



UNIVERSITÀ DI PISA



Sant'Anna
Scuola Universitaria Superiore Pisa



Consiglio Nazionale delle Ricerche

Book of Short Papers

SIS 2020



Società
Italiana di
Statistica

Editors: Alessio Pollice, Nicola Salvati and Francesco Schirripa Spagnolo

Copyright © 2020

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891910776

Contents

Specialized sessions

| | |
|--|----|
| Accounting for record linkage errors in inference (S2G-SIS)..... | 2 |
| Probabilistic record linkage with less than three matching variables. | 3 |
| <i>Tiziana Tuoto and Marco Fortini</i> | |
| Advanced methods for measuring and communicating uncertainty in official statistics | 9 |
| A model for measuring the accuracy in spatial price statistics using scanner data. | 10 |
| <i>Ilaria Benedetti and Federico Crescenzi</i> | |
| Communication of Uncertainty of Official Statistics. | 16 |
| <i>Edwin de Jonge and Gian Luigi Mazzi</i> | |
| Measuring uncertainty for infra-annual macroeconomic statistics. | 22 |
| <i>George Kapetanios, Massimiliano Marcellino and Gian Luigi Mazzi</i> | |
| Bayesian methods in biostatistics | 27 |
| Network Estimation of Compositional Data. | 28 |
| <i>Nathan Osborne, Christine B. Peterson and Marina Vannucci</i> | |
| Using co-data to empower genomics-based prediction and variable selection. | 34 |
| <i>Magnus M. Münch, Mirrelijn M. van Nee and Mark A. van de Wiel</i> | |
| Data integration versus privacy protection: a methodological challenge? | 40 |
| Statistical Disclosure Control for Integrated Data. | 41 |
| <i>Natalie Shlomo</i> | |
| The Integrated System of Statistic Registers: first steps towards facing privacy issues. | 47 |
| <i>Mauro Bruno and Roberta Radini</i> | |
| Trusted Smart Surveys: a possible application of Privacy Enhancing Technologies in Official Statistics. | 53 |
| <i>Fabio Ricciato, Kostas Giannakouris, Albrecht Wirthmann and Martina Hahn</i> | |
| Designing adaptive clinical trials | 59 |
| Optimal designs for multi-arm exponential trials. | 60 |
| <i>Rosamarie Frieri and Marco Novelli</i> | |
| Education: students' mobility and labour market..... | 66 |
| From measurement to explanatory approaches: an assessment of the attractiveness of the curricula programs supplied by Italian universities. | 67 |
| <i>Isabella Sullis, Silvia Columbu and Mariano Porcu</i> | |
| Pull factors for university students' mobility: a gravity model approach. | 73 |
| <i>Giovanni Boscaino and Vincenzo Giuseppe Genova</i> | |
| Spatial autoregressive gravity models to explain the university student mobility in Italy. | 79 |
| <i>Silvia Bacci, Bruno Bertaccini and Chiara Bocci</i> | |

| | |
|--|------------|
| Environmental Statistics (GRASPA-SIS) | 85 |
| A Time Clustering Model for Spatio-Temporal Data. <i>Clara Grazian, Gianluca Mastrantonio and Enrico Bibbona</i> | 86 |
| Reconstruction of sparsely sampled functional time series using frequency domain functional principal components. <i>Amira Elayouty, Marian Scott and Claire Miller</i> | 93 |
| Methods for High Dimensional Compositional Data Analysis | 98 |
| Algorithms for compositional tensors of third-order. <i>Violetta Simonacci</i> | 99 |
| High-dimensional regression with compositional covariates: a robust perspective. <i>Gianna Serafina Monti and Peter Filzmoser</i> | 105 |
| Three-way compositional analysis of energy intensity in manufacturing. <i>Valentin Todorov and Violetta Simonacci</i> | 111 |
| Modern Statistics for Physics Discoveries | 117 |
| Identification of high-energy λ -ray sources via nonparametric clustering. <i>Giovanna Menardi, Denise Costantin, and Federico Ferraccioli</i> | 118 |
| Statistical Analysis of Macroseismic Data for a better Evaluation of Earthquakes Attenuation Laws. <i>Marcello Chiodi, Antonino D'Alessandro, Giada Adelfio and Nicoletta D'Angelo</i> | 124 |
| Network Modelling in Biostatistics..... | 130 |
| Natural direct and indirect relative risk for mediation analysis. <i>Monia Lupparelli and Alessandra Mattei</i> | 131 |
| New issues on multivariate and univariate quantile regression | 137 |
| Mixtures of quantile regressions for longitudinal data: an R package. <i>Maria Francesca Marino, Maria Giovanna Ranalli and Marco Alfò</i> | 138 |
| Multivariate Mixed Hidden Markov Model for joint estimation of multiple quantiles. <i>Luca Merlo, Lea Petrella and Nikos Tzavidis</i> | 144 |
| Recent methodological advances in finite mixture modeling with applications (CLADAG-SIS) | 150 |
| Aggregating Gaussian mixture components. <i>Roberto Rocci</i> | 151 |
| Local and overall coefficients of determination for mixtures of generalized linear models. <i>Roberto Di Mari, Salvatore Ingrassia and Antonio Punzo</i> | 157 |
| Statistical Analysis of Satellite Data (SDS-SIS) | 163 |
| Functional Data Analysis for Interferometric Synthetic Aperture Radar Data Post-Processing: The case of Santa Barbara mud volcano. <i>Matteo Fontana, Alessandra Menafoglio, Francesca Cigna and Deodato Tapete</i> | 164 |
| Recent Contributions to the Understanding of the Uncertainty in Upper-Air Reference Measurements. <i>Alessandro Fassò</i> | 170 |
| Statistical models and methods for Business and Industry | 176 |
| Modelling and monitoring of complex 3D shapes: a novel approach for lattice structures. <i>Bianca Maria Colosimo, Marco Grasso and Federica Garghetti</i> | 177 |
| Open data powered territorial planning - Case study: The Turin historical center. <i>Silvia Casagrande, Gianmaria Origi, Alberto Pasanisi, Martina Tamburini, Pascal Terrien, Tania Cerquitelli and Alfonso Capozzoli</i> | 183 |
| Process optimization in Industry 4.0: Are all data analytics models useful? <i>Alberto Ferrer</i> | 189 |

| | |
|--|-----|
| Technology and demographic behaviours (AISP-SIS) | 195 |
| Internet and the Timing of Births. | 196 |
| <i>Maria Sironi, Osea Giuntella and Francesco C. Billari</i> | |
| The Internetization of Marriage: Effects of the Diffusion of High-Speed Internet on Marriage, Divorce, and Assortative Mating. | 202 |
| <i>Francesco C. Billari, Osea Giuntella and Luca Stella</i> | |

Solicited Sessions

| | |
|--|-----|
| Advanced Statistical Methods in Health Analytics | 209 |
| Assessing the impact of the intermediate event in a non-markovian illness-death model. | 210 |
| <i>Davide Paolo Bernasconi, Elena Tassistro, Maria Grazia Valsecchi and Laura Antolini</i> | |
| Big data and AI: challenges and opportunities in healthcare. | 216 |
| <i>Vieri Emiliani, Gian Luca Cattani and Fabrizio Selmi</i> | |
| Statistical methodology for volume-outcome studies. | 222 |
| <i>Marta Fiocco and Floor van Oudenhoven</i> | |
| Advances in textual data mining | 228 |
| Distance measures for exploring pairs of novels in a large corpus of Italian literature. | 229 |
| <i>Matilde Trevisani and Arjuna Tuzzi</i> | |
| Supervised vs Unsupervised Latent Dirichlet Allocation: topic detection in lyrics. | 235 |
| <i>Mariangela Sciandra, Alessandro Albano and Irene Carola Spera</i> | |
| Advances in the interaction between artificial intelligence and official statistics | 241 |
| Automated Land Cover Maps from Satellite Imagery by Deep Learning. | 242 |
| <i>Fabrizio De Fausti, Francesco Pugliese and Diego Zardetto</i> | |
| CROWD4SDG: Crowdsourcing for sustainable developments goals. | 248 |
| <i>Barbara Pernici</i> | |
| Permanent Population Census: evaluation of the effects of regional strategies on the process efficiency. The direct experience of Tuscany. | 253 |
| <i>Linda Porciani, Luisa Francovich, Luca Faustini and Alessandro Valentini</i> | |
| Capture-recapture methods | 259 |
| Bayesian Model Averaging for Latent Class Models in Capture-Recapture. | 260 |
| <i>Davide Di Cecco</i> | |
| Combining "signs of life" and survey data through latent class models to consider over-coverage in Capture-Recapture estimates of population counts. | 266 |
| <i>Marco Fortini, Antonella Bernardini, Marco Caputi and Nicoletta Cibella</i> | |
| Population size estimation with interval censored counts and external information. | 272 |
| <i>Alessio Farcomeni</i> | |
| Changes in environment extremes and their impacts | 278 |
| FPCA Clustering of rainfall events. | 279 |
| <i>Gianluca Sottile, Antonio Francipane, Leonardo Noto and Giada Adelfio</i> | |
| Trends in rainfall extremes in the Venice lagoon catchment. | 285 |
| <i>Ilaria Prosdocimi and Carlo Gaetan</i> | |

| | |
|--|------------|
| Copulas: models and inference | 291 |
| Analysis of district heating demand through different copula-based approaches. <i>F. Marta L. Di Lascio and Andrea Menapace</i> | 292 |
| CoVaR and backtesting: a comparison between a copula approach and parametric models. <i>Michele Leonardo Bianchi, Giovanni De Luca and Giorgia Riveccio</i> | 298 |
| Estimating Asymmetric Dependence via Empirical Checkerboard Copulas. <i>Wolfgang Trutschnig and Florian Griessenberger</i> | 304 |
| Strong Convergence of Multivariate Maxima. <i>Michael Falk, Simone A. Padoan and Stefano Rizzelli</i> | 310 |
| Data Science: when different expertise meet | 316 |
| Bayesian stochastic modelling of the temporal evolution of seismicity. <i>Elisa Varini and Renata Rotondi</i> | 317 |
| Cluster Analysis for the Characterization of Residential Personal Exposure to ELF Magnetic Field. <i>Gabriella Tognola, Silvia Gallucci, Marta Bonato, Emma Chiaramello, Isabelle Magne, Martine Souques, Serena Fiocchi, Marta Parazzini and Paolo Ravazzani</i> | 323 |
| Statistical Assessment and Validation of Ship Response in High Sea State by Computational Fluid Dynamics. <i>Andrea Serani, Matteo Diez and Frederick Stern</i> | 328 |
| Uncertainty Quantification for PDEs with random data using the Multi-Index Stochastic Collocation method. <i>Lorenzo Tamellini and Joakim Beck</i> | 334 |
| Emerging challenges in official statistics: new data sources and methods | 340 |
| Small area poverty indicators adjusted using local spatial price indices. <i>Stefano Marchetti, Luigi Biggeri, Caterina Giusti and Monica Pratesi</i> | 341 |
| Smart solutions for trusted smart statistics: the European big data hackathon experience. <i>Francesco Amato, Mauro Bruno, Tania Cappadozzi, Fabrizio De Fausti and Manuela Michelini</i> | 347 |
| The ESSnet Project Smart Surveys: new data sources and tools for Surveys of Official Statistics | 353 |
| Factorial and dimensional reduction methods for the construction of indicators for evaluation (SVQS-SIS)..... | 359 |
| A comparison of MBC with CLV and PCovR methods for dimensional reduction of the soccer players' performance attributes. <i>Maurizio Carpita, Enrico Ciavolino and Paola Pasca</i> | 360 |
| A framework of cumulated chi-squared type statistics for ordered correspondence analysis. New tools and properties. <i>Antonello D'Ambra, Pietro Amenta and Luigi D'Ambra</i> | 366 |
| Exploring drug consumption via an ultrametric correlation matrix. <i>Giorgia Zaccaria and Maurizio Vichi</i> | 372 |
| Ranking extraction in ordinal multi-indicator systems. <i>Marco Fattore and Alberto Arcagni</i> | 378 |
| Gender statistics | 384 |
| Gender differences in Italian STEM degree courses: a discrete-time competing-risks model. <i>Marco Enea and Massimo Attanasio</i> | 385 |
| Some Challenges and Results in Measuring Gender Inequality. <i>Fabio Crescenzi and Francesco Di Pede</i> | 391 |

| | |
|--|------------|
| How Deep is Your Plot? Young SIS and deep statistical learning (ySIS).. | 397 |
| A modal approach for clustering matrices. | 398 |
| <i>Federico Ferraccioli and Giovanna Menardi</i> | |
| A Note on Detection of Perturbations in Biological Networks. | 404 |
| <i>Vera Djordjilović</i> | |
| Bayesian inference for DAG-probit models. | 410 |
| <i>Federico Castelletti</i> | |
| Variational Bayes for Gaussian Factor Models under the Cumulative Shrinkage Process. | 416 |
| <i>Sirio Legramanti</i> | |
| Measuring poverty and vulnerability | 421 |
| Choosing the vulnerability threshold using the ROC curve. | 422 |
| <i>Chiara Gigliarano and Conchita D'Ambrosio</i> | |
| New advances in applications, a Bayesian nonparametric perspective | 428 |
| Bayesian Mixture Models for Latent Class Analysis. | 429 |
| <i>Raffaele Argiento, Bruno Bodin and Maria De Iorio</i> | |
| Non-Parametric Inference and Forecasting of Functional and Object Data | 435 |
| An interpretable estimator for the function-on-function linear regression model with application to the Canadian weather data. | 436 |
| <i>Fabio Centofanti and Matteo Fontana</i> | |
| Statistical process monitoring of multivariate profiles from ship operating conditions. | 440 |
| <i>Christian Capezza</i> | |
| Prior choice in Bayesian Modelling (SISbayes) | 446 |
| Bayesian Learning of Multiple Essential Graphs. | 447 |
| <i>Luca La Rocca, Federico Castelletti, Stefano Peluso, Francesco Claudio Stingo and Guido Consonni</i> | |
| Bayesian post-processing of Gibbs sampling output for variable selection. | 453 |
| <i>Stefano Cabras</i> | |
| Priors on precision parameters of IGRMF models. | 459 |
| <i>Aldo Gardini, Fedele Greco and Carlo Trivisano</i> | |
| Sequence Analysis: methods and applications | 465 |
| Internal migration, family formation and social stratification in Europe. A life course approach. | 466 |
| <i>Roberto Impicciatore, Gabriele Ballarino and Nazareno Panichella</i> | |
| Socio economic integration of migrants | 472 |
| A study on the characteristics of spouses who intermarry in Italy. | 473 |
| <i>Agnese Vitali and Romina Fraboni</i> | |
| Statistical Analysis for mobility and transportation | 479 |
| A multilevel Analysis of University attractiveness in the network flows from Bachelor to Master's degree. | 480 |
| <i>Silvia Columbu and Ilaria Primerano</i> | |
| Analysis of mobility data through a novel Cheng and Church algorithm for functional data. | 486 |
| <i>Marta Galvani, Agostino Torti and Alessandra Menafoglio</i> | |
| Bridge closures in a transportation network: analysis of the impacts in the region of Lombardy. | 491 |
| <i>Agostino Torti, Marika Arena, Giovanni Azzone, and Piercesare Secchi</i> | |

| | |
|--|------------|
| Statistical Methods and Applications in Social Network Analysis | 496 |
| A clustering procedure for ego-networks data: an application to Italian elders living in couple. <i>Elvira Pelle and Roberta Pappadà</i> | 497 |
| Analysing the mediating role of a network: a Bayesian latent space approach. <i>Chiara Di Maria, Antonino Abbruzzo and Gianfranco Lovison</i> | 503 |
| Network-time autoregressive models for valued network panel. <i>Viviana Amati</i> | 509 |
| University student mobility flows and network data structures. <i>Maria Prosperina Vitale, Giuseppe Giordano and Giancarlo Ragozini</i> | 515 |
| Statistical Methods in Psychometrics | 521 |
| A simple probabilistic model to evaluate questionable interim analysis strategies. <i>Francesca Freuli and Luigi Lombardi</i> | 522 |
| Incorporating Expert Knowledge in Structural Equation Models: Applications in Psychological Research. <i>Gianmarco Altoè, Claudio Zandonella Callegher, Enrico Toffalini and Massimiliano Pastore</i> | 528 |
| Predicting social media addiction from Instagram profiles: A data mining approach. <i>Antonio Calcagni, Veronica Cortellazzo, Francesca Guizzo, Paolo Girardi, Natale Canale</i> | 534 |
| Structural entropy based modeling for psychological measurement. <i>Enrico Ciavolino, Mario Angelelli, Paola Pasca and Omar Carlo Gioacchino Gelo</i> | 540 |
| Statistical modelling in environmental epidemiology | 546 |
| A Time Varying Coefficient Model to Estimate the Short-Term Effects of Air Pollution on Human Health. <i>Pasquale Valentini, Luigi Ippoliti and Clara Grazian</i> | 547 |
| Joint Analysis of Short and Long-Term Effects of Air Pollution. <i>Annibale Biggeri, Dolores Catelan, Giorgia Stoppa and Corrado Lagazio</i> | 551 |
| Statistical Modelling of Scientific Evidence for Forensic Investigation and Interpretation | 557 |
| DNA mixtures with related contributors. <i>Peter J. Green and Julia Mortera</i> | 558 |
| Forensic Statistics: How to estimate life expectancy after injury. <i>Jane L Hutton</i> | 564 |
| The additional contribution of combining genetic evidence from multiple samples in a complex case. <i>Giampietro Lago</i> | 570 |
| The history of forensic inference and statistics: a thematic perspective. <i>Franco Taroni and Colin Aitken</i> | 576 |
| Topological learning: interpretable representations of complex data..... | 581 |
| Comparing Neural Networks via Generalized Persistence. <i>Mattia G. Bergomi and Pietro Vertechi</i> | 582 |
| On the topological complexity of decision boundaries. <i>António Leitão and Giovanni Petri</i> | 588 |
| Persistence-based Kernels for Data Classification. <i>Ulderico Fugacci</i> | 594 |
| Topological and Mixed-type learning of Brain Activity. <i>Tullia Padellini, Pierpaolo Brutti, Riccardo Giubilei</i> | 600 |

Contributed papers and Posters

| | |
|---|------------|
| Bayesian Statistics | 607 |
| A Bayesian approach for modelling dependence among mixture densities. <i>Mario Beraha, Matteo Pegoraro, Riccardo Peli and Alessandra Guglielmi</i> | 608 |
| A change of glasses strategy to solve the rare type match problem. <i>Giulia Cereda and Fabio Corradi</i> | 614 |
| A new prior distribution on the simplex: the extended flexible Dirichlet. <i>Roberto Ascari, Sonia Migliorati and Andrea Ongaro</i> | 620 |
| ABC model choice via mixture weight estimation. <i>Gianmarco Caruso, Luca Tardella and Christian P. Robert</i> | 626 |
| An ABC algorithm for random partitions arising from the Dirichlet process. <i>Mario Beraha and Riccardo Corradin</i> | 632 |
| Bayesian Inference of Undirected Graphical Models from Count Data. <i>Pier Giovanni Bissiri, Monica Chiogna and Nguyen Thi Kim Hue</i> | 638 |
| Bayesian IRT models in NIMBLE. <i>Sally Paganin, Chris Paciorek and Perry de Valpine</i> | 644 |
| Bayesian modelling of Facebook communities via latent factor models. <i>Emanuele Aliverti</i> | 650 |
| Bayesian nonparametric adaptive classification with robust prior information. <i>Francesco Denti, Andrea Cappozzo and Francesca Greselin</i> | 655 |
| Choosing the right tool for the job: a systematic analysis of general purpose MCMC software. <i>Mario Beraha, Giulia Gualtieri, Eugenia Villa, Riccardo Vitali and Alessandra Guglielmi</i> | 661 |
| Empirical Bayes estimation for mixture models. <i>Catia Scricciolo</i> | 667 |
| Improving ABC via Large Deviations Theory. <i>Cecilia Viscardi, Michele Boreale and Fabio Corradi</i> | 673 |
| Learning Bayesian Networks for Nonparanormal Data. <i>Flaminia Musella and Vincenzina Vitale</i> | 679 |
| Measuring well-being combining different data sources: a Bayesian networks approach. <i>Federica Cugnata, Silvia Salini and Elena Siletti</i> | 685 |
| Penalising the complexity of extensions of the Gaussian distribution. <i>Diego Battagliese and Brunero Liseo</i> | 691 |
| Predictive discrepancy of credible intervals for the parameter of the Rayleigh distribution. <i>Fulvio De Santis and Stefania Gubbiotti</i> | 697 |
| Small-area statistical estimation of claim risk. <i>Francesca Fortunato, Fedele Greco and Pierpaolo Cristaudo</i> | 702 |
| Subject-specific Bayesian Hierarchical model for compositional data analysis. <i>Matteo Pedone and Francesco C. Stingo</i> | 708 |
| Wasserstein consensus for Bayesian sample size determination. <i>Michele Cianfriglia, Tullia Padellini and Pierpaolo Brutti</i> | 714 |
| Biostatistics | 720 |
| A comparison of the CAR and DAGAR spatial random effects models with an application to diabetes rate estimation in Belgium. <i>Vittoria La Serra, Christel Faes, Niel Hens and Pierpaolo Brutti</i> | 721 |
| A functional approach to study the relationship between dynamic covariates and survival outcomes: an application to a randomized clinical trial on osteosarcoma. <i>Marta Spreafico, Francesca Ieva and Marta Fiocco</i> | 727 |

| | |
|---|------------|
| A Statistical Approach to the Alignment of fMRI Data. <i>Angela Andreella, Ma Feilong, Yaroslav Halchenko, James Haxby and Livio Finos</i> | 733 |
| Adaptive clinical trials: Bayesian decision-theoretic and frequentist approaches for cost-effectiveness analysis. <i>Martin Forster and Marco Novelli</i> | 739 |
| Bootstrap corrected Propensity Score: Application for Anticoagulant Therapy in Haemodialysis Patients. <i>Maeregu W. Arisido, Fulvia Mecatti and Paola Rebora</i> | 745 |
| Combining multiple sources to overcome misclassification bias in epidemiological database studies. <i>Francesca Beraldi, Rosa Gini, Emanuela Dreassi, Leonardo Grilli and Carla Rampichini</i> | 751 |
| Deep Sparse Autoencoder-based Feature Selection for SNPs Validation in Prostate Cancer Radiogenomics. <i>Michela Carlotta Massi, Francesca Ieva, Anna Maria Paganoni, Andrea Manzoni, Paolo Zunino, Nicola Rares Franco, Tiziana Rancati and Catharine West</i> | 756 |
| Graphical models for count data: an application to single-cell RNA sequencing. <i>Nguyen Thi Kim Hue, Monica Chiogna and Davide Rizzo</i> | 762 |
| Interregional mobility, socio-economic inequality and mortality among cancer patients. <i>Claudio Rubino, Mauro Ferrante, Antonino Abbruzzo, Giovanna Fantaci and Salvatore Scondotto</i> | 768 |
| PET radiomics-based lesions representation in Hodgkin lymphoma patients. <i>Lara Cavinato, Martina Sollini, Margarita Kirienko, Matteo Biroli, Francesca Ricci, Letizia Calderoni, Elena Tabacchi, Cristina Nanni, Pier Luigi Zinzani, Stefano Fanti, Anna Guidetti, Alessandra Alessi, Paolo Corradini, Ettore Seregni, Carmelo Carlo-Stella, Arturo Chiti and Francesca Ieva</i> | 774 |
| Prediction of late radiotherapy toxicity in prostate cancer patients via joint analysis of SNPs sequences. <i>Nicola Rares Franco, Michela Carlotta Massi, Francesca Ieva, Anna Maria Paganoni, Andrea Manzoni, Paolo Zunino, Tiziana Rancati and Catharine West</i> | 780 |
| Predictive versus posterior probabilities for phase II trial monitoring. <i>Valeria Sambucini</i> | 785 |
| Profile networks for precision medicine. <i>Andrea Lazerini, Monia Lupporelli and Francesco C. Stingo</i> | 791 |
| Proton-Pump Inhibitor Provider Profiling via Funnel Plots and Poisson Regression. <i>Dario Delle Vedove, Francesca Ieva and Anna Maria Paganoni</i> | 797 |
| Selecting optimal thresholds in ROC analysis with clustered data. <i>Duc Khanh To, Gianfranco Adimari and Monica Chiogna</i> | 803 |
| Environment, Physics and Engineering | 809 |
| A hidden semi-Markov model for segmenting environmental toroidal data. <i>Francesco Lagona and Antonello Maruotti</i> | 810 |
| An experimental analysis on quality and security about green communication. <i>Vito Santarcangelo, Emilio Massa, Davide Scintu, Michele Di Lecce and Massimiliano Giacalone</i> | 816 |
| An improved sensitivity-data based method for probabilistic ecological risk assessment. <i>Sonia Migliorati and Gianna Serafina Monti</i> | 822 |
| Comparing predictive distributions in EMOS. <i>Giummolè Federica and Mameli Valentina</i> | 828 |
| Compositional analysis of fish communities in a fast changing marine ecosystem. <i>Pierfrancesco Alaimo Di Loro, Marco Mingione, Giovanna Jona Lasinio, Sara Martino and Francesco Colloca</i> | 834 |
| FDA dimension reduction techniques and components separation in Fourier-transform infrared spectroscopy. <i>Francesca Di Salvo, Elena Piacenza and Delia Francesca Chillura Martino</i> | 840 |
| Functional Data Analysis for Spectroscopy Data. <i>Mara S. Bernardi, Matteo Fontana, Alessandra Menafoglio, Diego Perugini, Alessandro Pisello, Marco Ferrari, Simone De Angelis, Maria Cristina De Sanctis and Simone Vantini</i> | 846 |
| Functional graphical model for spectrometric data analysis. <i>Laura Codazzi, Alessandro Colombi, Matteo Gianella, Raffaele Argiento, Lucia Paci and Alessia Pini</i> | 852 |
| Local LGCP estimation for spatial seismic processes. <i>Nicoletta D'Angelo, Marianna Siino, Antonino D'Alessandro and Giada Adelfio</i> | 857 |

| | |
|--|------------|
| Observation-driven models for storm counts. <i>Mirko Armillotta, Alessandra Luati and Monia Lupparelli</i> | 863 |
| Statistical control of complex geometries, with application to Additive Manufacturing. <i>Riccardo Scimone, Tommaso Taormina, Bianca Maria Colosimo, Marco Grasso, Alessandra Menafoglio, Piercesare Secchi</i> | 869 |
| Tree attributes map by 3P sampling in a design-based framework. <i>Lorenzo Fattorini and Sara Franceschi</i> | 875 |
| Unsupervised classification of texture images by gray-level spatial dependence matrices and genetic algorithms. <i>Roberto Baragona and Laura Bocci</i> | 880 |
| Finance, business and official statistics | 886 |
| A discrete choice approach to analyze contractual attributes in the durum wheat sector in Italy. <i>Stefano Ciliberti, Simone Del Sarto, Giulia Pastorelli, Angelo Frascarelli and Gaetano Martino</i> | 887 |
| A fuzzy approach to the measurement of the employment rate. <i>Bruno Cheli, Alessandra Coli and Andrea Regoli</i> | 893 |
| A proposal to model credit risk contagion using network count-based models. <i>Arianna Agosto and Daniel Felix Ahelegbey</i> | 898 |
| A similarity matrix approach to empower ESCO interfaces for testing, debugging and in support of users' experience. <i>Adham Kahlawi, Cristina Martelli, Lucia Buzzigoli, Laura Grassini</i> | 904 |
| Adding MIDAS terms to Linear ARCH models in a Quantile Regression framework. <i>Vincenzo Candila and Lea Petrella</i> | 910 |
| Company requirements in Italian tourism sector: an analysis for profiles. <i>Paolo Mariani, Andrea Marletta, Lucio Masserini and Mariangela Zenga</i> | 916 |
| Determinants of Firms' Default Risk after the 2008 and 2011 Economic Crises: a Latent Growth Models Approach. <i>Lucio Masserini, Matilde Bini and Alessandro Zeli</i> | 921 |
| Double Asymmetric GARCH-MIDAS model - new insights and results. <i>Alessandra Amendola, Vincenzo Candila and Giampiero M. Gallo</i> | 927 |
| European SMEs and Circular Economy Activities: Evaluating the Advantage on Firm Performance through the Estimation of Average Treatment Effects. <i>Luca Secondi</i> | 933 |
| Financial Spillover Measures to Assess the Stability of Basket-based Stablecoins. <i>Paolo Pagnottoni</i> | 939 |
| Forecasting Banknote Flows in Bdl Branches: Speed-up with Machine Learning. <i>Marco Brandi, Monica Fusaro, Tiziana Laureti and Giorgia Rocco</i> | 945 |
| Fully reconciled GDP forecasts from Income and Expenditure sides. <i>Luisa Bisaglia, Tommaso Di Fonzo and Daniele Girolimetto</i> | 951 |
| GLASSO Estimation of Commodity Risks. <i>Beatrice Foroni, Saverio Mazza, Giacomo Morelli and Lea Petrella</i> | 957 |
| Measuring the Effect of Unconventional Policies on Stock Market Volatility. <i>Giampiero M. Gallo, Demetrio Lacava and Edoardo Otranto</i> | 963 |
| Multidimensional versus unidimensional poverty measurement. <i>Michele Costa</i> | 969 |
| Multiple outcome analysis of European Agriculture in 2000-2016: a latent class multivariate trajectory approach. <i>Alessandro Magrini</i> | 975 |
| Nowcasting GDP using mixed-frequency based composite confidence indicators. <i>Maria Carannante, Raffaele Mattera, Michelangelo Misuraca, Germana Scepi and Maria Spano</i> | 981 |
| On the tangible and intangible assets of Initial Coin Offerings. <i>Paola Cerchiello and Anca Mirela Toma</i> | 987 |

| | |
|--|-------------|
| Seasonality variation of electricity demand: decompositions and tests. <i>Luigi Grossi and Mauro Mussini</i> | 993 |
| SMEs circular economy practices in the European Union: Implications for sustainability. <i>Nunzio Tritto, José G. Dias and Francesca Bassi</i> | 999 |
| Tax Incentives' Effect on the Provision of Occupational Welfare in Italian Enterprises. <i>Alessandra Righi</i> | 1005 |
| The determinants of eco-innovation: a country comparison using the community innovation survey. <i>Ida D'Attoma and Silvia Pacei</i> | 1011 |
| World ranking of urban sustainability through composite indicators. <i>Elena Grimaccia, Alessia Naccarato and Silvia Terzi</i> | 1017 |
| Machine Learning and Data Science..... | 1023 |
| A novel approach for Artificial Intelligence through Lorenz zonoids and Shapley Values. <i>Paolo Giudici and Emanuela Raffinetti</i> | 1024 |
| A warning signal for variable importance interpretation in tree-based algorithms. <i>Anna Gottard and Giulia Vannucci</i> | 1030 |
| Assessment of the effectiveness of digital flyers: analysis of viewing behavior using eye tracking. <i>Gianpaolo Zammarchi, Claudio Conversano and Francesco Mola</i> | 1036 |
| At risk mental status analysis: a comparison of model selection methods for ordinal target variable. <i>Elena Ballante, Silvia Molteni, Martina Mensi and Silvia Figini</i> | 1042 |
| Categorical Encoding for Machine Learning. <i>Agostino Di Ciaccio</i> | 1048 |
| Dynamic Quantile Regression Forest. <i>Mila Andreani and Lea Petrella</i> | 1054 |
| Estimating the UK Sentiment Using Twitter. <i>Stephan Schlosser, Daniele Toninelli and Michela Cameletti</i> | 1059 |
| Forecasting local rice prices from crowdsourced data in Nigeria. <i>Ilaria Lucrezia Amerise and Gloria Solano Hermosilla</i> | 1065 |
| Generalized Mixed Effects Random Forest: does Machine Learning help in predicting university student dropout? <i>Massimo Pellagatti, Chiara Masci, Francesca Ieva and Anna Maria Paganoni</i> | 1071 |
| HateViz: a textual dashboard Twitter data-driven. <i>Emma Zavarrone, Maria Gabriella Grassia, Marina Marino, Rocco Mazza and Nicola Canestrari</i> | 1077 |
| How to perform cyber risk assessment via cumulative logit models. <i>Silvia Facchinetti, Silvia Angela Osmetti and Claudia Tarantola</i> | 1083 |
| Machine learning prediction for accounting system. <i>Chiara Bardelli and Silvia Figini</i> | 1087 |
| Teaching statistics: an assessment framework based on Multidimensional IRT and Knowledge Space Theory. <i>Cristina Davino, Rosa Fabbriatore, Carla Galluccio, Daniela Pacella, Domenico Vistocco, Francesco Palumbo</i> | 1093 |
| The weight of words: textual data versus sentiment analysis in stock returns prediction. <i>Riccardo Ferretti and Andrea Sciandra</i> | 1099 |
| Unsupervised Energy Trees: clustering with complex and mixed-type variables. <i>Riccardo Giubilei, Tullia Padellini and Pierpaolo Brutti</i> | 1105 |
| Using anchoring vignettes to adjust self-reported life satisfaction: a nonparametric approach leading to a Semantic Differential scale. <i>Sara Garbin, Serena Berretta, Maria Iannario and Omar Paccagnella</i> | 1111 |
| Variable selection for robust model-based learning from contaminated data. <i>Andrea Cappozzo, Francesca Greselin and Thomas Brendan Murphy</i> | 1117 |

| | |
|--|-------------|
| Variable Selection in Text Regressions: Back to Lasso? <i>Marzia Freo and Alessandra Luati</i> | 1123 |
| Web Usage Mining and Website Effectiveness. <i>Maria Francesca Cracolici and Furio Urso</i> | 1129 |
| Models and methods - Categorical, Ordinal, Rank Data | 1135 |
| Aberration for the analysis of two-way contingency tables. <i>Roberto Fontana and Fabio Rapallo</i> | 1136 |
| An investigation of the paradoxical behaviour of κ -type inter-rater agreement coefficients for nominal data. <i>Amalia Vanacore and Maria Sole Pellegrino</i> | 1142 |
| Analyzing faking-good response data: Combination of a Replacement and a Binomial (CRB) distribution approach. <i>Luigi Lombardi and Antonio Calcagni</i> | 1148 |
| BOD – min range: A Robustness Analysis Method for Composite Indicators. <i>Emiliano Seri, Leonardo Salvatore Alaimo and Vittoria Carolina Malpassuti</i> | 1154 |
| Comparing classifiers for ordinal variables. <i>Silvia Golia and Maurizio Carpita</i> | 1160 |
| Discovering Interaction Effects Between Subject-Specific Covariates: A New Probabilistic Approach For Preference Data. <i>Alessio Baldassarre, Claudio Conversano, Antonio D'Ambrosio, Mark De Rooij and Elise Dusseldorp</i> | 1166 |
| Hybrid random forests for ordinal data. <i>Rosaria Simone and Gerhard Tutz</i> | 1171 |
| Model-based approach to biclustering ordinal data. <i>Monia Ranalli and Francesca Martella</i> | 1177 |
| New algorithms and goodness-of-fit diagnostics for ranked data modelling with the Extended Plackett-Luce distribution. <i>Cristina Mollica and Luca Tardella</i> | 1183 |
| Non-metric unfolding on augmented data matrix: a copula-based approach. <i>Marta Nai Ruscone and Antonio D'Ambrosio</i> | 1189 |
| Ordinal probability effect measures for dyadic analysis in cumulative models. <i>Maria Iannario and Domenico Vistocco</i> | 1194 |
| Simulated annealing for maximum rater agreement. <i>Fabio Rapallo and Maria Piera Rogantin</i> | 1200 |
| Models and methods – Regression..... | 1206 |
| A Clusterwise regression method for Distributional-valued Data. <i>Rosanna Verde, Francisco de A. T. de Carvalho and Antonio Balzanella</i> | 1207 |
| A nonparametric approach for nonlinear variable screening in high-dimensions. <i>Francesco Giordano, Sara Milito and Lucia Maria Parrella</i> | 1213 |
| Adjusted scores for inference in negative binomial regression. <i>Euloge C. Kenne Pagui, Alessandra Salvan and Nicola Sartori</i> | 1219 |
| Estimation of the treatment effect variance in a difference-in-differences framework. <i>Marco Doretti and Giorgio E. Montanari</i> | 1224 |
| Exploring multicollinearity in quantile regression. <i>Cristina Davino, Tormod Naes, Rosaria Romano and Domenico Vistocco</i> | 1230 |
| Generalized M-quantile random effects model. <i>Francesco Schirripa Spagnolo and Vincenzo Mauro</i> | 1236 |
| Goodness-of-fit assessment in linear quantile regression. <i>Ilaria Lucrezia Amerise and Agostino Tarsitano</i> | 1242 |
| Joint Redundancy Analysis by a multivariate linear predictor. <i>Laura Marcis and Renato Salvatore</i> | 1248 |

| | |
|--|-------------|
| M-quantile regression shrinkage and selection via the lasso. <i>M. Giovanna Ranalli, Lea Petrella and Francesco Pantalone</i> | 1254 |
| New insights into the Conditioning and Gain Score approaches in multilevel analysis. <i>Bruno Arpino, Silvia Bacci, Leonardo Grilli, Raffaele Guetto and Carla Rampichini</i> | 1260 |
| Simultaneous confidence regions and curvature measures in nonlinear models. <i>Claudia Furlan and Cinzia Mortarino</i> | 1265 |
| Models and methods – Sampling | 1271 |
| Design-based consistency of the Horvitz-Thompson estimator for spatial populations. <i>Lorenzo Fattorini, Marzia Marcheselli, Caterina Pisani and Luca Pratelli</i> | 1272 |
| Empirical likelihood in the statistical matching for informative samples. <i>Daniela Marella and Danny Pfeffermann</i> | 1278 |
| Evaluating a Hybrid One-Stage Snowball Sampling through Bootstrap Method on a Simulated Population. <i>Venera Tomaselli and Giulio Giacomo Cantone</i> | 1284 |
| How optimal subsampling depends on guessed parameter values. <i>Laura Deldossi and Chiara Tommasi</i> | 1290 |
| Indicators for risk of selection bias in non-probability samples. <i>Emilia Rocco and Alessandra Petrucci</i> | 1296 |
| On the behaviour of the maximum likelihood estimator for exponential models under a fixed and a two-stage design. <i>Caterina May and Chiara Tommasi</i> | 1302 |
| Pseudo-population based resamplings for two-stage design. <i>Pier Luigi Conti, Daniela Marella and Vincenzina Vitale</i> | 1308 |
| Models and methods - Theoretical Issues in Statistical Inference | 1314 |
| A new mixture model for three-way data. <i>Salvatore D. Tomarchio, Antonio Punzo and Luca Bagnato</i> | 1315 |
| A Sequential Test for the C _{pmk} Index. <i>Michele Scagliarini</i> | 1320 |
| Probability Interpretations and the Selection of the Most Effective Statistics Method. <i>Paolo Rocchi</i> | 1326 |
| Robust Composite Inference. <i>Valentina Mamei, Monica Musio, Erlis Ruli and Laura Ventura</i> | 1332 |
| Statistical hypothesis testing within the Generalized Error Distribution: Comparing the behavior of some nonparametric techniques. <i>Massimiliano Giacalone and Demetrio Panarello</i> | 1338 |
| Stochastic dependence with discrete copulas. <i>Fabrizio Durante and Elisa Perrone</i> | 1344 |
| Models and methods - Time Series and Longitudinal Data..... | 1350 |
| Bootstrap test in Poisson-INAR models. <i>Lucio Palazzo and Riccardo Ievoli</i> | 1351 |
| Continuous Time-Interaction Processes for Population Size Estimation. <i>Linda Altieri, Alessio Farcomeni, Danilo Alunni Fegatelli and Francesco Palini</i> | 1357 |
| Longitudinal data analysis using PLS-PM approach. <i>Rosanna Cataldo, Corrado Crocetta, Maria Gabriella Grassia and Marina Marino</i> | 1363 |
| Long-memory models for count time series. <i>Luisa Bisaglia, Massimiliano Caporin and Matteo Grigoletto</i> | 1369 |

| | |
|---|-------------|
| Combining multiple frequencies in Realized GARCH models. <i>Antonio Naimoli and Giuseppe Storti</i> | 1375 |
| Models with Time-Varying Parameters for Realized Covariance. <i>Luc Bauwens and Edoardo Otranto</i> | 1381 |
| Pitman-Yor mixture models for survival data stratification. <i>Riccardo Corradin, Luis Enrique Nieto Barajas and Bernardo Nipoti</i> | 1387 |
| Prediction is not everything, but everything is prediction. <i>Leonardo Egidi</i> | 1393 |
| The Generalized Dynamic Mixtures of Factor Analyzers for clustering multivariate longitudinal data. <i>Francesca Martella, Antonello Maruotti and Francesco Tursini</i> | 1399 |
| Trends and long-run relations in cointegrated time series observed with noise. <i>Angelica Gianfreda, Paolo Maranzano, Lucia Parisio and Matteo Pelagatti</i> | 1405 |
| Population and society | 1411 |
| A dimensionality assessment of refugees' vulnerability through an Item Response Theory approach. <i>Simone Del Sarto, Michela Gnalzi, Yara Maasri and Edouard Legoupil</i> | 1412 |
| Accounting for Interdependent Risks in Vulnerability Assessment of Refugees. <i>Daria Mendola, Anna Maria Parroco and Paolo Li Donni</i> | 1418 |
| Active ageing in China: What are the domains that most affect life satisfaction in the elderly? <i>Ilaria Rocco</i> | 1424 |
| Analyzing the waiting time of academic publications: a survival model. <i>Francesca De Battisti, Giuseppe Gerardi, Giancarlo Manzi and Francesco Porro</i> | 1430 |
| Clustering of food choices in a large sample of students using university canteen. <i>Valentina Lorenzoni, Isotta Triulzi, Irene Martinucci, Letizia Toncelli, Michela Natilli and Roberto Barale, Giuseppe Turchetti</i> | 1436 |
| Cruise passengers' expenditure at destinations: Review of survey techniques and data collection. <i>Caterina Sciortino, Stefano De Cantis, Mauro Ferrante and Szilvia Gyimóthy</i> | 1442 |
| Educational integration of foreign citizen children in Italy: a synthetic indicator. <i>Alessio Buonomo, Stefania Capecchi and Rosaria Simone</i> | 1448 |
| Estimating the Change in Housework Time of the Italian Woman after the Retirement of the Male Partner: An Approach Based on a Two-Regime Model Estimated by ML. <i>Giorgio Calzolari, Maria Gabriella Campolo, Antonino Di Pino and Laura Magazzini</i> | 1454 |
| First and Second Year Careers of STEM Students in Italy: A Geographical Perspective. <i>Antonella D'Agostino, Giulio Ghellini and Gabriele Lombardi</i> | 1460 |
| Future Scenarios and Support Interventions for the Family: Involving Experts' Participation through a Mixed-Method Research Study. <i>Mario Bolzan, Simone Di Zio, Manuela Scioni and Morena Tartari</i> | 1466 |
| Gender and Monetary Policy Preferences: a Diff-in-Diff Approach. <i>Donata Favaro, Anna Giraldo and Ina Gollikja</i> | 1472 |
| Headcount based indicators and functions to evaluate the effectiveness of Italian university education. <i>Silvia Terzi and Francesca Petrarca</i> | 1478 |
| Identify the speech code through statistics: a data-driven approach. <i>Andrea Briglia, Massimo Mucciardi and Jérémi Sauvage</i> | 1484 |
| Inspecting cause-specific mortality curves by simplicial functional data analysis. <i>Marco Stefanucci and Stefano Mazzucco</i> | 1490 |
| Intertemporal decision making and childless couples. <i>Daniela Bellani, Bruno Arpino and Daniele Vignoli</i> | 1495 |
| Italian Households' Material Deprivation: Multi-Objective Genetic Algorithm approach for categorical variables. <i>Laura Bocci and Isabella Mingo</i> | 1501 |

| | |
|--|------|
| LI-CoD Model. From Lifespan Inequality to Causes of Death. <i>Andrea Nigri and Susanna Levantesi</i> | 1507 |
| Modeling Well-Being through PLS-SEM and K-M. <i>Venera Tomaselli, Mario Fordellone and Maurizio Vichi</i> | 1513 |
| News life-cycle: a multiblock approach to the study of information. <i>Rosanna Cataldo, Marco Del Mastro, Maria Gabriella Grassia, Marina Marino and Rocco Mazza</i> | 1519 |
| Short-term rentals in a tourist town. <i>Silvia Bacci, Bruno Bertaccini, Gianni Dugheri, Paolo Galli, Antonio Giusti and Veronica Sula</i> | 1525 |
| Sportstat: a playful activity to developing statistical literacy. <i>Alessandro Valentini and Francesca Paradisi</i> | 1531 |
| Statistical modeling for some features of Airbnb activity. <i>Giulia Contu and Luca Frigau</i> | 1537 |
| Tertiary students with migrant background: evidence from a cohort enrolled at Sapienza University. <i>Cristina Giudici, Donatella Vicar and Eleonora Trappolini</i> | 1543 |
| The Causal Effect of Immigration Policies on Income Inequality. <i>Irene Crimaldi, Laura Forastiere, Fabrizia Mealli and Costanza Tortù</i> | 1549 |
| The job condition of academic graduates: a joint longitudinal analysis of AlmaLaurea and Mandatory Notices of the Ministry of Labour. <i>Maria Veronica Dorgali, Silvia Bacci, Bruno Bertaccini and Alessandra Petrucci</i> | 1557 |
| The joint effect of childcare services and flexible female employment on fertility rate in Europe. <i>Viviana Cocuccio and Massimo Mucciardi</i> | 1565 |
| The Left Behind Generation: How the current Early School Leavers affect tomorrow's NEETs? <i>Giovanni De Luca, Paolo Mazzocchi, Claudio Quintano and Antonella Rocca</i> | 1571 |
| The probability to be employed of young adults of foreign origin. <i>Alessio Buonomo, Francesca Di Iorio and Salvatore Strozza</i> | 1577 |
| The risk of inappropriateness in geriatric wards: a comparison among the Italian regions. <i>Paolo Mariani, Andrea Marietta, Marcella Mazzoleni and Mariangela Zenga</i> | 1583 |
| The role of the accumulation of poverty and unemployment for health disadvantages. <i>Annalisa Busetta, Daria Mendola, Emanuela Struffolino and Zachary Van Winkle</i> | 1589 |
| Unemployment and fertility in Italy. A regional level data panel analysis. <i>Gabriele Ruiu and Marco Breschi</i> | 1595 |
| University drop out and mobility in Italy. First evidences on first level degrees. <i>Nicola Tedesco and Luisa Salaris</i> | 1601 |
| Worthiness-based Scale Quantifying. <i>Giulio D'Epifanio</i> | 1607 |
| Young people in Southern Italy and the phenomenon of immigration: what is their perception? <i>Nunziata Ribecco, Angela Maria D'Ugento and Angela Labarile</i> | 1613 |

Modeling Well-Being through *PLS-SEM* and *K-M* *Misurare il benessere con il modello PLS-SEM e K-Means*

Venera Tomaselli¹, Mario Fordellone², Maurizio Vichi³

Abstract The concept of “fair and sustainable” well-being (BES) is employed to estimate the effects of 2008-2016 economic crisis on voting behaviour. Many studies suggest overcoming Gross Domestic Product as the only indicator to measure economic growth and development. So, in addition to strictly economic indicators, social welfare and performance of institutions, and public services indicators are used. With the aim to build a composite indicator of well-being, based on BES elementary indicators and domains, *Partial Least Squares-Structural Equation Modeling* and *K-means* clustering simultaneous method is employed to model the well-being classifying territorial micro-areas.

Abstract *Il benessere "equo e sostenibile" (BES) è impiegato per stimare gli effetti della crisi economica 2008-2016 sul comportamento di voto. Numerosi studi suggeriscono il superamento del prodotto interno lordo come unico indicatore per misurare la crescita e lo sviluppo economico. Pertanto, oltre ad indicatori strettamente economici, l'analisi tiene conto degli indicatori BES relativi al benessere sociale ed alle prestazioni delle istituzioni e dei servizi pubblici. Con l'obiettivo di costruire un indicatore composito di benessere, basato su indicatori elementari e domini del BES, il metodo simultaneo Partial Least Squares-Structural Equation Modeling e K-means clustering è utilizzato per l'analisi del benessere e la classificazione di micro aree territoriali.*

Key words: composite indicators, *Partial Least Squares-Structural Equation Models* and *K-means* method, well-being, territorial micro-areas

¹ Venera Tomaselli, Department of Political and Social Sciences, University of Catania IT; e-mail: venera.tomaselli@unict.it (*corresponding author*).

² Mario Fordellone, Department of Statistical Sciences, University “La Sapienza”, Rome IT; e-mail: mario.fordellone@uniroma1.it.

³ Maurizio Vichi, Department of Statistical Sciences, University “La Sapienza”, Rome IT; e-mail: maurizio.vichi@uniroma1.it.

1 Introduction

The concept of well-being is very complex and takes into account economic resources, public policies, quality of the environment, and many other topics. Well-being indicators, indeed, are proxies of unobservable, latent dimensions.

By questioning the use of Gross Domestic Product (GDP), the study aims to measure well-being through a multidimensional approach on the basis of the relationships among elementary indicators and domains proposed in the analysis of well-being through BES in order to define the territorial differences of well-being more reliably.

The indicators have been selected from 12 BES domains, grouped in turn into 6 topics: economy and labour market, education, environment, institutional performance, safety and quality of services. The crucial issues are:

- is the economic well-being level the only useful and meaningful measurement to explain the variability of territorial micro-areas?
- are indicators closely related to economic well-being and indicators of well-being related, for instance, to the quality of services and opportunities available in the territory, do not necessarily are correlated?
- is the well-being a composite concept based on a scaling of specific priorities?

In the following section 2, the well-being measurement is discussed; in section 3 the PLS-SEM and K-Means method is presented and in sections 4 the results obtained simultaneously through well-being composite indicator and classification method for the territorial micro-areas are delivered.

2 Well-being measurement

The measures of well-being, economic progress, and social welfare are adopted as drivers for designing public policies by decision makers (Jayawickreme et al., 2012; Layan, 2011; Sachs, 2012). The measures more accurately depict changes not only in individual living standards (Helliwell et al., 2012) but simultaneously also in comprehensive national economic growth (Diener et al., 2009).

Since around 2000 the Organisation for Economic Cooperation and Development (OECD) embarked on a global project to measure the well-being and progress of societies not just and not only through the economic performances. The project was involved in setting up and supporting the Commission on the Measurement of Economic Performance and Social Progress (CMEPSP), established by the President of France, Nicolas Sarkozy in 2008 and led by Stiglitz, Sen and Fitoussi.

The limits of GDP are reviewed in the report of the Commission because the GDP is not believed as a standard of the well-being of societies. The GDP, indeed, does not address economic inequality, happiness, quality of life, wellness, and other crucial societal parameters, and does not integrate environmental services into economic decisions (Stiglitz et al., 2010).

More recently, other scholars (Ven, 2015; Fleurbaey, 2015) have called for a new generation of multifaceted and more comprehensive well-being measures, better able to describe actual living standards and useful for a more accurate design of policies improving efficiency in resources assignment.

A series of measures of well-being, inspired at the Nussbaum-Sen approach to human capabilities and subjective well-being (Nussbaum and Sen, 1993), have been proposed in attempt to go beyond GDP with the aim to broaden the scope of effects in the assessment of policies. For instance, the Human Development Index by UNDP or the Better Life Initiative launched by OECD (2015) and many other approaches are based on the income, health, and education measurements of the countries' performance (for a review, see Fleurbaey and Blanchet, 2013).

From a technical view, many methods are employed to measure the well-being level through composite indicators. Nevertheless, no method is universally valid to select indicators based on theory-driven criteria, measure properly the concept, aggregate and normalise a set of input variables and define a weighting system (OECD, 2008).

The aim is to simplify the analysis of the multidimensional concept according with a formative or reflective measurement model, where elementary indicators are causes or effects of latent variables, respectively (Michalos, 2014; Simonetto, 2012).

Only some studies have focused on the construction of well-being composite indexes to evaluate and compare well-being specifically across the Italian provinces. Mazziotta and Pareto (2019) have obtained a global well-being index by aggregating 11 composite indices with AMPI (Mazziotta and Pareto, 2016) and have ranked Italian provinces for each dimension of well-being and have given an overall ranking.

Also Calcagnini and Perugini (2019) have proposed a composite indicator of well-being for the Italian provinces (NUTS-3) based on the methodology of the regional Index of Regional Quality of Development (QUARS) to analyse the extent to which the socio-economic heterogeneity in individual and contextual features within region affects the well-being among adjacent provinces.

In Italy, indicators of well-being have being used more and more for policy-making reasons at national, but also regional or local levels involving public institutions. From a theoretical point of view, the relationship of well-being assessment with policy-making process in healthcare, education, and training, or local services is the rationale to analyse the well-being measures at local level.

Since the policies of local governmental authorities have a direct and huge impact on the socio-economic context, the assessment of living standards at provincial level allows to evaluate economic, environmental, and social needs of the citizens at any level of government in order to implement and design decentralised policies. Following this, for territorial micro-areas the economic dimension could not be so crucial.

3 PLS–SEM-KM for composite indicator building

Partial Least Squares (PLS) methodologies are algorithmic tools with analytic proprieties aiming at solving problems about the stringent assumptions on data, e.g., distributional assumptions that are hard to observe in real life.

Tenenhaus et al. (2005) try to better clarify the terminology used in the PLS field through an interesting review of the literature, focusing the attention on the Structural Equation Models standpoint.

Given the $n \times J$ data matrix \mathbf{X} , the $n \times K$ membership matrix \mathbf{U} , the $K \times J$ centroids matrix \mathbf{C} , the $J \times P$ loadings matrix $\mathbf{\Lambda} = [\mathbf{\Lambda}_H, \mathbf{\Lambda}_L]$, and the errors matrices \mathbf{Z} , \mathbf{E} and \mathbf{D} , the PLS-SEM-KM model simultaneously identify a SEM model for variables and a partitioning KM model for units according to the following three equations:

$$\begin{aligned} \mathbf{H} &= \mathbf{H}\mathbf{B}^T + \mathbf{\Xi}\mathbf{\Gamma}^T + \mathbf{Z}, \\ \mathbf{X} &= \mathbf{Y}\mathbf{\Lambda}^T + \mathbf{E} = \mathbf{\Xi}\mathbf{\Lambda}_H^T + \mathbf{H}\mathbf{\Lambda}_L^T + \mathbf{E}, \\ \mathbf{X} &= \mathbf{U}\mathbf{C}\mathbf{\Lambda}\mathbf{\Lambda}^T = \mathbf{U}\mathbf{C}\mathbf{\Lambda}_H\mathbf{\Lambda}_H^T + \mathbf{U}\mathbf{C}\mathbf{\Lambda}_L\mathbf{\Lambda}_L^T + \mathbf{D} \end{aligned} \quad (1)$$

subject to constraints: (i) $\mathbf{\Lambda}^T\mathbf{\Lambda} = \mathbf{I}$; and (ii) $\mathbf{U} \in \{0,1\}$, $\mathbf{U}\mathbf{1}_K = \mathbf{1}_n$. Thus, the PLS-SEM-KM model includes the PLS-PM and the clustering KM equations (i.e., $\mathbf{X} = \mathbf{U}\mathbf{C}$ and then, $\mathbf{Y} = \mathbf{X}\mathbf{\Lambda}$ becomes $\mathbf{Y} = \mathbf{U}\mathbf{C}\mathbf{\Lambda}$).

The proposed methodology shows some important advantages with respect to the other proposed approaches for both cluster analysis and composite indicator construction: it is a new simultaneous approach.

In fact, identifies the best homogenous partition of the objects, represented by the best statistical relationships among latent and observed variables (Fordellone and Vichi, 2018; 2020).

4 Results: well-being composite indicator

The conceptual structure (Giovannini et al., 2012) of the BES considers 9 domains related to aspects that directly influence well-being, plus 3 instrumental or context domains.

In each well-being domain (ISTAT, 2018), elementary indicators and a synthesis through composite indicators related to each domain are integrated.

Our dataset consists in 109 units (Italian provinces) and 16 indicators organized in 9 different domains. Table 1 shows the relationships among MVs and LVs and the loadings estimated through PLS-SEM-KM.

Through the application of PLS-SEM-KM model, we have identified 3 homogeneous groups of Italian provinces. The optimal number of clusters identified corresponds to the maximum value of the pseudo- F function (around 1.1).

From this analysis, we can see that the theoretical polarity associated to each observed variable (see Table 1) is well described by the measurement PLS approach.

Modeling Well-Being through *PLS-SEM and K-M*

Whereas only the *Overcrowding of prisons* variable shows a non-significant loading, in fact also the sign is not correct.

Table 1: Measurement models matrix estimated through PLS-SEM-KM.

| MVs | LV1 | LV2 | LV3 | LV4 | LV5 | LV6 | LV7 | LV8 | LV9 | WB |
|-------------|-------|--------|--------|--------|-------|--------|-------|-------|-------|--------|
| MV1 | 0.795 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.721 |
| MV2 | 0 | 0.750 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.619 |
| MV3 | 0 | -0.678 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.917 |
| MV4 | 0 | 0 | -0.876 | 0 | 0 | 0 | 0 | 0 | 0 | -0.938 |
| MV5 | 0 | 0 | -0.887 | 0 | 0 | 0 | 0 | 0 | 0 | -0.881 |
| MV6 | 0 | 0 | 0.920 | 0 | 0 | 0 | 0 | 0 | 0 | 0.870 |
| MV7 | 0 | 0 | -0.917 | 0 | 0 | 0 | 0 | 0 | 0 | -0.914 |
| MV8 | 0 | 0 | 0 | -0.669 | 0 | 0 | 0 | 0 | 0 | -0.569 |
| MV9 | 0 | 0 | 0 | 0 | 0.050 | 0 | 0 | 0 | 0 | 0.202 |
| MV10 | 0 | 0 | 0 | 0 | 0 | -0.373 | 0 | 0 | 0 | -0.417 |
| MV11 | 0 | 0 | 0 | 0 | 0 | -0.216 | 0 | 0 | 0 | -0.140 |
| MV12 | 0 | 0 | 0 | 0 | 0 | -0.434 | 0 | 0 | 0 | -0.412 |
| MV13 | 0 | 0 | 0 | 0 | 0 | 0 | 0.105 | 0 | 0 | 0.226 |
| MV14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.277 | 0 | -0.088 |
| MV15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.691 | 0 | 0.704 |
| MV16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.748 | 0.822 |

In Table 2 the path coefficients' matrix of the estimated structural-PLS model is shown.

Table 2: Structural model matrix estimated by PLS-SEM-KM.

| | LV1 | LV2 | LV3 | LV4 | LV5 | LV6 | LV7 | LV8 | LV9 | WB |
|---------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| LV1-Health | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.324 |
| LV2-Education and training | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.053 |
| LV3-Work-life balance | | | | 0 | 0 | 0 | 0 | 0 | 0 | 0.666 |
| LV4-Economic well-being | | | | | 0 | 0 | 0 | 0 | 0 | 0.247 |
| LV5-Policy | | | | | | 0 | 0 | 0 | 0 | 0.117 |
| LV6-Safety | | | | | | | 0 | 0 | 0 | 0.163 |
| LV7-Cultural heritage | | | | | | | | 0 | 0 | 0.067 |
| LV8-Environment | | | | | | | | | 0 | 0.192 |
| LV9-Innovation and research | | | | | | | | | | 0.198 |
| Well-being composite indicator | | | | | | | | | | 0 |

From the structural model we can see that the Italian provincial well-being is highly affected by the *LV3-Work-life balance* construct (0.666), followed by *LV1-Health* (0.324) and *LV4-Economic well-being* (0.247) constructs. Whereas, the *LV6-Cultural heritage* (0.067) and *LV2-Education* (0.053) constructs are

Venera Tomaselli, Mario Fordellone, Maurizio Vichi

dimensions with non-significant coefficients. The overall fit of the structural model is good with a $R^2 = 0.74$. In terms of clustering results, the three clusters identified by the PLS-SEM-KM algorithm describe three groups of well-being, i.e., *high*, *medium*, and *low* levels. In this way is very easy to classify the Italian provinces through a “BES ranking”.

5 Conclusions

The BES 2018 report by ISTAT has confirmed that in 2015 and in 2016 an improvement in many areas of well-being has been observed, even if territorial differences remain stable both in levels and dynamics.

The differences appear in some cases as real structural differences between North and South of Italy. Then, the different well-being levels measured through a composite well-being indicator in the territories allow to explain the variability in territorial micro areas.

References

1. Diener, E., Lucas, R., Schimmack, U., & Helliwell, J. (2009). Well-Being for public policy. Oxford, UK: Oxford University Press. doi.org/10.1093/acprof:oso/9780195334074.001.0001.
2. Fleurbaey, M. (2015). Beyond income and wealth. *Rev. Income Wealth*, 61(2), 199-219.
3. Fleurbaey, M., & Blanchet, D. (2013). Beyond GDP: Measuring welfare and assessing sustainability. Oxford University Press.
4. Fordellone, M., & Vichi, M. (2018). Structural Equation Modeling and simultaneous clustering through the Partial Least Squares algorithm. arXiv preprint: arXiv:1810.07677v1.
5. Fordellone, M., & Vichi, M. (2020). Finding groups in structural equation modeling through the partial least squares algorithm. *Comput Stat Data An* doi: <https://doi.org/10.1016/j.csda.2020.106957>
6. Giovannini, E., Morrone, A., Rondinella, T., & Sabbadini, L. L. (2012). L'iniziativa CNEL-ISTAT per la misurazione del Benessere Equo e Sostenibile in Italia. *Autonomie locali e servizi sociali*, 1, 125-136.
7. Helliwell, J. F., Layard, R., & Sachs, J. (2012). Some policy implications. In J. F. Helliwell, R. Layard & J. Sachs (Eds.) *World Happiness Report* (pp. 90-107), New York, NY: The Earth Institute, Columbia University
8. ISTAT (2018). BES 2018: Il benessere equo e sostenibile in Italia. Istituto Nazionale di Statistica.
9. Jayawickreme, E., Forgeard, M. J., & Seligman, M. E. (2012). The engine of well-being. *Rev Gen Psychol*, 16(4), 327-342. doi.org/10.1037/a0027990.
10. Mazziotta, M., & Pareto, A. (2016). On a generalized non-compensatory composite index for measuring socio-economic phenomena. *Soc Indic Res*, 127(3), 983-1003.
11. Mazziotta, M., & Pareto, A. (2019). Use and misuse of PCA for measuring well-being. *Soc Indic Res*, 142(2), 451-476.
12. Nussbaum, M. & Sen, A. (1993). *The Quality of Life*. Oxford: Oxford University Press.
13. OECD (2015). *How's life? 2015 Measuring well-being*. Paris, FR: OECD Publishing. (Available online: <http://dx.doi.org/10.1787/9789264121164-en>).
14. Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2010). *Mismeasuring our lives: why GDP doesn't add up*. New York, NJ: The New Press.
15. Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y. M., & Lauro, N. C. (2005). PLS path modelling. *Comput Stat Data An*, 48(1), 159-205.
16. Ven, P. (2015). Introduction to the symposium on new measures of well-being: Perspectives from statistical offices. *Rev Income Wealth*, 61(1), 1-3.