Book of Short Papers SIS 2018

Editors: Antonino Abbruzzo - Eugenio Brentari

Marcello Chiodi - Davide Piacentino

Copyright © 2018

PUBLISHED BY PEARSON

WWW.PEARSON.COM

First printing, November 2018, ISBN-9788891910233

SIS 2018
49 TH SCIENTIFIC MEETING OF THE ITALIAN STATISTICAL SOCIETY
PALERMO 20-22 JUNE

Contents

1 Preface 17

2 Plenary Sessions 19

2.1 A new paradigm for rating data models. *Domenico Piccolo* 19

2.2 Statistical challenges and opportunities in modelling coupled behaviour-disease dynamics of vaccine refusal. *Plenary/Chris T. Bauch* 32

3 Specialized Sessions 45

3.1 3.1 - Bayesian Nonparametric Learning 45

3.1.1 Bayesian nonparametric covariate driven clustering. *Raffaele Argiento, Ilaria Bianchini, Alessandra Guglielmi and Ettore Lanzarone* 46

3.1.2 A Comparative overview of Bayesian nonparametric estimation of the size of a population. *Luca Tardella and Danilo Alunni Fegatelli* 56

3.1.3 Logit stick-breaking priors for partially exchangeable count data. *Tommaso Rigon* 64

3.2 BDsports - Statistics in Sports 72

3.2.1 A paired comparison model for the analysis of on-field variables in football matches. *Gunther Schauberger and Andreas Groll* 72

3.2.2 Are the shots predictive for the football results?. *Leonardo Egidi, Francesco Pauli, Nicola Torelli* 81

3.2.3 Zero-inflated ordinal data models with application to sport (in)activity. *Maria Iannario and Rosaria Simone* 89

3.3 Being young and becoming adult in the Third Millennium: definition issues and processes analysis 97

3.3.1 Do Social Media Data predict changes in young adults' employment status? Evidence from Italy. *Andrea Bonanomi and Emiliano Sironi* 97

3.3.2 Parenthood: an advanced step in the transition to adulthood. *Cinzia Castagnaro, Antonella Guarneri and Eleonora Meli* 106

3.4 Economic Statistics and Big Data 114

3.4.1 Improvements in Italian CPI/HICP deriving from the use of scanner data. *Alessandro Brunetti, Stefania Fatello, Federico Polidoro, Antonella Simone* 114

3.4.2 Big data and spatial price comparisons of consumer prices. *Tiziana Laureti and Federico Polidoro* 123

3.5 Financial Time Series Analysis 131

3.5.1 Dynamic component models for forecasting trading volumes. *Antonio Naimoli and Giuseppe Storti* 131

3.5.2 Conditional Quantile-Located VaR. *Giovanni Bonaccolto, Massimiliano Caporin and Sandra Paterlini* 140

3.6 Forensic Statistics 146

3.6.1 Cause of effects: an important evaluation in Forensic Science. *Fabio Corradi and Monica Musio* 146

3.6.2 Evaluation and reporting of scientific evidence: the impact of partial probability assignments. *Silvia Bozza, Alex Biedermann, Franco Taroni* 155

3.7 Missing Data Handling in Complex Models 161

3.7.1 Dependence and sensitivity in regression models for longitudinal responses subject to dropout. *Marco Alfo' and Maria Francesca Marino* 161

3.7.2 Multilevel analysis of student ratings with missing level-two covariates: a comparison of imputation techniques. *Maria Francesca Marino e Carla Rampichini* 170

3.7.3 Multilevel Multiple Imputation in presence of interactions, non-linearities and random slopes. *Matteo Quartagno and James R. Carpenter* 175

3.8 Monitoring Education Systems. Insights from Large Scale Assessment Surveys 183

3.8.1 Educational Achievement of Immigrant Students. A Cross-National Comparison Over-Time Using OECD-PISA Data. *Mariano Porcu* 183

3.9 New Perspectives in Time Series Analysis 192

3.9.1 Generalized periodic autoregressive models for trend and seasonality varying time series. *Francesco Battaglia and Domenico Cucina and Manuel Rizzo* 192

3.10 Recent Advances in Model-based Clustering 201

3.10.1 Flexible clustering methods for high-dimensional data sets. *Cristina Tortora and Paul D. McNicholas* 201

3.10.2 A Comparison of Model-Based and Fuzzy Clustering Methods. *Marco Alfo', Maria Brigida Ferraro, Paolo Giordani, Luca Scrucca, and Alessio Serafini* 208

3.10.3 Covariate measurement error in generalized linear models for longitudinal data: a latent Markov approach. *Roberto Di Mari, Antonio Punzo, and Antonello Maruotti* 216

3.11 Statistical Modelling 224

3.11.1 A regularized estimation approach for the three-parameter logistic model. *Michela Battauz and Ruggero Bellio* 224

3.11.2 Statistical modelling and GAMLSS. *Mikis D. Stasinopoulos and Robert A. Rigby and Fernanda De Bastiani* 233

3.12 Young Contributions to Statistical Learning 239

3.12.1 Introducing spatio-temporal dependence in clustering: from a parametric to a nonparametric approach . *Clara Grazian, Gianluca Mastrantonio and Enrico Bibbona* 239

3.12.2 Bayesian inference for hidden Markov models via duality and approximate filtering distributions. *Guillaume Kon Kam King, Omiros Papaspiliopoulos and Matteo Ruggiero* 248
3.12.3 K-means seeding via MUS algorithm. *Leonardo Egidi, Roberta Pappada`, Francesco Pauli, Nicola Torelli* 256

4 Sollicited Sessions 263

4.1 Advances in Discrete Latent Variable Modelling 263
4.1.1 A joint model for longitudinal and survival data based on a continuous-time latent Markov model. *Alessio Farcomeni and Francesco Bartolucci* 264
4.1.2 Modelling the latent class structure of multiple Likert items: a paired comparison approach. *Brian Francis* 273
4.1.3 Dealing with reciprocity in dynamic stochastic block models. *Francesco Bartolucci, Maria Francesca Marino, Silvia Pandolfi* 281
4.1.4 Causality patterns of a marketing campaign conducted over time: evidence from the latent Markov model. *Fulvia Pennoni, Leo Paas and Francesco Bartolucci* 289

4.2 Complex Spatio-temporal Processes and Functional Data 297
4.2.1 Clustering of spatio-temporal data based on marked variograms. *Antonio Balzanella and Rosanna Verde* 297
4.2.2 Space-time earthquake clustering: nearest-neighbor and stochastic declustering methods in comparison. *Elisa Varini, Antonella Peresan, Renata Rotondi, and Stefania Gentili* 304
4.2.3 Advanced spatio-temporal point processes for the Sicily seismicity analysis. *Marianna Siino and Giada Adelfio* 312
4.2.4 Spatial analysis of the Italian seismic network and seismicity. *Antonino D'Alessandro, Marianna Siino, Luca Greco and Giada Adelfio* 320

4.3 Dimensional Reduction Techniques for Big Data Analysis 328
4.3.1 Clustering Data Streams via Functional Data Analysis: a Comparison between Hierarchical Clustering and K-means Approaches. *Fabrizio Maturo, Francesca Fortuna, and Tonio Di Battista* 328
4.3.2 Co-clustering algorithms for histogram data. *Francisco de A.T. De Carvalho and Antonio Balzanella and Antonio Irpino and Rosanna Verde* 338
4.3.3 A network approach to dimensionality reduction in Text Mining. *Michelangelo Misuraca, Germana Scepi and Maria Spano* 344
4.3.4 Self Organizing Maps for distributional data. *Rosanna Verde and Antonio Irpino* 352

4.4 Enviromental Processes, Human Activities and their Interactions 353
4.4.1 Estimation of coral growth parameters via Bayesian hierarchical non-linear models. *Crescenza Calculli, Barbara Cafarelli and Daniela Cocchi* 353
4.4.2 A Hierarchical Bayesian Spatio-Temporal Model to Estimate the Short-term Effects of Air Pollution on Human Health. *Fontanella Lara, Ippoliti Luigi and Valentini Pasquale* 361
4.4.3 A multilevel hidden Markov model for space-time cylindrical data. *Francesco Lagona and Monia Ranalli* 367
4.4.4 Estimation of entropy measures for categorical variables with spatial correlation. *Linda Altieri, Giulia Roli* 373

4.5 Innovations in Census and in Social Surveys 381
4.5.1 A micro-based approach to ensure consistency among administrative sources and to improve population statistics. *Gianni Corsetti, Sabrina Prati, Valeria Tomeo, Enrico Tucci* 381
4.5.2 Demographic changes, research questions and data needs: issues about migrations. *Salvatore Strozza and Giuseppe Gabrielli* 392

4.5.3 Towards more timely census statistics: the new Italian multiannual dissemination programme. *Simona Mastroluca and Mariangela Verrascina* 400

4.6 Living Conditions and Consumption Expenditure in Time of Crises 409

4.6.1 Household consumption expenditure and material deprivation in Italy during last economic crises. *Ilaria Arigoni and Isabella Siciliani* 409

4.7 Network Data Analysis and Mining 418

4.7.1 Support provided by elderly Italian people: a multilevel analysis. *Elvira Pelle, Giulia Rivellini and Susanna Zaccarini* 418

4.7.2 Data mining and analysis of comorbidity networks from practitioner prescriptions. *Giancarlo Ragozini, Giuseppe Giordano, Sergio Pagano, Mario De Santis, Pierpaolo Cavallo* 426

4.7.3 Overlapping mixture models for network data (manet) with covariates adjustment. *Saverio Ranciati and Giuliano Galimberti and Ernst C. Wit and Veronica Vinciotti* 434

4.8 New Challenges in the Measurement of Economic Insecurity, Inequality and Poverty 440

4.8.1 Social protection in mitigating economic insecurity. *Alessandra Coli* 440

4.8.2 Changes in poverty concentration in U.S. urban areas. *Francesco Andreoli and Mauro Mussini* 450

4.8.3 Evaluating sustainability through an input-stateoutput framework: the case of the Italian provinces. *Achille Lemmi, Laura Neri, Federico M. Pulselli* 458

4.9 New Methods and Models for Ordinal Data 466

4.9.1 Weighted and unweighted distances based decision tree for ranking data. *Antonella Plaia, Simona Buscemi, Mariangela Sciandra* 466

4.9.2 A dissimilarity-based splitting criterion for CUBREMOT. *Carmela Cappelli, Rosaria Simone and Francesca Di Iorio* 474

4.9.3 Constrained Extended Plackett-Luce model for the analysis of preference rankings. *Cristina Mollica and Luca Tardella* 480

4.9.4 A prototype for the analysis of time use in Italy. *Stefania Capecchi and Manuela Michelini* 487

4.10 New Perspectives in Supervised and Unsupervised Classification 493

4.10.1 Robust Updating Classification Rule with applications in Food Authenticity Studies. *Andrea Cappozzo, Francesca Greselin and Thomas Brendan Murphy* 493

4.10.2 A robust clustering procedure with unknown number of clusters. *Francesco Dotto and Alessio Farcomeni* 500

4.10.3 Issues in joint dimension reduction and clustering methods. *Michel van de Velden, Alfonso Iodice D'Enza and Angelos Markos* 508

4.11 New Sources, Data Integration and Measurement Challenges for Estimates on Labour Market Dynamics 514

4.11.1 The development of the Italian Labour register: principles, issues and perspectives . *C. Baldi, C. Ceccarelli, S. Gigante, S. Pacini* 514

4.11.2 Digging into labour market dynamics: toward a reconciliation of stock and flows short term indicators. *F. Rapiti, C. Baldi, D. Ichim, F. Pintaldi, M. E. Pontecorvo, R. Rizzi* 523

4.11.3 How effective are the regional policies in Europe? The role of European Funds. *Gennaro Punzo, Mariateresa Ciommi, and Gaetano Musella* 531

4.11.4 Labour market condition in Italy during and after the financial crises: a segmented regression analysis approach of interrupted time series. *Lucio Masserini and Matilde Bini* 539

4.12 Quantile and Generalized Quantile Methods 547
4.12.1 Multiple quantile regression for risk assessment. *Lea Petrella and Valentina Raponi* 547
4.12.2 Parametric Modeling of Quantile Regression Coefficient Functions. *Paolo Frumento and Matteo Bottai* 550
4.12.3 Modelling the effect of Traffic and Meteorology on Air Pollution with Finite Mixtures of M-quantile Regression Models. *Simone Del Sarto, Maria Francesca Marino, Maria Giovanna Ranalli and Nicola Salvati* 552
4.12.4 Three-level M-quantile model for small area poverty mapping. *Stefano Marchetti and Nicola Salvati* 560
4.13 Recent Advances on Extreme Value Theory 560
4.13.1 Extremes of high-order IGARCH processes. *Fabrizio Laurini* 560
4.14 Spatial Economic Data Analysis 569
4.14.1 Spatial heterogeneity in principal component analysis: a study of deprivation index on Italian provinces. *Paolo Postiglione, M. Simona Andreano, Roberto Benedetti, Alfredo Cartone* 569
4.15 Spatial Functional Data Analysis 578
4.15.1 Object oriented spatial statistics for georeferenced tensor data. *Alessandra Menafoglio and Davide Pigoli and Piercesare Secchi* 578
4.15.2 A Spatio-Temporal Mixture Model for Urban Crimes. *Ferretti Angela, Ippoliti Luigi and Valentini Pasquale* 585
4.16 Statistical Methods for Service Quality 591
4.16.1 Cumulative chi-squared statistics for the service quality improvement: new properties and tools for the evaluation. *Antonello D'Ambra, Antonio Lucadamo, Pietro Amenta, Luigi D'Ambra* 591
4.16.2 A robust multinomial logit model for evaluating judges' performances. *Ida Camminatiello and Antonio Lucadamo* 600
4.16.3 Complex Contingency Tables and Partitioning of Three-way Association Indices for Assessing Justice CourtWorkload. *Rosaria Lombardo, Yoshio Takane and Eric J Beh* 607
4.16.4 Finding the best paths in university curricula of graduates to improve academic guidance services. *Silvia Bacci and Bruno Bertaccini* 615
4.17 Statistical Modelling for Business Intelligence Problems 623
4.17.1 A nonlinear state-space model for the forecasting of field failures. *Antonio Pievatolo* 623
4.17.2 Does Airbnb affect the real estate market? A spatial dependence analysis. *Mariangela Guidolin and Mauro Bernardi* 632
4.17.3 Bayesian Quantile Trees for Sales Management. *Mauro Bernardi and Paola Stolfi* 640
4.17.4 Discrimination in machine learning algorithms. *Roberta Pappadá and Francesco Pauli* 648
4.18 Statistical models for sports data 656
4.18.1 The study of relationship between financial performance and points achieved by Italian football championship clubs via GEE and diagnostic measures. *Anna Crisci, Sarnacchiaro Pasquale e Luigi D?Ambra* 656
4.18.2 Exploring the Kaggle European Soccer database with Bayesian Networks: the case of the Italian League Serie A. *Maurizio Carpita and Silvia Golia* 665
4.18.3 A data-mining approach to the Parkour discipline. *Paola Pasca, Enrico Ciavolino and Ryan L. Boyd* 673
4.18.4 Players Movements and Team Shooting Performance: a Data Mining approach for Basketball. *Rodolfo Metulini* 681

4.19 Supporting Regional Policies through Small Area Statistical Methods 689
4.19.1 Survey-weighted Unit-Level Small Area Estimation. *Jan Pablo Burgard and Patricia Dörr* 689
4.19.2 Robust and model-assisted small area estimation methods: an application to the Banca d'Italia Survey of Industrial and Service Firms. *Bottone Marco, Casciano Maria Cristina, Fabrizi Enrico, Filiberti Salvatore, Neri Andrea and Salvati Nicola* 698
4.20 The Second Generation at School 706
4.20.1 Resilient students with migratory background. *Anna Di Bartolomeo and Giuseppe Gabrielli* 706
4.20.2 Residential Proximity to Attended Schools among Immigrant-Origin Youths in Bologna. *Federica Santangelo, Debora Mantovani and Giancarlo Gasperoni* 715
4.20.3 From school to ... future: strategies, paths and perspectives of immigrant immediate descendants in Naples . *Giustina Orientale Caputo and Giuseppe Gargiulo* 723
4.21 Tourism Destinations, Household, Firms 731
4.21.1 The Pricing Behaviour of Firms in the On-line Accommodation Market: Evidence from a Metropolitan City. *Andrea Guizzardi and Flavio Maria Emanuele Pons* 731
4.21.2 The Migration-Led-Tourism Hypothesis for Italy: A Survey. *Carla Massidda, Romano Piras and Ivan Etzo* 741
4.21.3 Tourism Statistics: development and potential uses. *Fabrizio Antolini* 749
4.21.4 Tourism attractiveness in Italy. Some empirical evidence comparing origin-destination domestic tourism flows. *Francesca Giambona, Emanuela Dreassi, and Alessandro Magrini* 757
4.22 What's Happening in Africa 765
4.22.1 Environmental shocks and internal migration in Tanzania. *Maria Francesca Marino, Alessandra Petrucci, and Elena Pirani* 765
4.22.2 Determinants and geographical disparities of BMI in African Countries: a measurement error small area approach. *Serena Arima and Silvia Polettini* 773

5 Contributed Sessions 781
5.1 Advanced Algorithms and Computation 781
5.1.1 Brexit in Italy. *Francesca Greco, Livia Celardo, Leonardo Salvatore Alaimo* . 782
5.1.2 Distance based Depth-Depth classifier for directional data. *Giuseppe Pandolfo and Giovanni C. Porzio* 789
5.1.3 Approximate Bayesian Computation for Forecasting in Hydrological models. *Jonathan Romero-Cuéllar, Antonino Abbruzzo, Giada Adelfio and Félix Francés* 793
5.1.4 Customer Churn prediction based on eXtreme Gradient Boosting classifier. *Mohammed Hassan Elbedawi Omar and Matteo Borrotti* 799
5.1.5 HPC-accelerated Approximate Bayesian Computation for Biological Science. *Ritabrata Dutta* 805
5.1.6 PC Algorithm for Gaussian Copula Data. *Vincenzina Vitale and Paola Vicard* 813
5.2 Advances in Clustering Techniques 819
5.2.1 On the choice of an appropriate bandwidth for modal clustering. *Alessandro Casa, José E. Chacón and Giovanna Menardi* 819
5.2.2 Unsupervised clustering of Italian schools via non-parametric multilevel models. *Chiara Masci, Francesca Ieva and Anna Maria Paganoni* 826
5.2.3 Chiara Masci, Francesca Ieva and Anna Maria Paganoni. *Laura Bocci and Donatella Vicari* 832
5.2.4 Robust Reduced k-Means and Factorial k-Means by trimming. *Luca Greco and Antonio Lucadamo and Pietro Amenta* 837
5.2.5 Dirichlet processes, posterior similarity and graph clustering. *Stefano Tonellato* 843

5.2.6 Bootstrap ClustGeo with spatial constraints. *Veronica Distefano, Valentina Mameli, Fabio Della Marra* 849

5.3 Advances in Statistical Models 855

5.3.1 Regression modeling via latent predictors. *Francesca Martella and Donatella Vicari* 855

5.3.2 Analysis of dropout in engineering BSc using logistic mixed-effect models. *Luca Fontana and Anna Maria Paganoni* 862

5.3.3 dgLARS method for relative risk regression models. *Luigi Augugliaro and Angelo M. Mineo* 868

5.3.4 A Latent Class Conjoint Analysis for analysing graduates profiles. *Paolo Mariani, Andrea Marletta, Lucio Masserini and Mariangela Zenga* 874

5.3.5 A longitudinal analysis of the degree of accomplishment of anti-corruption measures by Italian municipalities: a latent Markov approach. *Simone Del Sarto, Michela Gnaldi, Francesco Bartolucci* 880

5.3.6 Modelling the effect of covariates for unbiased estimates in ecological inference methods. *Venera Tomaselli, Antonio Forcina and Michela Gnaldi* 886

5.4 Advances in Time Series 892

5.4.1 Filtering outliers in time series of electricity prices. *Ilaria Lucrezia Amerise* . . . 892

5.4.2 Time-varying long-memory processes. *Luisa Bisaglia and Matteo Grigoletto* 899

5.4.3 Statistical Analysis of Markov Switching DSGE Models. *Maddalena Cavicchioli* 905

5.4.4 Forecasting energy price volatilities and comovements with fractionally integrated MGARCH models. *Malvina Marchese and Francesca Di Iorio* 910

5.4.5 Improved bootstrap simultaneous prediction limits. *Paolo Vidoni* 916

5.5 Data Management 922

5.5.1 Using web scraping techniques to derive co-authorship data: insights from a case study. *Domenico De Stefano, Vittorio Fuccella, Maria Prosperina Vitale, Susanna Zaccarin* 922

5.5.2 Dealing with Data Evolution and Data Integration: An approach using Rarefaction. *Luca Del Core, Eugenio Montini, Clelia Di Serio, Andrea Calabria* 929

5.5.3 Monitoring event attendance using a combination of traditional and advanced surveying tools. *Mauro Ferrante, Amit Birenboim, Anna Maria Milito, Stefano De Cantis* 935

5.5.4 Indefinite Topological Kernels. *Tullia Padellini and Pierpaolo Brutti* 941

5.5.5 Data Integration in Social Sciences: the earnings intergenerational mobility problem. *Veronica Ballerini, Francesco Bloise, Dario Briscolini and Michele Raitano* 947

5.5.6 An innovative approach for the GDPR compliance in Big DAta era. *M. Giacalone, C. Cusatelli, F. Fanari, V. Santarcangelo, D.C. Sinitó* 953

5.6 Developments in Graphical Models 959

5.6.1 An extension of the glasso estimator to multivariate censored data. *Antonino Abbruzzo and Luigi Augugliaro and Angelo M. Mineo* 959

5.6.2 Bayesian Estimation of Graphical Log-Linear Marginal Models. *Claudia Tarantola, Ioannis Ntzoufras and Monia Lupparelli* 966

5.6.3 Statistical matching by Bayesian Networks. *Daniela Marella and Paola Vicard and Vincenzina Vitale* 972

5.6.4 Sparse Nonparametric Dynamic Graphical Models. *Fabrizio Poggioni, Mauro Bernardi, Lea Petrella* 978

5.6.5 Non-communicable diseases, socio-economic status, lifestyle and well-being in Italy: An additive Bayesian network model. *Laura Maniscalco and Domenica Matranga* 984

5.6.6 Using Almost-Dynamic Bayesian Networks to Represent Uncertainty in Complex Epidemiological Models: a Proposal. *Sabina Marchetti* 990

5.7 Educational World **996**

5.7.1 How to improve the Quality Assurance System of the Universities: a study based on compositional analysis . *Bertaccini B., Gallo M., Simonacci V., and Menini T.* . 996

5.7.2 Evaluation of students' performance at lower secondary education. An empirical analysis using TIMSS and PISA data.. *G. Graziosi, T. Agasisti, K. De Witte and F. Pauli* 1001

5.7.3 Testing for the Presence of Scale Drift: An Example. *Michela Battauz* 1007

5.7.4 The evaluation of Formative Tutoring at the University of Padova. *Renata Clerici, Lorenza Da Re, Anna Giraldo, Silvia Meggiolaro* 1012

5.7.5 Benefits of the Erasmus mobility experience: a discrete latent variable analysis. *Silvia Bacci, Valeria Caviezel and Anna Maria Falzoni* 1017

5.7.6 University choice and the attractiveness of the study area. Insights from an analysis based on generalized mixed-effect models. *Silvia Columbu, Mariano Porcu and Isabella Sulis* 1023

5.8 Environment **1029**

5.8.1 The climate funds for energy sustainability: a counterfactual analysis. *Alfonso Carfora and Giuseppe Scandurra* 1029

5.8.2 Exploratory GIS Analysis via Spatially Weighted Regression Trees. *Carmela Iorio, Giuseppe Pandolfo, Michele Staiano, and Roberta Siciliano* 1036

5.8.3 A functional regression control chart for profile monitoring. *Fabio Centofanti, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini* 1042

5.8.4 Understanding pro-environmental travel behaviours in Western Europe. *Gennaro Punzo, Rosalia Castellano, and Demetrio Panarello* 1047

5.9 Family & Economic issues **1053**

5.9.1 Measuring Economic Uncertainty: Longitudinal Evidence Using a Latent Transition Model. *Francesca Giambona, Laura Grassini and Daniele Vignoli* 1053

5.9.2 Intentions to leave Italy or to stay among foreigners: some determinants of migration projects. *Ginevra Di Giorgio, Francesca Dota, Paola Muccitelli and Daniele Spizzichino* 1060

5.9.3 Wages differentials in association with individuals, enterprises and territorial characteristics. *S. De Santis, C. Freguja, A. Masi, N. Pannuzi, F. G. Truglia* 1066

5.9.4 The Transition to Motherhood among British Young Women: Does housing tenure play a role?. *Valentina Tocchioni, Ann Berrington, Daniele Vignoli and Agnese Vitali* 1072

5.10 Finance & Insurance **1078**

5.10.1 Robust statistical methods for credit risk. *A. Corbellini, A. Ghiretti, G. Morelli and A. Talignani* 1078

5.10.2 Depth-based portfolio selection. *Giuseppe Pandolfo, Carmela Iorio and Antonio D'Ambrosio* 1085

5.10.3 Estimating large-scale multivariate local level models with application to stochastic volatility. *Matteo Pelagatti and Giacomo Sbrana* 1091

5.11 Health and Clinical Data **1097**

5.11.1 Is retirement bad for health? A matching approach. *Elena Pirani, Marina Ballerini, Alessandra Mattei, Gustavo De Santis* 1097

5.11.2 The emergency department utilisation among the immigrant population resident in Rome from 2005 to 2015. *Eleonora Trappolini, Laura Cacciani, Claudia Marino, Cristina Giudici, Nera Agabiti, Marina Davoli* 1104

5.11.3 Multi-State model with nonparametric discrete frailty. *Francesca Gasperoni, Francesca Ieva, Anna Maria Paganoni, Chris Jackson and Linda Sharples* 1111

5.11.4 A Functional Urn Model for CARA Designs. *Giacomo Aletti, Andrea Ghiglietti, and William F. Rosenberger* 1117

5.11.5 Assessment of the INLA approach on gerarchic bayesian models for the spatial disease distribution: a real data application. *Paolo Girardi, Emanuela Bovo, Carmen Stocco, Susanna Baracco, Alberto Rosano, Daniele Monetti, Silvia Rizzato, Sara Zamberlan, Enrico Chinellato, Ugo Fedeli, Massimo Rugge* 1123

5.12 Medicine 1129

5.12.1 Hidden Markov Models for disease progression. *Andrea Martino, Andrea Ghiglietti, Giuseppina Guatteri, Anna Maria Paganoni* 1129

5.12.2 A simulation study on the use of response-adaptive randomized designs. *Anna Maria Paganoni, Andrea Ghiglietti, Maria Giovanna Scarale, Rosalba Miceli, Francesca Ieva, Luigi Mariani, Cecilia Gavazzi and Valeria Edefonti* 1136

5.12.3 The relationship between health care expenditures and time to death: focus on myocardial infarction patients. *Luca Grassetti and Laura Rizzi* 1142

5.12.4 A multivariate extension of the joint models. *Marcella Mazzoleni and Mariangela Zenga* 1148

5.12.5 Multipurpose optimal designs for hypothesis testing in normal response trials. *Marco Novelli and Maroussa Zagoraiou* 1154

5.12.6 Additive Bayesian networks for an epidemiological analysis of swine diseases. *Marta Pittavino and Reinhard Furrer* 1160

5.13 Population Dynamics 1166

5.13.1 Employment Uncertainty and Fertility: a Meta-Analysis of European Research Findings. *Giammarco Alderotti, Daniele Vignoli and Michela Baccini* 1166

5.13.2 What Shapes Population Age Structures in the Long Run. *Gustavo De Santis and Giambattista Salinari* 1172

5.13.3 The impact of economic development on fertility: a complexity approach in a cross-country analysis. *NiccolóInnocenti, Daniele Vignoli and Luciana Lazzeretti* 1178

5.13.4 A Probabilistic Cohort-Component Model for Population Fore-casting - The Case of Germany. *Patrizio Vanella and Philipp Deschermeier* 1183

5.13.5 Mortality trends in Sardinia 1992-2015: an ecological study. *Vanessa Santos Sanchez, Gabriele Ruiu Marco Breschi, Lucia Pozzi* 1187

5.14 Recent Developments in Bayesian Inference 1193

5.14.1 Posterior distributions with non explicit objective priors. *Erlis Ruli, Nicola Sartori and Laura Ventura* 1193

5.14.2 A predictive measure of the additional loss of a non-optimal action under multiple priors. *Fulvio De Santis and Stefania Gubbiotti* 1200

5.14.3 Bayesian estimation of number and position of knots in regression splines. *Gioia Di Credico, Francesco Pauli and Nicola Torelli* 1206

5.14.4 The importance of historical linkages in shaping population density across space. *Ilenia Epifani and Rosella Nicolini* 1212

5.15 Recent Developments in Sampling 1218

5.15.1 Species richness estimation exploiting purposive lists: A proposal. *A. Chiarucci, R.M. Di Biase, L. Fattorini, M. Marcheselli and C. Pisani* 1218

5.15.2 Design-based exploitation of big data by a doubly calibrated estimator. *Maria Michela Dickson, Giuseppe Espa and Lorenzo Fattorini* 1225

5.15.3 Design-based mapping in environmental surveys. *L. Fattorini, M. Marcheselli and C. Pisani* 1231

5.15.4 Testing for independence in analytic inference. *Pier Luigi Conti and Alberto Di Iorio* 1237

5.15.5 On the aberrations of two-level Orthogonal Arrays with removed runs. *Roberto Fontana and Fabio Rapallo* 1243

5.16 Recent Developments in Statistical Modelling 1249

5.16.1 Quantile Regression Coefficients Modeling: a Penalized Approach. *Gianluca Sottile, Paolo Frumento and Matteo Bottai* 1249

5.16.2 Simultaneous calibrated prediction intervals for time series. *Giovanni Fonseca, Federica Giummolé and Paolo Vidoni* 1256

5.16.3 Reversibility and (non)linearity in time series. *Luisa Bisaglia and Margherita Gerolimetto* 1262

5.16.4 Heterogeneous effects of subsidies on farms' performance: a spatial quantile regression analysis. *Marusca De Castris and Daniele Di Gennaro* 1268

5.16.5 On the estimation of high-dimensional regression models with binary covariates. *Valentina Mameli, Debora Slanzi and Irene Poli* 1275

5.17 Social Indicators 1281

5.17.1 Can a neighbour region influence poverty? A fuzzy and longitudinal approach. *Gianni Betti, Federico Crescenzi and Francesca Gagliardi* 1281

5.17.2 Weight-based discrimination in the Italian Labor Market: how do ethnicity and gender interact? *Giovanni Busetta, Maria Gabriella Campolo, and Demetrio Panarello* 1288

5.17.3 The Total Factor Productivity Index as a Ratio of Price Indexes. *Lisa Crosato and Biancamaria Zavanella* 1294

5.17.4 Monetary poverty indicators at local level: evaluating the impact of different poverty thresholds. *Luigi Biggeri, Caterina Giusti and Stefano Marchetti* . . . 1300

5.17.5 A gender inequality assessment by means of the Gini index decomposition. *Michele Costa* 1306

5.18 Socio-Economic Statistics 1312

5.18.1 The NEETs during the economic crisis in Italy, Young NEETs in Italy, Spain and Greece during the economic crisis. *Giovanni De Luca, Paolo Mazzocchi, Claudio Quintano, Antonella Rocca* 1312

5.18.2 Camel or dromedary? A study of the equilibrium distribution of income in the EU countries. *Crosato L., Ferretti C., Ganugi P.* 1319

5.18.3 Small Area Estimation of Inequality Measures. *Maria Rosaria Ferrante and Silvia Pacei* 1325

5.18.4 Testing the Learning-by-Exporting at Micro-Level in light of influence of "Statistical Issues" and Macroeconomic Factors. *Maria Rosaria Ferrante and Marzia Freo* 1330

5.18.5 The mobility and the job success of the Sicilian graduates *Ornella Giambalvo and Antonella Plaia and Sara Binassi* 1336

5.19 Statistical Analysis of Energy Markets 1342

5.19.1 Forecasting Value-at-Risk for Model Risk Analysis in Energy Markets. *Angelica Gianfreda and Giacomo Scandolo* 1342

5.19.2 Prediction interval of electricity prices by robust nonlinear models. *Lisa Crosato, Luigi Grossi and Fany Nan* 1349

5.19.3 Bias Reduction in a Matching Estimation of Treatment Effect. *Maria Gabriella Campolo, Antonino Di Pino and Edoardo Otranto* 1354

5.20 Statistical Inference and Testing Procedures 1360

5.20.1 Comparison of exact and approximate simultaneous confidence regions in nonlinear regression models. *Claudia Furlan and Cinzia Mortarino* 1360

5.20.2 Tail analysis of a distribution by means of an inequality curve. *E. Taufer, F. Santi, G. Espa and M. M. Dickson* 1367

5.20.3 Nonparametric penalized likelihood for density estimation. *Federico Ferraccioli, Laura M. Sangalli and Livio Finos* 1373

5.20.4 Rethinking the Kolmogorov-Smirnov Test of Goodness of Fit in a Compositional Way. *G.S. Monti, G. Mateu-Figueras, M. I. Ortego, V. Pawlowsky-Glahn and J. J. Egozcue* 1379

5.20.5 Stochastic Dominance for Generalized Parametric Families. *Tommaso Lando and Lucio Bertoli-Barsotti* 1385

5.21 Statistical Models for Ordinal Data 1390

5.21.1 A comparative study of benchmarking procedures for interrater and intrarater agreement studies. *Amalia Vanacore and Maria Sole Pellegrino* 1390

5.21.2 Measuring the multiple facets of tolerance using survey data. *Caterina Liberati and Riccarda Longaretti and Alessandra Michelangeli* 1397

5.21.3 Modified profile likelihood in models for clustered data with missing values. *Claudia Di Caterina and Nicola Sartori* 1401

5.21.4 Worthiness Based Social Scaling. *Giulio D'Epifanio* 1407

5.21.5 Direct Individual Differences Scaling for Evaluation of Research Quality. *Gallo M., Trendafilov N., and Simonacci V.* 1412

5.21.6 A test for variable importance. *Rosaria Simone* 1416

5.22 Statistical Models New Proposals 1422

5.22.1 Decomposing Large Networks: An Approach Based on the MCA based Community Detection. *Carlo Drago* 1422

5.22.2 On Bayesian high-dimensional regression with binary predictors: a simulation study. *Debora Slanzi, Valentina Mameli and Irene Poli* 1429

5.22.3 On the estimation of epidemiological parameters from serological survey data using Bayesian mixture modelling. *Emanuele Del Fava, Piero Manfredi, and Ziv Shkedy* 1435

5.22.4 An evaluation of KL-optimum designs to discriminate between rival copula models. *Laura Deldossi, Silvia Angela Osmetti, Chiara Tommasi* 1441

5.22.5 Variational Approximations for Frequentist and Bayesian Inference. *Luca Maestrini and Matt P. Wand* 1447

5.22.6 Node-specific effects in latent space modelling of multidimensional networks. *Silvia D'Angelo and Marco Alfó and Thomas Brendan Murphy* 1453

5.23 Statistics for Consumer Research 1459

5.23.1 A panel data analysis of Italian hotels. *Antonio Giusti, Laura Grassini, Alessandro Viviani* 1459

5.23.2 A Bayesian Mixed Multinomial Logit Model for Partially Microsimulated Data on Labor Supply. *Cinzia Carota and Consuelo R. Nava* 1466

5.23.3 Comparison between Experience-based Food Insecurity scales. *Federica Onori, Sara Viviani and Pierpaolo Brutti* 1472

5.23.4 Sovereign co-risk measures in the Euro Area. *Giuseppe Arbia, Riccardo Bramante, Silvia Facchinetti, Diego Zappa* 1478

5.23.5 Simultaneous unsupervised and supervised classification modeling for clustering, model selection and dimensionality reduction. *Mario Fordellone and Maurizio Vichi* 1484

5.23.6 Consumers' preference for coffee consumption: a choice experiment including organoleptic characteristics and chemical analysis *Rossella Berni, Nedka D. Nikiforova and Patrizia Pinelli* 1491

5.24 Statistics for Earthquakes 1498

5.24.1 How robust is the skill score of probabilistic earthquake forecasts? *Alessia Caponera and Maximilian J. Werner* 1498

5.24.2 Functional linear models for the analysis of similarity of waveforms. *Francesca Di Salvo, Renata Rotondi and Giovanni Lanzano* 1505

5.24.3 Detection of damage in civil engineering structure by PCA on environmental vibration data. *G. Agró, V. Carlisi, R. Mantione* 1511

5.25 Statistics for Financial Risks 1517

5.25.1 Conditional Value-at-Risk: a comparison between quantile regression and copula functions. *Giovanni De Luca and Giorgia Rivieccio* 1517

5.25.2 Systemic events and diffusion of jumps. *Giovanni Bonaccolto, Nancy Zambon and Massimiliano Caporin* 1523

5.25.3 Traffic Lights for Systemic Risk Detectio. *Massimiliano Caporin, Laura Garcia-Jorcano, Juan-Angel Jiménez-Martin* 1529

5.25.4 Bayesian Quantile Regression Treed. *Mauro Bernardi and Paola Stolfi* 1536

5.25.5 Model Selection in Weighted Stochastic Block models. *Roberto Casarin, Michele Costola, Erdem Yenerdag* 1541

5.26 Tourism & Cultural Participation 1545

5.26.1 The determinants of tourism destination competitiveness in 2006-2016: a partial least square path modelling approach. *Alessandro Magrini, Laura Grassini* 1545

5.26.2 Participation in tourism of Italian residents in the years of the economic recession. *Chiara Bocci, Laura Grassini, Emilia Rocco* 1552

5.26.3 Cultural Participation in the digital Age in Europe: a multilevel cross-national analysis. *Laura Bocci and Isabella Mingo* 1558

5.26.4 Tourist flows and museum admissions in Italy: an integrated analysis. *Lorenzo Cavallo, Francesca Petrei, Maria Teresa Santoro* 1565

5.26.5 Posterior Predictive Assessment for Item Response Theory Models: A Proposal Based on the Hellinger Distance. *Mariagiulia Matteucci and Stefania Mignani* 1571

5.27 Well-being & Quality of Life 1577

5.27.1 Is Structural Equation Modelling Able to Predict Well-being? *Daniele Toninelli and Michela Cameletti* 1577

5.27.2 The well-being in the Italian urban areas: a local geographic variation analysis. *Eugenia Nissi and Annalina Sarra* 1584

5.27.3 Comparing Composite Indicators to measure Quality of Life: the Italian "Sole 24 Ore" case. *Gianna Agró, Mariantonietta Ruggieri and Erasmo Vassallo* . . . 1590

5.27.4 Quality of working life in Italy: findings from Inapp survey. *Paolo Emilio Cardone* 1596

5.27.5 Well-being indices: what about Italian scenario? *Silvia Facchinetti and Elena Siletti* 1603

5.27.6 How can we compare rankings that are expected to be similar? An example based on composite well being indicators. *Silvia Terzi e Luca Moroni* 1609

6 Poster Sessions 1617

6.0.1 A distribution curves comparison approach to analyze the university moving students performance. *Giovanni Boscaino, Giada Adelfio, Gianluca Sottile* . . . 1617

6.0.2 A Partial Ordering Application in Aggregating Dimensions of Subjective Well-being. *Paola Conigliaro* 1624

6.0.3 A note on objective Bayes analysis for graphical vector autoregressive models. *Lucia Paci and Guido Consonni* 1630

6.0.4 Bayesian Population Size Estimation with A Single Sample. *Pierfrancesco Alaimo Di Loro and Luca Tardella* 1636

6.0.5 Classification of the Aneurisk65 dataset using PCA for partially observed functional data. *Marco Stefanucci, Laura Sangalli and Pierpaolo Brutti* 1642

6.0.6 Deep Learning to the Test: an Application to Traffic Data Streams. *Nina Deliu and Pierpaolo Brutti* 1647

6.0.7 Estimating the number of unseen species under heavy tails. *Marco Battiston, Federico Camerlenghi, Emanuele Dolera and Stefano Favaro* 1653

6.0.8 How to measure cybersecurity risk. *Silvia Facchinetti, Paolo Giudici and Silvia Angela Osmetti* 1659

6.0.9 Implementation of an innovative technique to improve Sauvignon Blanc wine quality. *Filippa Bono, Pietro Catanaia and Mariangela Vallone* 1663
6.0.10 Investigating the effect of drugs consumption on survival outcome of Heart Failure patients using joint models: a case study based on regional administrative data. *Marta Spreafico, Francesca Gasperoni, Francesca Ieva* 1669
6.0.11 Mapping the relation between University access test and student's university performance. *Vincenzo Giuseppe Genova, Antonella Plaia* 1675
6.0.12 Multivariate analysis of marine litter abundance through Bayesian space-time models. *C. Calculli, A. Pollice, L. Sion, and P. Maiorano* 1681
6.0.13 Power Priors for Bayesian Analysis of Graphical Models of Conditional Independence in Three Way Contingency Tables. *Katerina Mantzouni, Claudia Tarantola and Ioannis Ntzoufras* 1685
6.0.14 Random Garden: a Supervised Learning Algorithm. *Ivan Luciano Danesi, Valeria Danese, Nicolo' Russo and Enrico Tonini* 1691
6.0.15 Spatiotemporal Prevision for Emergency Medical System Events in Milan. *Andrea Gilardi, Riccardo Borgoni, Andrea Pagliosa, Rodolfo Bonora* 1697
6.0.16 Spatial segregation immigrant households in Messina. *Angelo Mazza and Massimo Mucciardi* 1703
6.0.17 Supervised Learning for Link Prediction in Social Networks. *Riccardo Giubilei, Pierpaolo Brutti* 1707
6.0.18 Women's empowerment and child mortality: the case of Bangladesh. *Chiara Puglisi, Annalisa Busetta* 1713

SIS 2018
49 TH SCIENTIFIC MEETING OF THE ITALIAN STATISTICAL SOCIETY
PALERMO 20-22 JUNE

1. Preface

This book includes the papers presented at the "49th Meeting of the Italian Statistic Society". The conference has registered 445 participants, 350 reports divided into 4 plenary sessions, 20 specialised sessions, 25 sessions solicited, 27 sessions spontaneous, 2 poster sessions. The high number of participants, the high quality of the interventions, the productive spirit of the conference, the ability to respect the time table, are the main indices of the full success of this conference. The meeting hosted also, as plenary sessions, the ISTAT annual report 2018, and a round table on statistics and job markets. Methodological plenary sessions concerned with ordinal data, the dynamics of climate change and models in biomedicine.
Moreover, two related events were held: Start-up Research (SUR) and Stats Under the Stars (SUS4). The SUS4 event attracted many sponsors of statistical, financial, editorial fields as well as numerous students, not only from Italy but also from abroad (Groningen, Tyumen, Barcelona, and Valencia): 98 students for a total of 25 teams. The SUR was a 2-day meeting where small research groups of young scholars, advised by senior researchers with a well-established experience in different areas of Statistics, was asked to develop innovative methods and models to analyse a common dataset from the Neurosciences.

Is retirement bad for health? A matching approach

Il pensionamento fa male alla salute? Una analisi causale

Elena Pirani, Marina Ballerini, Alessandra Mattei, Gustavo De Santis

Abstract The aim of this paper is to assess the causal impact of the transition from work to retirement on individual health in various European countries in recent years. The health effects of this transition are far from clear: the specialized literature reports both positive and negative consequences, however, most of the early studies focus on associations rather than causal relationships. We estimate causal effects of retirement on three measures of health and well-being – self-rated health, depression, quality of life – using a propensity score matching approach under the assumption of selection on observables on data coming from SHARE, the longitudinal Survey on Health, Ageing and Retirement in Europe, in the years between 2004 and 2016. Our results suggest that the transition from work to retirement negatively affect self-rated health almost everywhere in Europe; nevertheless, the quality of life seems to improve, especially in Continental and Mediterranean countries.

Abstract L'obiettivo di questo lavoro è valutare l'impatto causale del pensionamento sulla salute individuale in vari paesi europei. La letteratura riporta effetti sia positivi che negativi del pensionamento sulla salute e benessere degli individui, tuttavia, la maggior parte degli studi precedenti si concentra sulle associazioni piuttosto che sulle relazioni causali. Considerando tre misure di salute – salute percepita, depressione e qualità della vita – sulla base dei dati delle indagini SHARE svolte tra il 2004 e il 2016, stimiamo gli effetti causali del pensionamento sulla salute utilizzando l'approccio del propensity score matching sotto l'ipotesi di assenza di confondimento. I risultati suggeriscono che il pensionamento ha effetti negativi sulla salute percepita in tutti i paesi europei, mentre la qualità della vita sembra beneficiarne, soprattutto nei paesi dell'Europa continentale e mediterranea.

Key words: Retirement; Europe; Share; Causal inference.

---

[1] Elena Pirani, University of Florence, elena.pirani@unifi.it;
Marina Ballerini, University of Florence, marinaballerini.21@gmail.com;
Alessandra Mattei, University of Florence, mattei@disia.unifi.it;
Gustavo De Santis, University of Florence, desantis@disia.unifi.it.

1 Retirement and health: a complex connection

The effects on health of the transition from work to retirement are unclear. Several scholars argue that retirement itself is a stressful event (e.g., Carp 1967; MacBride 1976), which can lead to a break with support networks and friends, and may be accompanied by feelings of loneliness, uselessness, or obsolesce (MacBride 1976). Others claim instead that retirement is a health-preserving life event: it is a relief from work-related stress (Eibich 2015), and encourages health-improving behaviors – such as quit smoking – or increased physical activity (Eibich 2015; Insler 2014).

A strand of the literature reports a significant increase in health after retirement (e.g. Blake and Garrouste 2012; Charles 2004; Coe and Zamarro 2011; Insler 2014; Latif 2013; Neuman 2008), whereas other researchers find significant negative effects on both objective and subjective health measures (e.g. Behncke 2012; Dave et al. 2008; Sahlgren 2012), and also on cognitive functions (Bonsang et al. 2012; Mazzonna and Peracchi 2012). Bound and Waidmann (2007) showed a short-term positive relationship between retirement and health for men but not for women.

In a large part of the previous studies, the focus was on the *association* between health and retirement, and comparisons between the retired and those still working were usually not adjusted for health characteristics before retirement: this adjustment is instead crucial for drawing inference on the causal effects of retirement on health (Coe and Zamarro 2011).

In this paper, we aim to assess the *causal* impact on health of the transition from work to retirement by applying a propensity score matching approach under the assumption of selection on observable to the data of the Survey on Health, Ageing and Retirement in Europe (SHARE). We refer to several European countries, and analyse the heterogeneity of the causal effects across different geographical areas.

2 Data and method

Our empirical analyses were based on SHARE, the Survey on Health, Ageing and Retirement in Europe, which is a panel including five regular waves plus a wave on people's life histories (wave 3, SHARELIFE, which, however, is not considered here because it collected very different information compared to the regular waves and excluded some of the variables we need). In each wave, SHARE data cover the key areas of life (health, socio-economic status social and family networks, etc.) of more than 60,000 individuals aged 50 or over. We focused on the period between 2004 and 2016 and on a subset of the SHARE countries that participated in at least three consecutive waves (Austria, Belgium, Czech Republic, Denmark, Estonia, France, Germany, Italy, the Netherland, Slovenia, Sweden, Switzerland, and Spain), and for which we had all the information we needed for our analysis.

We selected three health and well-being indicators. As a general measure of health, we used self-rated health, dichotomizing the original 5-point scale into good

(excellent, very good, good) and poor (fair, poor) perceived health (preliminary analyses considering the original formulation proved consistent results). Because of its subjective nature, self-rated health may change across populations (Prinja et al., 2012); however, various studies proved its power in predicting objective health conditions (Egidi and Spizzichino, 2006), physical and emotional well-being (Bayliss et al., 2012), and even mortality (Idler and Benyamini, 1997). Moreover, we used a composite indicator of depression constructed from the 12 basic items of the EURO-D scale (Prince et al., 1999): depressed mood, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness. This scale, which ranges from 0 (not depressed) to 12 (very depressed), was developed in an effort to derive a common scale of depression symptoms, especially in later life, based on different indicators in several European countries. Finally, we considered a theoretically grounded measure of quality of life, i.e., a composite indicator based on four subscales corresponding to four life domains (Hyde et al., 2003; Mehrbrodt et al., 2017): control (C), autonomy (A), self-realization (S) and pleasure (P). After reversing the original scale, this indicator, which was proved to represent a reliable indicator of quality of life in the context of research on ageing (Wiggins et al., 2008), scored between 0 (high quality of life) and 36 (low quality of life).

Because our objective was the estimation of a causal effect of entering retirement on individual well-being in the short run, we focused on the subsample of those who were in the labor market (employed or self-employed) in the waves *t=1,2,4* of the survey and were either in the labor market or retired in the waves *t=2,4,5*. Those who got out of the labor market for other reasons (e.g., unemployment) were discarded from our subsample.

We followed the "potential outcome" approach (e.g., Imbens and Rubin, 2015). For each unit i, $i=1,\ldots,n$, we considered a vector X_i of background variables. Let D_i denote the treatment variable indicator, equal to 1 if unit i retires between two consecutive waves of the survey – waves *w1* and *w2*; waves *w2* and *w4*, or waves *w4* and *w5* – and zero otherwise. Under the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980), for each unit i there are two potential outcomes at a future point in time after treatment (at waves *w4*, *w5* or *w6*): the value of the health outcome Y if unit i retired – $Y_i(1)$ – and the value of Y at the same future point in time if the unit did not retire – $Y_i(0)$. The causal effect of the transition from work to retirement for each unit is defined as a comparison of the treatment and control potential outcomes, $Y_i(1)$ and $Y_i(0)$, typically their difference. In this paper we focus on Average Treatment Effects on the Treated (ATT effects), that is, the effects of the transition from work to retirement averaged over the subpopulation of units who actually retired: $\mathrm{ATT} = \mathrm{E}[Y_i(1) - Y_i(0) \mid D=1]$.

Because we used observational data, we needed to introduce some assumptions on the treatment-assignment mechanism to draw inference on the causal effects of interest. We assumed unconfoundedness (or selection on observables), which implies that, conditioning on the observed covariates, an experimental-like context is reproduced. Formally, unconfoundedness requires that the treatment assignment is independent of the potential outcomes: $D_i \perp (Y_i(0), Y_i(1)) | X_i$. We also assumed that

there was sufficient overlap in the joint distribution of the covariates between treated and control subjects: $0 < P(D_i=1|X_i=x) < 1$ for each *i*. Under these assumptions, we applied a statistical matching technique, the purpose of which was to select a sub-group of control subjects (who did not retire between two subsequent waves) who were, in all respects, as similar as possible to the treated subjects, i.e., those who retired from work. We matched individuals based on the propensity score, or the probability of entering retirement conditional on the observed covariates. The propensity score is a balancing score, that is, covariates are independent of the treatment conditional on the propensity score. Moreover, if the unconfoundeness and the overlap assumptions hold conditional on covariates, they also hold conditional on the propensity score (Rosenbaum and Rubin 1983). Therefore, matching on the propensity score is sufficient to remove confounding.

The variables on which we constructed our propensity score included individual socio-demographic characteristics (e.g., age, living arrangement, relatives alive, level of education) and health-related behaviours and health conditions (e.g., smoking and drinking, mobility index, as well as self-rated health, quality of life, and depression). We also introduced the type and the sector of work, even if we acknowledge that these aspects only partially account for important aspects of working life, such as stressing factors or autonomy in decisions, which could importantly contribute to individual's health and wellbeing. We imposed an exact matching on country of residence and gender. On the basis of the estimated propensity score, for each of the 1124 retired individuals we selected as a match the closest individual – i.e., a person of the same sex, from the same country and with very similar pre-treatment characteristics – among the 6250 potential controls (1-to-1 nearest neighbor matching; Abadie and Imbens 2011).

3 Preliminary results

The check of the covariate balance (i.e., similarity of treated and controls individuals in terms of covariates) proved that the matching procedure was successful: after matching, the differences between the two groups (retired an non-retired) in terms of socio-demographic, work-related and health covariates either disappeared (best case) or were drastically reduced (not shown here). We proceeded then to the estimation of the causal effect of retirement on health, by computing the Average Treatment Effects on the Treated (ATT) using a matching estimator.

The first part of Table 1 reports the ATT effects of retirement on the three well-being indicators considered, computed for all the SHARE countries together. Our results show a worsening of self-rated health after retirement; conversely, they also convey the impression of a slightly improvement in terms of (less) depression and (higher) quality of life. Note however that these two latter estimated effects are very small in absolute terms and not statistically significant.

In order to account for the heterogeneity of people living in the different European countries, we estimated the ATT effects separately for the Nordic

countries (Denmark and Sweden), Continental countries (Austria, Belgium, France, Germany, the Netherland, and Switzerland), Mediterranean countries (Spain and Italy), and East European countries (Czech Republic, Slovenia, Estonia). The negative effect of retirement on self-rated health persisted for all European areas, even if with a loss of significance for East European countries.

Considering European countries altogether masks some territorial differences for the other two well-being indicators: both depression and quality of life levels increase for retired people in Continental and Mediterranean countries (although the effect is not statistically significant for the second group, maybe due to the small sample size). On the contrary, in Nordic countries retirement seems to determine a detrimental effect also on depression and quality of life (but again results are not statistically significant), whereas for Eastern European countries the effect on the two well-being variables diverges.

Table 1: Estimated ATT effects and their standard errors, for all countries, and by groups of countries. SHARE 2004-2016.

			Self-rated health	Depression	Quality of life
All countries	treated (n=1124),	ATT	0.09	-0.12	-0.17
	matched controls (n=1124)	*Std Err*	*0.01*	*0.07*	*0.21*
By group of countries:					
Nordic	treated (n=194),	ATT	0.09	0.13	0.29
	matched controls (n=194)	*Std Err*	*0.03*	*0.15*	*0.44*
Continental	treated (n=474),	ATT	0.08	-0.10	-0.91
	matched controls (n=474)	*Std Err*	*0.02*	*0.11*	*0.31*
Mediterranean	treated (n=116),	ATT	0.12	-0.14	-0.13
	matched controls (n=116)	*Std Err*	*0.05*	*0.22*	*0.68*
East European	treated (n=340),	ATT	0.04	-0.24	0.68
	matched controls (n=340)	*Std Err*	*0.03*	*0.15*	*0.44*

In our analysis the richness of background (i.e., pre-treatment) information allowed us to adjust treatment comparisons for a large set of pre-treatment characteristics – in terms of health, life-style behaviors, socio-demographic characteristics and factors linked to the (previous) working condition – and thus the assumption of selection on observable appears to be plausible. Under this assumption, we found that the effects of retirement on health vary not only according to the context of reference, but also depending on the specific health/well-being indicator considered.

It is thus worth to investigate more in detail the mechanisms through which retirement affects the various dimensions of health and well-being. Specifically, aspects such as family types and intergenerational relationships, social relationships, embeddedness in social network, and job characteristics will be examined in our future research.

Acknowledgements: We thank the CREW group for early feedback on this paper (CREW=Care, Retirement and Wellbeing of Older People across Different Welfare Regimes; https://crew-more-years-better-lives.org/).

References

1. Abadie A, Imbens GW. (2011). Bias-corrected matching estimators for average treatment effects. *J Bus Econ Stat.*; 29(1), 1–11.
2. Bayliss E.A., Ellis J.L., Shoup J.A., Zeng C., McQuillan D.B., Steiner, J.F. (2012). Association of patient-centered outcomes with patient-reported and icd-9-based morbidity measures. *Ann. Fam. Med.,* 10, 126-133.
3. Behncke S. (2012). Does retirement trigger ill health?, *Health Economics*, 21, 282-300
4. Blake H., and Garrouste C. (2012). Collateral effects of a pension reform in France, *Health, Econometrics and Data Group (HEDG) Working Papers* 12/16, University of York
5. Bonsang E., Adam S., and Perelman S. (2012), Does retirement affect cognitive functioning?, *Journal of Health Economics*, 31, 490– 501
6. Bound J., and Waidmann T. (2007). Estimating the Health Effects of Retirement, University of Michigan *Retirement Research Center working paper* 2007-168
7. Carp F.M. (1967). Retirement crisis, *Science* 157, 102–103
8. Charles K. (2004). Is retirement depressing? Labor force inactivity and psychological well-being in later life, *Research in Labor Economics*, 23, 269-299
9. Coe N.B., and Zamarro G. (2011), Retirement effects on health in Europe, *Journal of Health Economics*, 30, 77-86
10. Dave D., Rashad I. and Spasojevic J. (2008). The Effects of Retirement on Physical and Mental Health Outcomes, *Southern Economic Journal*, 75, 2, 497-523
11. Egidi V., Spizzichino D. (2006). Perceived health and mortality: a multidimensional analysis of ECHP Italian Data. Genus LXII (3e4), 135e154
12. Eibich P. (2015). Understanding the effect of retirement on health: Mechanisms and Heterogeneity, *J. Health Econ*., 43, 1–12
13. Hyde M., Wiggins R.D., Higgs P., & Blane D.B. (2003). A measure of quality of life in early old age: the theory, development and properties of a needs satisfaction model (CASP-19). *Aging & mental health*, 7(3), 186-194.
14. Idler L.E., Benyamini Y. (1997). Self- rated health and mortality: a review of 27 community studies. *J. Health Soc. Behav.*, 38, 21-37.
15. Imbens GW, Rubin DB (2015). *Causal inference for statistics, social, and biomedical sciences: an introduction*. 1st ed. New York: Cambridge University Press.
16. Insler M. (2014). The Health Consequences of Retirement, *The Journal of Human Resources*, 49, 1
17. Latif E. (2013). The impact of retirement on mental health in Canada, *Journal of Mental Health Policy and Economics*, 16(1), 35-46.
18. MacBride A. (1976). Retirement as a life crisis: myth or reality? *Canadian Psychiatric Association Journal* 72, 547–556
19. Mazzonna F. and Peracchi F. (2014). Unhealthy Retirement?, *EIEF Working Paper 14/09*
20. Mehrbrodt T., Gruber S. & Wagner S. (2017). Scales and Multi-Item Indicators. Share Manuals, retrieved on www.share-project.org/fileadmin/pdf_documentation/SHARE_Scales_and_Multi-Item_Indicators.pdf
21. Neuman K. (2008). Quit Your Job and Get Healthier? The Effect of Retirement on Health, *Journal of Labor Research*, 29, 177-201
22. Prince M.J., Reischies F., Beekman A.T., et al.(1999). Development of the EURO-D scale-a European, Union initiative to compare symptoms of depression in 14 European centres. *British Journal of Psychiatry*,174, 330–8.
23. Rosenbaum P.R, Rubin D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
24. Rubin DB. (1980). Discussion of "Randomization analysis of experimental data in the Fisher randomization test", by Basu, *Journal of the American Statistical Association*, 75, 591–593.
25. Sahlgren G. (2012), Work 'til you drop: Short- and longer-term health effects of retirement in Europe, *IFN Working Paper* n. 928
26. Wiggins R.D., Netuveli G., Hyde M., Higgs P., Blane D. (2008). The Evaluation of a Self-enumerated Scale of Quality of Life (CASP-19) in the Context of Research on Ageing: A Combination of Exploratory and Confirmatory Approaches. *Social Indicators Research*, 89, 61–77.