysis | 927 |
| <i>Nicoletta D'Angelo, Giada Adelfio, and Jorge Mateu</i> | |
| Spatial explorative analysis of thyroid cancer in Sicilian volcanic areas | 933 |
| <i>Francesca Bitonti and Angelo Mazza</i> | |
| Using geo-spatial topic modelling to understand the public view of Italian Twitter users: a climate change application | 939 |
| <i>Yuri Calleo and Francesco Pilla</i> | |
| Comparing local structures of spatio-temporal point processes on linear networks | 945 |
| <i>Nicoletta D'Angelo, Giada Adelfio, and Jorge Mateu</i> | |
| DISTATIS-based spatio-temporal clustering approach: an application to business cycles' time series | 951 |
| <i>Raffaele Mattera and Germana Scepi</i> | |
| Developments in composite indicators | 957 |
| Bayesian Networks for monitoring the gender gap | 958 |
| <i>Flaminia Musella, Lorenzo Giammei, Silvana Romio, Fulvia Mecatti and Paola Vicard</i> | |
| An Alternative Aggregation Function for the UNDP Human Development Index | 964 |
| <i>Manuela Scioni and Paola Annoni</i> | |
| An ultrametric model for building a composite indicator system to study climate change in European countries | 970 |
| <i>Giorgia Zaccaria and Pasquale Sarnacchiaro</i> | |
| Functional Weighted Malmquist Productive Index: a proposal for a dynamic composite indicator | 975 |
| <i>Annalina Sarra, Eugenia Nissi and Tonio Di Battista</i> | |
| CFA & PLS-PM for UX-AI Product infused | 981 |
| <i>Emma Zavarrone and Rosanna Cataldo</i> | |
| Fertility, adulthood, and economic uncertainty | 987 |
| Uncertainty and fertility intentions: a comparison between the Great Recession and the Covid-19 crisis | 988 |
| <i>Chiara Ludovica Comolli</i> | |
| Interpreting the relationship between life course trajectories and explanatory factors. An example on the transition to adulthood | 996 |
| <i>Danilo Bolano, Matthias Studer and Reto Buergin</i> | |

| | |
|--|-------------|
| The relationship between economic news and fertility: the case of Germany <i>Maria Francesca Morabito, Raffaele Guetto, Matthias Vollbracht and Daniele Vignoli</i> | 1002 |
| Leaving home among Millennials in Italy: does economic uncertainty matter? <i>Silvia Meggiolaro and Fausta Ongaro</i> | 1008 |
| Adverse pregnancy outcomes in The United Kingdom following unexpected job loss <i>Alessandro Di Nallo and Selin Koksal</i> | 1014 |
| Bayesian modelling and inference 2 | 1020 |
| A Bayesian beta linear model to analyze fuzzy rating responses <i>Antonio Calcagni, Massimiliano Pastore, Gianmarco Altoe and Livio Finos</i> | 1021 |
| A Mixture Model for Multi-Source Cyber-Vulnerability Assessment <i>Mario Angelelli, Serena Arima and Christian Catalano</i> | 1028 |
| Hierarchical Bayesian models for analysing fish biomass data <i>Rita Fici, Antonino Abbruzzo, Luigi Augugliaro and Giacomo Milisenda</i> | 1034 |
| Insights into the derivative-based method for nonlinear mediation models <i>Claudio Rubino and Chiara Di Maria</i> | 1040 |
| An exploration of Approximate Bayesian Computation (ABC) and dissimilarities <i>Laura Bondi, Marco Bonetti and Raffaella Piccarreta</i> | 1046 |
| Advances in Categorical and Preference data | 1052 |
| On the predictability of a class of ordinal data models <i>Rosaria Simone and Domenico Piccolo</i> | 1053 |
| Multivariate analysis of binary ordinal data using graphical models <i>Camilla Caroni, Fabio Alberto Comazzi, Andrea Deretti and Federico Castelletti</i> | 1059 |
| Multinomial Thompson Sampling for adaptive experiments with rating scales <i>Nina Deliu</i> | 1065 |
| Ranking extraction in nested partially ordered data systems <i>Marco Fattore, Barbara Cavalletti, Matteo Corsi and Alessandro Avellone</i> | 1071 |
| Towards the definition of distance measures in the preference-approval structures <i>Alessandro Albano, Mariangela Sciandra and Antonella Plaia</i> | 1077 |
| Covid-19 Assessment and Evaluation 1 | 1083 |
| Covid-19 impact assessment and inequality decomposition methods <i>Federico Attili and Michele Costa</i> | 1084 |

| | |
|--|-------------|
| Multiversal methods for model selection: COVID-19 vaccine coverage and relative risk reduction <i>Venera Tomaselli and Giulio Giacomo Cantone</i> | 1090 |
| Efficiency and feasibility of two stage sampling designs for estimating SARS-CoV-2 epidemic <i>Pietro Demetrio Falorsi, Vincenzo Nardelli and Giuseppe Arbia</i> | 1096 |
| Evaluating the impacts of Covid-19 on the overall Italian death process via Functional Data Analysis <i>Riccardo Scimone, Alessandra Menafoglio, Laura M. Sangalli and Piercesare Secchi</i> | 1102 |
| Developing countries, migration and migrants | 1107 |
| Domestic violence in Africa: a glance through the DHS survey <i>Micaela Arcaio, Daria Mendola and Anna Maria Parroco</i> | 1108 |
| Inequalities in undernutrition among Roma and non-Roma children in Western Balkans: an analysis of the determinants <i>Annalisa Busetta, Valeria Cetorelli and Chiara Puglisi</i> | 1114 |
| The manual, communicative and quantitative abilities of native and foreign workers according to their level of education in Italy <i>Camilla Pangallo, Oliviero Casacchia and Corrado Polli</i> | 1120 |
| HIV Prevalence in some African Territories: Socio-Economic Drivers <i>Micaela Arcaio, Daria Mendola and Anna Maria Parroco</i> | 1126 |
| A longitudinal cross country comparison of migrant integration policies via Mixture of Matrix-Normals <i>Leonardo Salvatore Alaimo, Francesco Amato and Emiliano Serì</i> | 1132 |
| Education and job placement | 1138 |
| Measuring happiness at work with categorical Principal Component Analysis <i>Ulpiana Kocollari, Maddalena Cavicchioli and Fabio Demaria</i> | 1139 |
| Early and accurate: a Machine Learning approach to predict students' final outcome with registry data <i>Lidia Rossi, Marta Cannistrà and Tommaso Agasisti</i> | 1146 |
| Students' experience with distance learning during Covid 19 pandemic in Southern Italy <i>Angela Maria D'Ugento and Nunziata Ribecco</i> | 1153 |
| Time series methods and Applications | 1159 |
| Trend and cycle decomposition in nonlinear time series <i>Maddalena Cavicchioli</i> | 1160 |
| Asymptotic properties of the SETAR parameters: a new approach <i>Marcella Niglio and Guy Mélard</i> | 1166 |
| Food prices forecast using post-sampled crowdsourced data with Reg-ARMA model: the case of Nigeria <i>Ilaria Lucrezia Amerise, Gloria Solano Hermosilla, Vincenzo Nardelli and Giuseppe Arbia</i> | 1172 |

| | |
|---|-------------|
| Universal change point testing for dependent data <i>Federica Spoto, Alessia Caponera and Pierpaolo Brutti</i> | 1178 |
| Change point detection in fruit bioimpedance using a three-way panel model <i>F. Marta L. Di Lascio and Selene Perazzini</i> | 1184 |
| Bayesian modelling and inference 3 | 1190 |
| A dynamic power prior approach to non-inferiority trials for normal means with unknown variance <i>Francesco Mariani, Fulvio De Santis and Stefania Gubbiotti</i> | 1191 |
| Bayesian Change-Point Detection for a Brownian Motion with a Total Miss Criterion <i>Bruno Buonaguidi</i> | 1197 |
| On the comparison of alternative Bayesian measures of posterior discrepancy <i>Fulvio De Santis and Stefania Gubbiotti</i> | 1203 |
| A Bayesian Test for the comparison of two independent populations <i>Mara Manca, Silvia Columbu and Monica Musio</i> | 1209 |
| A contribution to the L. J. Savage problem <i>Francesco Bertolino, Silvia Columbu and Mara Manca</i> | 1215 |
| Methods for Complex Data | 1221 |
| Optimization of delayed rejection adaptive metropolis <i>Daniele Raffo and Antonietta Mira</i> | 1222 |
| Dealing with multicollinearity and outliers in multinomial logit model: a simulation study <i>Ida Camminatiello and Antonio Lucadamo</i> | 1228 |
| A tool to validate the assumptions on ratios of nearest neighbors' distances: the Consecutive Ratio Paths <i>Francesco Denti and Antonietta Mira</i> | 1233 |
| Dimensionality reduction and visualization for interval-valued data via midpoints-ranges principal component analysis <i>Viviana Schisa, Alfonso Iodice D'Enza and Francesco Palumbo</i> | 1239 |
| Data-driven design-based mapping of forest resources <i>Sara Franceschi, Rosa Maria Di Biase, Lorenzo Fattorini, Marzia Marcheselli and Caterina Pisani</i> | 1245 |
| Environmental data and Climate change | 1252 |
| Ensemble model output statistics for temperature forecasts in Veneto <i>Gaetan Carlo, Giummole Federica, Mameli Valentina and Siad Si Mokrane</i> | 1253 |
| State of the urban Environment in Italy. A comparative analysis of selected composite indicators <i>Giuseppe Lecardane</i> | 1259 |

| | |
|---|-------------|
| A Functional Data Analysis approach for Climate Model Selection: the case study of Campania Region <i>Veronica Villani, Elvira Romano and Paola Mercogliano</i> | 1266 |
| Evolution of scientific literature on climate change: a bibliometric analysis <i>Gianpaolo Zammarchi, Giulia Contu, Maurizio Romano</i> | 1273 |
| Energy and material demand of the Italian Regions <i>Flora Fullone, Giulia Iorio, Assunta Lisa Carulli</i> | 1279 |
| Health and survivorship | 1285 |
| Increasing Inequalities in Mortality by Socioeconomic Position in Italy <i>Chiara Ardito, Nicolás Zengarini, Roberto Leombruni, Angelo d'Errico and Giuseppe Costa</i> | 1286 |
| The role of health conditions in the relationship between socio- economic status and well-being: the counterfactual approach in mediation models <i>Sara Manzella and Margherita Silan</i> | 1296 |
| Excess economic burden of multimorbidity: a population-based study in Italy <i>Chiara Seghieri, Niccolò Borri, Gaia Bertarelli and Sabina Nuti</i> | 1302 |
| Depression-free life expectancy among 50 and older Americans by gender, race/ethnicity and education: the effect of marital disruption <i>Alessandro Feraldi and Cristina Giudici</i> | 1308 |
| Disability-free grandparenthood in Italy. Trends and gender differences <i>Margherita Moretti, Elisa Cisotto and Alessandra De Rose</i> | 1314 |
| Advances in regression models | 1320 |
| Semiparametric M-quantile regression for modelling georeferenced housing price data <i>Riccardo Borgoni, Antonella Carcagni, Alessandra Michelangeli, Nicola Salvati and Francesco Schirripa Spagnolo</i> | 1321 |
| Resampling-based inference for high-dimensional regression <i>Anna Vesel, Jelle J. Goeman, Angela Andreella and Livio Finos</i> | 1327 |
| Quantile regression coefficient modeling for counts to evaluate the productivity of university students <i>Viviana Carcaiso and Leonardo Grilli</i> | 1333 |
| Adaptive smoothing spline using non-convex penalties <i>Daniele Cuntreza and Vito M.R. Muggeo</i> | 1339 |
| Conditional tests for generalized linear models <i>Riccardo De Santis, Jelle J. Goeman, Anna Vesely and Livio Finos</i> | 1345 |

| | |
|--|-------------|
| Methods and applications in economics and finance | 1351 |
| Mixed models for anomaly detection in anti-money laundering aggregate reports <i>Stefano Iezzi and Marianna Siino</i> | 1352 |
| On the drivers of Greenwashing risk: evidence from Eurostoxx600 <i>Yana Kostiuk, Costanza Bosone and Paola Cerchiello</i> | 1358 |
| Modelling Financial Returns with Finite Mixtures of GED <i>Pierdomenico Dutillo and Stefano Antonio Gattone</i> | 1364 |
| Risk Parity strategy for portfolio construction: a kurtosis-based approach <i>Maria Debora Braga, Consuelo Rubina Nava and Maria Grazia Zoia</i> | 1370 |
| Fully reconciled probabilistic GDP forecasts from Income and Expenditure sides <i>Tommaso Di Fonzo and Daniele Girolimetto</i> | 1376 |
| Latent Class models | 1382 |
| Latent thresholds model in classification tasks <i>Giuseppe Mignemi, Andrea Spoto and Antonio Calcagni</i> | 1383 |
| Adaptive filters for time-varying correlation parameters <i>Michele Lambardi di San Miniato, Ruggero Bellio, Luca Grassetti and Paolo Vidoni</i> | 1389 |
| Bayesian structural learning for Latent Class Model with an application to Record Linkage <i>Davide Di Cecco</i> | 1395 |
| Multilevel Latent Class modelling to advise students in self-learning platforms: an application in the context of learning Statistics <i>Roberto Fabbriatore, Zsuzsa Bakk, Roberto Di Mari, Mark de Rooij and Francesco Palumbo</i> | 1401 |
| Latent Markov models with associated mixed responses <i>Alfonso Russo and Alessio Farcomeni</i> | 1407 |
| Methods for health studies | 1413 |
| Beyond the fragility index <i>Piero Quatto and Enrico Ripamonti</i> | 1414 |
| Evaluation of the diagnostic-therapeutic paths for schizophrenic patients through state sequences analysis <i>Laura Savaré, Giovanni Corrao and Francesca Ieva</i> | 1419 |
| Optimal timing of bone-marrow transplant in myelodysplastic syndromes through multi-state modeling and microsimulation <i>Caterina Gregorio, Marta Spreafico and Francesca Ieva</i> | 1425 |
| A fully Bayesian approach for sample size determination of Poisson clinical trials <i>Susanna Gentile and Valeria Sambucini</i> | 1431 |

| | |
|--|-------------|
| Compartmental models in epidemiology: Application on Smoking Habits in Tuscany <i>Alessio Lachi, Cecilia Viscardi, Maria Chiara Malevolti, Giulia Carreras and Michela Baccini</i> | 1437 |
| Covid-19 Assessment and Evaluation 2 | 1443 |
| We are in the same storm but not in the same boat: Impact of COVID-19 on UK households <i>Demetrio Panarello and Giorgio Tassinari</i> | 1444 |
| A network approach to investigate learning experiences and social support in higher education <i>Ilaria Primerano, Maria Carmela Catone, Giuseppe Giordano, Maria Prosperina Vitale</i> | 1450 |
| Physical and cultural activity, internet use and anxiety of Italian university students during the pandemic <i>Giovanni Busetta, Maria Gabriella Campolo and Demetrio Panarello</i> | 1456 |
| The digital divide in Italy before and during the pandemic phase <i>Laura Zannella</i> | 1462 |
| Covid-19 and financial professional advice <i>Marianna Brunetti and Rocco Ciciretti</i> | 1468 |
| Bayesian modelling and inference 4 | 1472 |
| Bayesian functional mixed effects model for sports data <i>Patric Dolmeta, Raffaele Argiento and Silvia Montagna</i> | 1473 |
| Bayesian Optimization with Machine Learning for Big Data Applications in the Cloud <i>Bruno Guindani, Danilo Ardagna and Alessandra Guglielmi</i> | 1479 |
| Confidence distributions and fusion inference for intractable likelihoods <i>Elena Bortolato and Laura Ventura</i> | 1485 |
| Wasserstein distance and applications to Bayesian nonparametrics <i>Marta Catalano, Hugo Lavenant, Antonio Lijoi and Igor Prunster</i> | 1491 |
| Network Analysis and community detection | 1497 |
| Community detection in networks: a heuristic version of Girvan Newman algorithm <i>Ilaria Bombelli and Lorenzo Di Rocco</i> | 1498 |
| Geographically weighted regression for spatial network data: an application to traffic volumes estimation <i>Andrea Gilardi, Riccardo Borgoni and Jorge Mateu</i> | 1504 |
| Asymmetric Spectral Clustering: a comparison between symmetrizations <i>Cinzia Di Nuzzo and Donatella Vicari</i> | 1510 |
| Community detection of seismic point processes <i>Valeria Policastro, Nicoletta D'Angelo and Giada Adelfio</i> | 1516 |

| | |
|---|-------------|
| An Explorative analysis of Different Distance Metrics to Compare Unweighted Undirected Networks <i>Anna Simonetto, Matteo Ventura and Gianni Gilloli</i> | 1522 |
| Gender, attitudes and family ties | 1528 |
| Parents of a disabled child in Italy: less healthy but more civically engaged <i>Nicoletta Balbo and Danilo Bolano</i> | 1529 |
| Searching the nexus between women's empowerment and female genital cutting (FGC) <i>Patrizia Farina, Liva Ortensi, Thomas Pettinato and Enrico Ripamonti</i> | 1535 |
| Social stratification, gender, and attitudes towards voluntary childlessness in Europe: A double machine learning approach <i>Danilo Bolano and Francesco C. Billari</i> | 1539 |
| Integrating structuralism and diffusionism to explain the new Italian emigration <i>Francesca Bitonti</i> | 1545 |
| On the effects of rooted family ties in business networks: The South of Italy in the 19th century <i>Roberto Rondinelli, Giancarlo Ragozini and Maria Carmela Schisani</i> | 1551 |
| Methods and Applications in Clustering | 1557 |
| A semi-supervised clustering method to extract information from the electronic Word Of Mouth <i>Giulia Contu, Luca Frigau, Maurizio Romano and Marco Ortu</i> | 1558 |
| Spectral approach for clustering three-way data <i>Cinzia Di Nuzzo and Salvatore Ingrassia</i> | 1564 |
| Double clustering with a matrix-variate regression model: finding groups of athletes and disciplines in decathlon's data <i>Mattia Stival, Mauro Bernardi, Manuela Cattelan and Petros Dellaportas</i> | 1570 |
| Classification of the population dynamics <i>Federico Bacchi and Laura Neri</i> | 1576 |
| Locating γ -Ray Sources on the Celestial Sphere via Modal Clustering <i>Anna Montin, Alessandra R. Brazzale and Giovanna Menardi</i> | 1582 |
| Sampling and Official Statistics | 1588 |
| Fisher's Noncentral Hypergeometric Distribution for Population Size Estimation <i>Veronica Ballerini and Brunero Liseo</i> | 1589 |
| Small area models for skew and kurtotic distributions <i>Maria Rosaria Ferrante and Lorenzo Mori</i> | 1595 |

| | |
|--|-------------|
| The use of remotely sensed data in sampling designs for forest monitoring | 1601 |
| <i>Chiara Bocci, Gherardo Chirici, Giovanni D'Amico, Saverio Francini and Emilia Rocco</i> | |
| Analyzing different causes of one-inflation in capture recapture models for criminal populations | 1607 |
| <i>Davide Di Cecco, Andrea Tancredi and Tiziana Tuoto</i> | |
| Administrative database and official statistics: an IT and statistical procedure | 1613 |
| <i>Caterina Marini and Vittorio Nicolardi</i> | |
| Spatial modeling and Analyses | 1619 |
| Spatial statistics analysis using microdata: an application at agricultural sector | 1620 |
| <i>Daniela Fusco, Maria Antonietta Liguori, Valerio Moretti and Francesco Giovanni Truglia</i> | |
| Bayesian spatial modeling of extreme precipitation | 1627 |
| <i>Federica Stolf</i> | |
| A proposal to adjust local Moran's I for measuring residential segregation | 1632 |
| <i>Antonio De Falco and Antonio Irpino</i> | |
| Accurate directional inference for gaussian graphical models | 1637 |
| <i>Claudia Di Caterina, Nancy Reid and Nicola Sartori</i> | |
| Advances in Classification | 1643 |
| Measures of interrater agreement based on the standard deviation | 1644 |
| <i>Giuseppe Bove</i> | |
| A Comparison of accuracy measures for Classification tasks | 1650 |
| <i>Amalia Vanacore and Maria Sole Pellegrino</i> | |
| Iterative Threshold-based Naive Bayes Classifier: an efficient Tb-NB improvement | 1656 |
| <i>Maurizio Romano, Gianpaolo Zammarchi and Giulia Contu</i> | |
| Reprogramming FairGANs with Variational Auto-Encoders: A New Transfer Learning Model | 1662 |
| <i>Beatrice Nobile, Gabriele Santin, Bruno Lepri and Pierpaolo Brutti</i> | |
| Robust statistics | 1669 |
| Combinatorial Analysis of Factorial Designs with Ordered Factors | 1670 |
| <i>Roberto Fontana and Fabio Rapallo</i> | |
| Robustifying the Rasch model with the forward search | 1676 |
| <i>Anna Comotti and Francesca Greselin</i> | |
| A novel estimation procedure for robust CP model fitting | 1682 |
| <i>Valentin Todorov, Violetta Simonacci, Michele Gallo and Nikolay Trendafilov</i> | |

| | |
|---|-------------|
| A robust approach for functional ANOVA with application to additive manufacturing <i>Fabio Centofanti, Bianca Maria Colosimo, Marco Luigi Grasso, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini</i> | 1688 |
| Modeling unconditional M-quantiles in a regression framework <i>Luca Merlo, Lea Petrella and Nicola Salvati</i> | 1692 |
| Model-based clustering | 1696 |
| Bayesian mixtures of semi-Markov models <i>Rosario Barone and Andrea Tancredi</i> | 1697 |
| Specification of informative priors for capture-recapture finite mixture models <i>Pierfrancesco Alaimo Di Loro, Gianmarco Caruso, Marco Mingione, Giovanna Jona Lasinio and Luca Tardella</i> | 1703 |
| Clustering multivariate categorical data: a graphical model-based approach <i>Francesco Rettore, Michele Russo, Luca Zerman and Federico Castelletti</i> | 1709 |
| The Gaussian mixture model-based clustering for the comparative analysis of the Healthcare Digitalization Index in the Italian local health authorities <i>Margaret Antonicelli, Michele Rubino and Filomena Maggino</i> | 1715 |
| Student performance evaluation | 1721 |
| Rasch model versus Rasch Mixture model: strengthens and limits in identifying factors affecting students' performance in mathematics <i>Clelia Cascella</i> | 1722 |
| Does taking additional Maths classes improve university performance? <i>Martina Vittorietti, Andrea Priulla and Massimo Attanasio</i> | 1728 |
| University dropout and churn in Italy: an analysis over time <i>Barbara Barbieri, Mariano Porcu, Luisa Salaris, Isabella Sullis, Nicola Tedesco and Cristian Usala</i> | 1734 |
| The ANOVI for detecting the impact of education and employment on income inequality <i>Elena Fabrizi, Alessio Guandalini and Alessandra Spagnoli</i> | 1740 |
| What causes juvenile crime? a case-control study <i>Elena Dalla Chiara and Federico Perali</i> | 1747 |
| Methods and Applications in Survival analysis | 1753 |
| Recursive partitioning for survival data <i>Ambra Macis</i> | 1754 |
| Detecting survival patterns in a digital learning platform <i>Marta Cannistrà, Mara Soncin and Federico Frattini</i> | 1760 |
| An extension of proper Bayesian bootstrap ensemble tree models to survival analysis <i>Elena Ballante</i> | 1766 |

| | |
|---|-------------|
| Modelling time to university dropout by means of time-dependent frailty COX PH models <i>Mirko Giovio, Paola Mussida and Chiara Masci</i> | 1771 |
| Family history in survival and disease development <i>Maria Veronica Vinattieri and Marco Bonetti</i> | 1777 |
| Text mining | 1783 |
| Topics & metaverse: an explorative analysis <i>Emma Zavarrone, Alessia Forciniti, Emanuele Parisi, Maria Gabriella Grassia</i> | 1784 |
| Applying Topic Models to bibliographic search: some results in basketball domain <i>Manlio Migliorati and Eugenio Brentari</i> | 1791 |
| Exploiting Text Mining and Network Analysis for future scenarios development: an application on remote working <i>Yuri Calleo, Simone Di Zio and Vanessa Russo</i> | 1797 |
| Emotion recognition in Italian political language to predict positionings and crises government <i>Alessia Forciniti and Emma Zavarrone</i> | 1803 |
| What does your self-description reveal about you? <i>Riccardo Ricciardi</i> | 1809 |
| Variable selection and complete matrix approaches | 1815 |
| A Statistical Approach for the Completion of Input-Output Tables <i>Rodolfo Metulini, Giorgio Gnecco, Francesco Biancalani and Massimo Riccaboni</i> | 1816 |
| On multivariate records over sequences of random vectors with Marshall-Olkin dependence of components <i>A. Khorrami Chokami and Simone A. Padoan</i> | 1822 |
| The joint censored gaussian graphical lasso model <i>Gianluca Sottile, Luigi Augugliaro and Veronica Vinciotti</i> | 1829 |
| Variable selection with unbiased estimation: the cdf penalty <i>Daniele Cuntrera, Vito M.R. Muggeo and Luigi Augugliaro</i> | 1835 |
| Automatic variable selection for MIDAS regressions: an application <i>Consuelo Rubina Nava, Luigi Riso and Maria Grazia Zoia</i> | 1841 |
| Distribution Theory and Estimation | 1847 |
| A general framework for unit distributions <i>Francesca Condino, Filippo Domma and Bozidar V. Popovic</i> | 1848 |
| Prediction intervals based on multiplicative model combinations <i>Valentina Marneli and Paolo Vidoni</i> | 1854 |
| Some advances on pairwise likelihood estimation in ordinal data latent variable models <i>Giuseppe Alfonzetti and Ruggero Bellio</i> | 1860 |

| | |
|--|-------------|
| Functional Data Analysis | 1866 |
| A new functional clustering method: the Functional Clustering and Dimension Reduction model <i>Adelia Evangelista and Stefano Antonio Gattone</i> | 1867 |
| Nonparametric functional prediction bands: theory with an application to bike sharing mobility demand in the city of Milan <i>Jacopo Diquigiovanni, Matteo Fontana and Simone Vantini</i> | 1873 |
| An R package for the statistical process monitoring of functional data <i>Christian Capezza, Fabio Centofanti, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini</i> | 1878 |
| Trend filtering for functional regression <i>Federico Ferraccioli, Alessandro Casa and Marco Stefanucci</i> | 1884 |
| Conformal prediction for spatio-functional regression models <i>Diana, Romano, Irpino</i> | 1890 |
| Tourism and sport studies | 1895 |
| Assessing satisfaction of tourists visiting Italian museums: evidence from the eWOM <i>Daria Mendola and Valentina Oddo</i> | 1896 |
| COVID-19 pandemic and tourism demand: a comparison between Spain and Italy <i>Caterina Sciortino, Ludovica Venturella and Stefano De Cantis</i> | 1902 |
| A compositional analysis of tourism in Europe <i>Francesco Porro</i> | 1908 |
| Improving administrative data quality on tourism using Big Data <i>Antonella Bianchino, Armando d'Aniello and Daniela Fusco</i> | 1914 |
| Geographical variations of socio-demographic issues | 1920 |
| Elderly HCE and health care need: comparing spatially unexplained levels <i>Irene Torrini, Laura Rizzi and Luca Grasseti</i> | 1921 |
| Measuring sustainable development at the regional level. The case of Italy <i>Marianna Bartiromo and Enrico Ivaldi</i> | 1927 |
| Socio-economic deprivation and COVID-19 infection: a Bayesian spatial modelling approach <i>Antonino Abbruzzo, Andrea Mattaliano, Alessandro Arrigo, Salvatore Scondotto and Mauro Ferrante</i> | 1933 |
| Applications in Economics | 1939 |
| The measurement of economic security through relative indicators <i>Alessandro Gallo, Silvia Pacei and Maria Rosaria Ferrante</i> | 1940 |

| | |
|--|-------------|
| A regional analysis of the efficiency by energy's producers in Italy <i>Gianna Greca, Giuseppe Cinquegrana and Giovanni Fosco</i> | 1946 |
| On investigating social and financial aspects of Cardano <i>Stefano Vacca, Marco Ortu, Gianpaolo Zammarchi and Giuseppe Destefanis</i> | 1953 |
| Combined permutation test on the effect of age of micro enterprises on the propensity to Circular Economy <i>Stefano Bonnini and Michela Borghesi</i> | 1959 |
| Comparison of Two Different Approaches to Measure Economic Access to Food and Insecurity: an Application to Mexican data <i>Stefano Marchetti, Luca Secondi and Adrian Vargas-Lopez</i> | 1965 |
| Image analysis and visual methods | 1971 |
| Bias correction of the maximum likelihood estimator for Emax model at the interim analysis <i>Caterina May and Chiara Tommasi</i> | 1972 |
| Visual and automated methods in digital microscopy to evaluate fungal colonisation on plant roots <i>Ivan Sciascia, Andrea Crosino and Andrea Genre</i> | 1977 |
| From satellite images to road pavement type: an object-oriented classification approach <i>Arianna Burzacchi, Matteo Landrò and Simone Vantini</i> | 1983 |
| Valid inference for group analysis of functionally aligned fMRI images <i>Angela Andreella, Riccardo De Santis and Livio Finos</i> | 1987 |
| Topological persistence for astronomical image segmentation <i>Riccardo Ceccaroni, Pierpaolo Brutti, Marco Castellano, Adriano Fontana and Emiliano Merlin</i> | 1993 |
| Statistical assessment and empirical estimation | 1999 |
| Confidence regions for optimal sensitivity and specificity of a diagnostic test <i>Gianfranco Adimari, Duc-Khanh To and Monica Chiogna</i> | 2000 |
| On the sensitiveness to the memory parameter in the network of tennis <i>Alberto Arcagni, Vincenzo Candila and Rosanna Grassi</i> | 2006 |
| Two-part model with measurement error <i>Maria Felice Arezzo, Serena Arima, and Giuseppina Guagnano</i> | 2011 |
| Statistical assessment of practical significance <i>Andrea Ongaro, Sonia Migliorati, and Enrico Ripamont</i> | 2017 |
| Autoregressive and mixed effects models | 2023 |
| Asymptotic Properties of the Nonlinear Least Squares Estimator in HE-HAR Models <i>Emilija Dzuverovic and Edoardo Otranto</i> | 2024 |

| | |
|--|-------------|
| A note on testing for threshold non-linearity in presence of heteroskedasticity in time series <i>Simone Giannerini and Greta Goracci</i> | 2030 |
| The conditional autoregressive Whart-G model <i>Massimiliano Caporin and Marco Girardi</i> | 2036 |
| Semi-parametric generalized linear mixed effects models for binary response for the analysis of heart failure hospitalizations <i>Alessandra Ragni, Chiara Masci, Francesca Ieva and Anna Maria Paganoni</i> | 2042 |
| Issues in Data science | 2048 |
| etree: Classification and Regression With Structured and Mixed-Type Data in R <i>Riccardo Giubilei, Tullia Padellini and Pierpaolo Brutti</i> | 2049 |
| Deep Learning framework for ungrouping coarsely aggregated vital rates <i>Andrea Nigri</i> | 2055 |
| Inside the metaverse: analysis of the state of the art and development of a new usage approach based on quality and ethics <i>Vito Santarcangelo, Emilio Massa, Saverio Gianluca Crisafulli, Antonio Ruoto, Angelo Lamacchia, Alessandro D'Alcantara, Alessandro Verderame and Massimiliano Giacalone</i> | 2061 |

Does taking additional Maths classes improve university performance?

A clustered case-control study based on Italian high-school and university data

Martina Vittorietti, Andrea Priulla, Massimo Attanasio

Abstract Several recent studies in educational literature showed how students' skills in maths affect their success at higher levels of education. The aim of this paper is to evaluate the effect of taking additional maths class at high school on first-year performance of Italian university students. However, university performance and the choice of the high-school depend on several factors that make this evaluation challenging. Using information coming from three different sources, we carry out a multilevel propensity score procedure to estimate the average treatment effect between the applied sciences track and the traditional scientific one. After balancing for school- and student-level covariates, the results of a logistic regression model suggest no difference between the two school tracks.

Abstract Studi recenti nell'ambito dell'istruzione secondaria hanno mostrato come le abilità matematiche degli studenti abbiano un effetto sul loro futuro successo scolastico e universitario. L'obiettivo di questo lavoro è valutare se studiare più matematica a scuola è associato ad una migliore performance al primo anno di università. La performance universitaria e la scelta della scuola dipendono tuttavia da molteplici fattori che rendono complessa questa analisi. Utilizzando dati provenienti da tre diverse fonti, applichiamo un multilevel propensity score matching per stimare l'effetto medio di due trattamenti: "il liceo scientifico delle scienze applicate" e il "liceo scientifico tradizionale". Dopo aver bilanciato le caratteristiche della scuola e degli studenti, i risultati di un modello logistico mostrano che non emergono differenze tra i due percorsi scolastici.

Key words: propensity score, fine balance, educational data, multilevel propensity score, optimal network flow

Martina Vittorietti, Andrea Priulla, Massimo Attanasio
Department of Economics, Business and Statistics, University of Palermo, Italy, e-mail: martina.vittorietti@unipa.it, andrea.priulla@unipa.it, massimo.attanasio@unipa.it

1 Introduction

Mathematics is perceived as the foundation for scientific and technological knowledge that is cherished by societies worldwide. Due to its relevance and application, mathematics is widely regarded as one of the most important school subjects and a central major of the school curriculum [8]. Mathematics underpins the study of many disciplines at university, not only in science, technology, engineering and mathematics (STEM) but also in agriculture, pharmacy and economics [10]. In Italy, there is at least one mathematics or mathematics related course in most of the degree courses, and it is often regarded as one of the toughest exams. Several surveys at both international level, such as the PISA tests promoted by OECD, and at national level, such as INVALSI tests, are conducted for investigating mathematical skills in high school. Italy is characterized by a high school system organized in separate tracks, which sorts students according to their academic or vocational orientation. Despite the existence of different tracks, the high school could be described as tripartite, with a 5-year academic oriented generalist education provided by lyceums, with further distinctions in humanities, science activities, languages, pedagogical sciences, 5-year technical schools, and 5-year vocational schools. In 2010, Italy underwent a reform of the secondary school education. One of the main innovations was the introduction of a new school curriculum: the applied sciences curriculum. The main difference with respect to the traditional scientific track was the substitution of the Latin language class with classes in scientific subjects such as mathematics, physics, natural sciences, and the addition of computer science as a new subject. In [12] the author compares the in-going and out-going mathematical skills of the students of the two different curricula, underlying the strengths and the weaknesses of both tracks, but as far as the authors' knowledge there are no results about the academic performances related to this new high school track. Therefore, the main objective of this paper is to compare the academic performances of the students coming from traditional scientific or applied sciences tracks, in order to answer the following question: "Does taking additional maths classes improve the university performance?" Before answering this question, it should be noted that university performance lies above an incredibly intricate net of multi-dimensional factors. Therefore, the question in the present form is ill-posed because of social and cultural factors that affect the choice of studying more mathematics or more correctly the choice of the applied sciences track, that eventually will affect university choice and the university performance too [1]. Among the social and cultural factors, the most relevant are: gender [3, 6], social level, location of the high school [2, 11], socio-economic status of the family [4, 16]. For investigating the difference between the two tracks, a balancing for all the possible factors related to both the school track and the academic performance is needed. Analogously to clinical studies, we investigate the effect of the *treatment* "Studying more maths". It is possible to see this study as a clustered case-control study, in which the cases are the students who attended the applied sciences track, and the control are the students who attended the traditional Scientific track. We propose a multi-level balancing approach of the socio-demographic characteristics using a multi-level propen-

Does taking additional Maths classes improve university performance?

sity score model. This procedure aims to eliminating the selection bias due to the fact that, in observational studies, the choice of the high school track, is not random, because it is related to socio-economic and cultural factors[4]. The number of ECTS achieved by the two group of students, treated and controls, in their first academic year is then compared and the average treatment effect (ATE) is estimated.

2 Data

The dataset used in this paper is built by linking three distinct administrative sources: aggregate data coming from the National Archive of Schools (ANS-S); micro-data on high school career from the National Evaluation Institute for the School System (INV);

As most of the educational data, the data at hand presents a hierarchical structure. In fact, it is possible to distinguish variables based on two levels:

1. *school-level*: macro-region, university enrollment rate and the Economic, Social, Cultural Status index (ESCS) used by INVALSI [14];
2. *student-level*: gender, ESCS, high school career's regularity.

3 Methods

Educational data often presents a multilevel structure, so, it is common to speak of clustered treatments. In clustered-treatment observational studies, treatments are applied to entire cluster (schools) or to individuals (students or teachers), or to multiple levels simultaneously (schools and students) [13]. Let S be the number of matched pairs of schools $s = 1, \dots, S$. Each school presents both classes belonging to the applied sciences and the traditional scientific track, $j = 1, 2$, the treatment and the control respectively, for a total of $2S$ units. Each school track of each school s_j contains $n_{s_j} > 1$ students, $i = 1, \dots, n_{s_j}$. Each pair is matched on a vector of observed pre-treatment covariates x_{s_j} . Let X_{s_j} be the matrix whose rows consists of the x_{s_j} vectors for each student i in the j -th track of the s -th school with support $\mathbf{X} \subset \mathbf{R}$. A student i in school track j of school s is described by both observed covariates and possibly unobserved covariates u_{s_j} . In our study, treatment assignment occurs at class level as whole classes are assigned to treatment or to control. If the j -th class in pair s receives the treatment $Z_{s_j} = 1$, whereas if it receives the control $Z_{s_j} = 0$, thus $Z_{s_1} + Z_{s_2} = 1$, for each s as each pair contains the treatment classes and the control classes. Following the designed structure, the statistical analysis strategy uses a propensity score model as a matching strategy [9, 15]. In particular, we follow the optimal multilevel matching using network flows proposed in [13], in which first we match the schools on school-related covariates and then we match students using individual socio-demographic characteristics. After the matching procedure, we fit a logit regression model with random effects on the matched data, to evaluate the

effect of the treatment on the probability of achieving at least 37 ECTS. We chose the cut-point of 37 ECTS because empirical evidences suggest that 37 is the best cut-point predictor of university success

4 Results

In Table 1 we assess the balance among the students covariates before and after the matching procedure.

| | Unmatched | | | | Matched | | | |
|-------------------------------|-----------|---------|-------|---------|---------|---------|-------|---------|
| | Treated | Control | SMD | P-value | Treated | Control | SMD | P-value |
| <i>Male %</i> | 71,0% | 50,0% | 0,44 | 0,00 | 66,9% | 65,0% | 0,04 | 0,17 |
| <i>Student's ESCS</i> | 0,39 | 0,52 | -0,14 | 0,00 | 0,48 | 0,52 | -0,05 | 0,08 |
| <i>HS final mark</i> | 78,87 | 81,56 | -0,22 | 0,00 | 79,41 | 79,88 | -0,04 | 0,16 |
| <i>INVALSI math result</i> | 245,61 | 245,20 | 0,01 | 0,34 | 248,83 | 248,42 | 0,01 | 0,66 |
| <i>INVALSI italian result</i> | 226,04 | 233,24 | -0,22 | 0,00 | 231,05 | 232,26 | -0,04 | 0,18 |
| <i>INVALSI english result</i> | 451,84 | 459,11 | -0,13 | 0,00 | 460,38 | 461,73 | -0,02 | 0,40 |

Table 1 Student covariates comparison in the treatment and control groups before and after the matching.

According to the standardized mean difference (SMD) [5] before the matching procedure, the major imbalance is in terms of gender, while no imbalance is recorded for the INVALSI math results. After the matching procedure all the student covariates are balanced, in fact no SMD is statistically significant.

In Table 2, the results of two logistic models with random effects on the university field of study fitted on unmatched and matched data are reported. The main differences are: before matching, the treatment effect was negative and significant, suggesting that the applied sciences students have worse performance with respect to their peers of the traditional scientific; after matching, the treatment is not anymore effective. Moreover, in Figure 1, it is possible to see how the treatment varies after the matching procedure with respect to the field of study: if on unmatched data the treatment had a positive effect for the STEM courses and negative on no STEM courses, after the matching the effect is reversed.

5 Conclusions

In this paper we compare the university performance of students that attended two different school tracks, to assess the effect of different skills. The main results show that the two groups significantly differ by gender, socio-economic level, final high school grade and scores in English and Italian INVALSI tests, while no difference are observed in maths test scores. After having balanced both school and students

Does taking additional Maths classes improve university performance?

| Variable | Unmatched | | Matched | |
|--|-----------|--------|----------|--------|
| | Estimate | Pvalue | Estimate | Pvalue |
| <i>Intercept</i> | -5,48 | 0,00 | -5,98 | 0,00 |
| <i>HS diploma (ref="Scientific")</i> | -0,11 | 0,05 | -0,04 | 0,56 |
| <i>HS final mark</i> | 0,07 | 0,00 | 0,07 | 0,00 |
| <i>Genere (ref="Female")</i> | -0,14 | 0,00 | -0,22 | 0,01 |
| <i>HS macro-region (ref="North") - Center</i> | -0,61 | 0,00 | -0,59 | 0,00 |
| <i>HS macro-region (ref="North") - South and Islands</i> | -0,98 | 0,00 | -0,92 | 0,00 |
| <i>Student's ESCS</i> | 0,06 | 0,00 | 0,04 | 0,26 |
| <i>INVALSI math result</i> | 0,01 | 0,00 | 0,01 | 0,00 |
| <i>INVALSI italian result</i> | 0,00 | 0,95 | 0,00 | 0,16 |
| <i>INVALSI english result</i> | 0,00 | 0,00 | 0,00 | 0,54 |

Table 2 Parameters of logit regression models with random effects before and after matching.

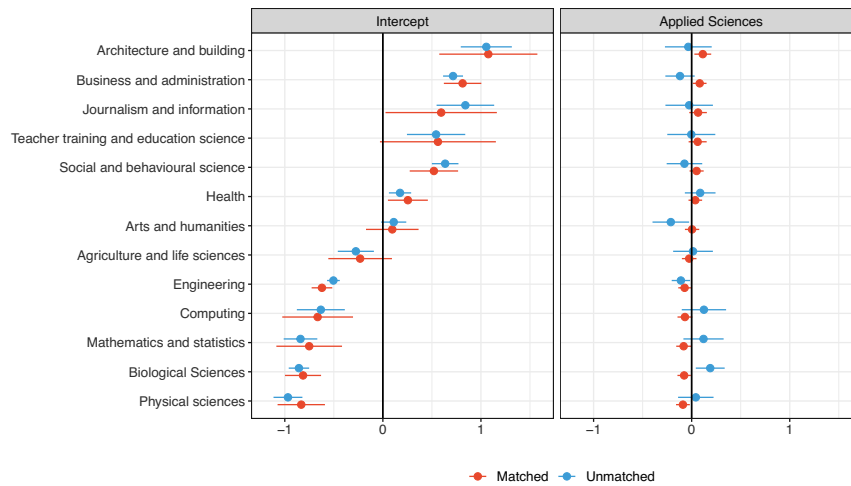


Fig. 1 Comparison of the intercept and the treatment (Applied Sciences) parameters for each university field of study estimated using both the unmatched and the matched data.

covariates, the results of a logit regression model with random effects with respect to the university field of study suggest that there is no difference between the two tracks. However, some considerations are needed. We evaluate the difference among the two different school tracks just with respect to their academic performance. In reality, it already exists a difference in the enrollment rates of the students belonging to the two groups, strongly in favor of the traditional scientific school track. This tunes the matching approach towards a specific direction: it selects the best students of the applied sciences school track. The Italian high school system is still hierarchical and the traditional scientific track is then considered as characterized by a better “social context” compared to the applied science track. Our procedure can not balance for this social context, which is an unobservable variable.

Acknowledgements This paper has been supported from Italian Ministerial grant PRIN 2017 “From high school to job placement: micro-data life course analysis of university student mobility and its impact on the Italian North-South divide.”, n. 2017HBTK5P.

References

1. Aina, Carmen, Massimiliano Bratti, and Enrico Lippo. “Ranking high schools using university student performance in Italy.” *Economia Politica* 38.1 (2021): 293-321.
2. Azzolini, Davide, and Loris Vergolini. “Tracking, inequality and education policy. Looking for a recipe for the Italian case.” *Scuola democratica* 2 (2014): 0-0.
3. Cascella, Clelia, Chiara Giberti, and Giorgio Bolondi. “An analysis of Differential Item Functioning on INVALSI tests, designed to explore gender gap in mathematical tasks.” *Studies in Educational Evaluation* 64 (2020): 100819.
4. Checchi, Daniele, and Luca Flabbi. “Intergenerational mobility and schooling decisions in Germany and Italy: The impact of secondary school tracks.” (2007).
5. Cohen, Jacob. “Statistical power analysis for the social sciences.” Hillsdale, NJ: Lawrence Erlbaum (1988).
6. Contini, Dalit, Di Tommaso, Maria L., and Mendolia, Silvia. “The gender gap in mathematics achievement: Evidence from Italian data.” *Economics of Education Review* 58 (2017): 32-42.
7. D’Amore, Bruno, et al. “La didattica della matematica: strumenti per capire e per intervenire.” *Atti del Convegno Nazionale omonimo* (2014): 3-4.
8. Hagan, John Ekow, et al. “Students’ perception towards mathematics and its effects on academic performance.” *Asian Journal of Education and Social Studies* 8.1 (2020): 8-14.
9. Keele, Luke, Matthew Lenard, and Lindsay Page. “Matching Methods for Clustered Observational Studies in Education.” *Journal of Research on Educational Effectiveness* 14.3 (2021): 696-725.
10. Nicholas, Jackie, et al. “Mathematics preparation for university: entry, pathways and impact on performance in first year science and mathematics subjects.” *International Journal of Innovation in Science and Mathematics Education* 23.1 (2015).
11. Panichella, Nazareno, and Moris Triventi. “Social inequalities in the choice of secondary school: Long-term trends during educational expansion and reforms in Italy.” *European Societies* 16.5 (2014): 666-693.
12. Petolicchio, Annamaria. “La Riforma del Liceo Scientifico: il curriculum Scienze Applicate.” *FORMAZIONE & INSEGNAMENTO. Rivista internazionale di Scienze dell’educazione e della formazione* 14.2 (2016): 187-194.
13. Pimentel, Samuel D., et al. “Optimal multilevel matching using network flows: An application to a summer reading intervention.” *The Annals of Applied Statistics* 12.3 (2018): 1479-1505.
14. Ricci, Roberto. “The Economic, Social, and Cultural Background: a continuous index for the Italian Students of the fifth grade.” *Mimeo* (2010).
15. Rosenbaum, Paul R. “Modern algorithms for matching in observational studies.” *Annual Review of Statistics and Its Application* 7 (2020): 143-176.
16. Xie, Yu, Michael Fang, and Kimberlee Shauman. “STEM education.” *Annual review of sociology* 41 (2015): 331-357.