

SIS 2017
Statistics and Data Science:
new challenges, new generations

28–30 June 2017
Florence (Italy)

Proceedings of the Conference
of the Italian Statistical Society

edited by
Alessandra Petrucci
Rosanna Verde

FIRENZE UNIVERSITY PRESS
2017

SIS 2017. Statistics and Data Science: new challenges, new generations : 28-30 June 2017 Florence (Italy) : proceedings of the Conference of the Italian Statistical Society / edited by Alessandra Petrucci, Rosanna Verde. – Firenze : Firenze University Press, 2017.

(Proceedings e report ; 114)

<http://digital.casalini.it/9788864535210>

ISBN 978-88-6453-521-0 (online)

Peer Review Process

All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Committees of the individual series. The works published in the FUP catalogue are evaluated and approved by the Editorial Board of the publishing house. For a more detailed description of the refereeing process we refer to the official documents published on the website and in the online catalogue of the FUP (www.fupress.com).

Firenze University Press Editorial Board

A. Dolfi (Editor-in-Chief), M. Boddi, A. Bucelli, R. Casalbuoni, M. Garzaniti, M.C. Grisolia, P. Guarnieri, R. Lanfredini, A. Lenzi, P. Lo Nostro, G. Mari, A. Mariani, P.M. Mariano, S. Marinai, R. Minuti, P. Nanni, G. Nigro, A. Perulli, M.C. Torricelli.

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0: <https://creativecommons.org/licenses/by/4.0/legalcode>)

CC 2017 Firenze University Press
Università degli Studi di Firenze
Firenze University Press
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

SIS2017 Committees

Scientific Program Committee:

Rosanna Verde (chair), Università della Campania “Luigi Vanvitelli”
Maria Felice Arezzo, Sapienza Università di Roma
Antonino Mazzeo, Università di Napoli Federico II
Emanuele Baldacci, Eurostat
Pierpaolo Brutti, Sapienza Università di Roma
Marcello Chiodi, Università di Palermo
Corrado Crocetta, Università di Foggia
Giovanni De Luca, Università di Napoli Parthenope
Viviana Egidi, Sapienza Università di Roma
Giulio Ghellini, Università degli Studi di Siena
Ippoliti Luigi, Università di Chieti-Pescara “G. D’Annunzio”
Matteo Mazziotto, ISTAT
Lucia Paci, Università Cattolica del Sacro Cuore
Alessandra Petrucci, Università degli Studi di Firenze
Filomena Racioppi, Sapienza Università di Roma
Laura M. Sangalli, Politecnico di Milano
Bruno Scarpa, Università degli Studi di Padova
Cinzia Viroli, Università di Bologna

Local Organizing Committee:

Alessandra Petrucci (chair), Università degli Studi di Firenze
Gianni Betti, Università degli Studi di Siena
Fabrizio Cipollini, Università degli Studi di Firenze
Emanuela Dreassi, Università degli Studi di Firenze
Caterina Giusti, Università di Pisa
Leonardo Grilli, Università degli Studi di Firenze
Alessandra Mattei, Università degli Studi di Firenze
Elena Pirani, Università degli Studi di Firenze
Emilia Rocco, Università degli Studi di Firenze
Maria Cecilia Verri, Università degli Studi di Firenze

Supported by:

Università degli Studi di Firenze
Università di Pisa
Università degli Studi di Siena
ISTAT
Regione Toscana
Comune di Firenze
BITBANG srl

Index

Preface	XXV
Alexander Agapitov, Irina Lackman, Zoya Maksimenko <i>Determination of basis risk multiplier of a borrower default using survival analysis</i>	1
Tommaso Agasisti, Alex J. Bowers, Mara Soncin <i>School principals' leadership styles and students achievement: empirical results from a three-step Latent Class Analysis</i>	7
Tommaso Agasisti, Sergio Longobardi, Felice Russo <i>Poverty measures to analyse the educational inequality in the OECD Countries</i>	17
Mohamed-Salem Ahmed, Laurence Broze, Sophie Dabo-Niang, Zied Gharbi <i>Quasi-Maximum Likelihood Estimators For Functional Spatial Autoregressive Models</i>	23
Giacomo Aletti, Alessandra Micheletti <i>A clustering algorithm for multivariate big data with correlated components</i>	31
Emanuele Aliverti <i>A Bayesian semiparametric model for terrorist networks</i>	37

Giorgio Alleva <i>Emerging challenges in official statistics: new sources, methods and skills</i>	43
Rémi André, Xavier Luciani and Eric Moreau <i>A fast algorithm for the canonical polyadic decomposition of large tensors</i>	45
Maria Simona Andreano, Roberto Benedetti, Paolo Postiglione, Giovanni Savio <i>On the use of Google Trend data as covariates in nowcasting: Sampling and modeling issues</i>	53
Francesco Andreoli, Mauro Mussini <i>A spatial decomposition of the change in urban poverty concentration</i>	59
Margaret Antonicelli, Vito Flavio Covella <i>How green advertising can impact on gender different approach towards sustainability</i>	65
Rosa Arboretti, Eleonora Carrozzo, Luigi Salmaso <i>Stratified data: a permutation approach for hypotheses testing</i>	71
Marika Arena, Anna Calissano, Simone Vantini <i>Crowd and Minorities: Is it possible to listen to both? Monitoring Rare Sentiment and Opinion Categories about Expo Milano 2015</i>	79
Maria Felice Arezzo, Giuseppina Guagnano <i>Using administrative data for statistical modeling: an application to tax evasion</i>	83
Monica Bailot, Rina Camporese, Silvia Da Valle, Sara Letardi, Susi Osti <i>Are Numbers too Large for Kids? Possible Answers in Probable Stories</i>	89

Index	IX
Simona Balbi, Michelangelo Misuraca, Germana Scepi <i>A polarity-based strategy for ranking social media reviews</i>	95
A. Balzanella, S.A. Gattone, T. Di Battista, E. Romano, R. Verde <i>Monitoring the spatial correlation among functional data streams through Moran's Index</i>	103
Oumayma Banouar, Said Raghay <i>User query enrichment for personalized access to data through ontologies using matrix completion method</i>	109
Giulia Barbati, Francesca Ieva, Francesca Gasperoni, Annamaria Iorio, Gianfranco Sinagra, Andrea Di Lenarda <i>The Trieste Observatory of cardiovascular disease: an experience of administrative and clinical data integration at a regional level</i>	115
Francesco Bartolucci, Stefano Peluso, Antonietta Mira <i>Marginal modeling of multilateral relational events</i>	123
Francesca Bassi, Leonardo Grilli, Omar Paccagnella, Carla Rampichini, Roberta Varriale <i>New Insights on Students Evaluation of Teaching in Italy</i>	129
Mauro Bernardi, Marco Bottone, Lea Petrella <i>Bayesian Quantile Regression using the Skew Exponential Power Distribution</i>	135
Mauro Bernardi <i>Bayesian Factor-Augmented Dynamic Quantile Vector Autoregression</i>	141

Bruno Bertaccini, Giulia Biagi, Antonio Giusti, Laura Grassini <i>Does data structure reflect monuments structure? Symbolic data analysis on Florence Brunelleschi Dome</i>	149
Gaia Bertarelli and Franca Crippa, Fulvia Mecatti <i>A latent markov model approach for measuring national gender inequality</i>	157
Agne Bikauskaite, Dario Buono <i>Eurostat's methodological network: Skills mapping for a collaborative statistical office</i>	161
Francesco C. Billari, Emilio Zagheni <i>Big Data and Population Processes: A Revolution?</i>	167
Monica Billio, Roberto Casarin, Matteo Iacopini <i>Bayesian Tensor Regression models</i>	179
Monica Billio, Roberto Casarin, Luca Rossini <i>Bayesian nonparametric sparse Vector Autoregressive models</i>	187
Chiara Bocci, Daniele Fadda, Lorenzo Gabrielli, Mirco Nanni, Leonardo Piccini <i>Using GPS Data to Understand Urban Mobility Patterns: An Application to the Florence Metropolitan Area</i>	193
Michele Boreale, Fabio Corradi <i>Relative privacy risks and learning from anonymized data</i>	199
Giacomo Bormetti, Roberto Casarin, Fulvio Corsi, Giulia Livieri <i>A stochastic volatility framework with analytical filtering</i>	205

Index	XI
Alessandro Brunetti, Stefania Fatello, Federico Polidoro <i>Estimating Italian inflation using scanner data: results and perspectives</i>	211
Guénael Cabanes, Younès Bennani, Rosanna Verde, Antonio Irpino <i>Clustering of histogram data : a topological learning approach</i>	219
Renza Campagni, Lorenzo Gabrielli, Fosca Giannotti, Riccardo Guidotti, Filomena Maggino, Dino Pedreschi <i>Measuring Wellbeing by extracting Social Indicators from Big Data</i>	227
Maria Gabriella Campolo, Antonino Di Pino <i>Assessing Selectivity in the Estimation of the Causal Effects of Retirement on the Labour Division in the Italian Couples</i>	235
Stefania Capecchi, Rosaria Simone <i>Composite indicators for ordinal data: the impact of uncertainty</i>	241
Stefania Capecchi, Domenico Piccolo <i>The distribution of Net Promoter Score in socio-economic surveys</i>	247
Massimiliano Caporin, Francesco Poli <i>News, Volatility and Price Jumps</i>	253
Carmela Cappelli, Rosaria Simone, Francesca di Iorio <i>Growing happiness: a model-based tree</i>	261
Paolo Emilio Cardone <i>Inequalities in access to job-related learning among workers in Italy: evidence from Adult Education Survey (AES)</i>	267

- Alessandro Casa, Giovanna Menardi
Signal detection in high energy physics via a semisupervised nonparametric approach
 273
- Claudio Ceccarelli, Silvia Montagna, Francesca Petrarca
Employment study methodologies of Italian graduates through the data linkage of administrative archives and sample surveys
 279
- Ikram Chairi, Amina El Gonnouni, Sarah Zouinina, Abdelouahid Lyhyaoui
Prediction of Firm's Creditworthiness Risk using Feature Selection and Support Vector Machine
 285
- Sana Chakri, Said Raghay, Salah El Hadaj
Contribution of extracting meaningful patterns from semantic trajectories
 293
- Chieppa A., Ferrara R., Gallo G., Tomeo V.
Towards The Register-Based Statistical System: A New Valuable Source for Population Studies
 301
- Shirley Coleman
Consulting, knowledge transfer and impact case studies of statistics in practice
 305
- Michele Costa
The evaluation of the inequality between population subgroups
 313
- Michele Costola
Bayesian Non-Negative l_1 -Regularised Regression
 319
- Lisa Crosato, Caterina Liberati, Paolo Mariani, Biancamaria Zavanella
Industrial Production Index and the Web: an explorative cointegration analysis
 327

Index	XIII
Francesca Romana Crucinio, Roberto Fontana <i>Comparison of conditional tests on Poisson data</i>	333
Riccardo D'Alberto, Meri Raggi <i>Non-parametric micro Statistical Matching techniques: some developments</i>	339
Stefano De Cantis, Mauro Ferrante, Anna Maria Parroco <i>Measuring tourism from demand side</i>	345
Lucio De Capitani, Daniele De Martini <i>Optimal Ethical Balance for Phase III Trials Planning</i>	351
Claudia De Vitiis, Alessio Guandalini, Francesca Inglese, Marco D. Terribili <i>Sampling schemes using scanner data for the consumer price index</i>	357
Ermelinda Della Valle, Elena Scardovi, Andrea Iacobucci, Edoardo Tignone <i>Interactive machine learning prediction for budget allocation in digital marketing scenarios</i>	365
Marco Di Marzio, Stefania Fensore, Agnese Panzera, Charles C. Taylor <i>Nonparametric classification for directional data</i>	371
Edwin Diday <i>Introduction to Symbolic Data Analysis and application to post clustering for comparing and improving clustering methods by the Symbolic Data Table that they induce</i>	379
Carlo Drago <i>Identifying Meta Communities on Large Networks</i>	387

- Neska El Haouij, Jean-Michel Poggi, Raja Ghozi, Sylvie Sevestre Ghalila, Mériem Jaidane
Random Forest–Based Approach for Physiological Functional Variable Selection for Drivers Stress Level Classification
 393
- Silvia Facchinetti, Silvia A. Osmetti
A risk index to evaluate the criticality of a product defectiveness
 399
- Federico Ferraccioli, Livio Finos
Exponential family graphical models and penalizations
 405
- Mauro Ferrante, Giovanna Fantaci, Anna Maria Parroco, Anna Maria Milito, Salvatore Scondotto
Key–indicators for maternity hospitals and newborn readmission in Sicily
 411
- Ferretti Camilla, Ganugi Piero, Zammori Francesco
Change of Variables theorem to fit Bimodal Distributions
 417
- Francesco Finazzi, Lucia Paci
Space–time clustering for identifying population patterns from smartphone data
 423
- Annunziata Fiore, Antonella Simone, Antonino Virgillito
IT Solutions for Analyzing Large-Scale Statistical Datasets: Scanner Data for CPI
 429
- Michael Fop, Thomas Brendan Murphy, Luca Scrucca
Model–based Clustering with Sparse Covariance Matrices
 437
- Maria Franco-Villoria, Marian Scott
Quantile Regression for Functional Data
 441

- Gallo M., Simonacci V., Di Palma M.A.
Three-way compositional data: a multi-stage trilinear decomposition algorithm
445
- Francesca Gasperoni, Francesca Ieva, Anna Maria Paganoni, Chris Jackson, Linda Sharples
Nonparametric shared frailty model for classification of survival data
451
- Stefano A. Gattone, Angela De Sanctis
Clustering landmark-based shapes using Information Geometry tools
457
- Alan E. Gelfand, Shinichiro Shirota
Space and circular time log Gaussian Cox processes with application to crime event data
461
- Abdelghani Ghazdali
Blind source separation
469
- Massimiliano Giacalone, Antonio Ruoto, Davide Liga, Maria Pilato, Vito Santarcangelo
An innovative approach for Opinion Mining : the Plutchick analysis
479
- Massimiliano Giacalone, Demetrio Panarello
A G.E.D. method for market risk evaluation using a modified Gaussian Copula
485
- Chiara Gigliarano, Francesco Maria Chelli
Labour market dynamics and recent economic changes: the case of Italy
491
- Giuseppe Giordano, Giancarlo Ragozini, Maria Prosperina Vitale
On the use of DISTATIS to handle multiplex networks
499

- Michela Gnaldi, Silvia Bacci, Samuel Greiff, Thiemo Kunze
Profiles of students on account of complex problem solving (CPS) strategies exploited via log-data
505
- Michela Gnaldi, Simone Del Sarto
Characterising Italian municipalities according to the annual report of the prevention-of-corruption supervisor: a Latent Class approach
513
- Silvia Golia
A proposal of a discretization method applicable to Rasch measures
519
- Anna Gottard
Tree-based Non-linear Graphical Models
525
- Sara Hbali, Youssef Hbali, Mohamed Sadgal, Abdelaziz El Fazziki
Sentiment Analysis for micro-blogging using LSTM Recurrent Neural Networks
531
- Stefano Maria Iacus, Giuseppe Porro, Silvia Salini, Elena Siletti
How to Exploit Big Data from Social Networks: a Subjective Well-being Indicator via Twitter
537
- Francesca Ieva
Network Analysis of Comorbidity Patterns in Heart Failure Patients using Administrative Data
543
- Antonio Irpino, Francisco de A.T. De Carvalho, Rosanna Verde
Automatic variable and components weighting systems for Fuzzy cmeans of distributional data
549
- Michael Jauch, Paolo Giordani, David Dunson
A Bayesian oblique factor model with extension to tensor data
553

- Johan Koskinen, Chiara Broccatelli, Peng Wang, Garry Robins
Statistical analysis for partially observed multilayered networks
561
- Francesco Lagona
Copula-based segmentation of environmental time series with linear and circular components
569
- Alessandro Lanteri, Mauro Maggioni
A Multiscale Approach to Manifold Estimation
575
- Tiziana Laureti, Carlo Ferrante, Barbara Dramis
Using scanner and CPI data to estimate Italian sub-national PPPs
581
- Antonio Lepore
Graphical approximation of Best Linear Unbiased Estimators for Extreme Value Distribution Parameters
589
- Antonio Lepore, Biagio Palumbo, Christian Capezza
Monitoring ship performance via multi-way partial least-squares analysis of functional data
595
- Caterina Liberati, Lisa Crosato, Paolo Mariani, Biancamaria Zavanella
Dynamic profiling of banking customers: a pseudo-panel study
601
- Giovanni L. Lo Magno, Mauro Ferrante, Stefano De Cantis
A comparison between seasonality indices deployed in evaluating unimodal and bimodal patterns
607
- Rosaria Lombardo, Eric J Beh
Three-way Correspondence Analysis for Ordinal-Nominal Variables
613

- Monia Lupparelli, Alessandra Mattei
Log-mean linear models for causal inference
 621
- Badiaa Lyoussi, Zineb Selihi, Mohamed Berraho, Karima El Rhazi, Youness El Achhab, Adiba El Marrakchi, Chakib Nejjari
Research on the Risk Factors accountable for the occurrence of degenerative complications of type 2 diabetes in Morocco: a prospective study
 627
- Valentina Mameli, Debora Slanzi, Irene Poli
Bootstrap group penalty for high-dimensional regression models
 633
- Stefano Marchetti, Monica Pratesi, Caterina Giusti
Improving small area estimates of households' share of food consumption expenditure in Italy by means of Twitter data
 639
- Paolo Mariani, Andrea Marletta, Mariangela Zenga
Gross Annual Salary of a new graduate: is it a question of profile?
 647
- Maria Francesca Marino, Marco Alfò
Dynamic random coefficient based drop-out models for longitudinal responses
 653
- Antonello Maruotti, Jan Bulla
Hidden Markov models: dimensionality reduction, atypical observations and algorithms
 659
- Chiara Masci, Geraint Johnes, Tommaso Agasisti
A flexible analysis of PISA 2015 data across countries, by means of multilevel trees and boosting
 667

Index	XIX
Lucio Masserini, Matilde Bini <i>Impact of the 2008 and 2012 financial crises on the unemployment rate in Italy: an interrupted time series approach</i>	673
Angelo Mazza, Antonio Punzo, Salvatore Ingrassia <i>An R Package for Cluster-Weighted Models</i>	681
Antonino Mazzeo, Flora Amato <i>Methods and applications for the treatment of Big Data in strategic fields</i>	687
Letizia Mencarini, Viviana Patti, Mirko Lai, Emilio Sulis <i>Happy parents' tweets</i>	693
Rodolfo Metulini, Marica Manisera, Paola Zuccolotto <i>Space-Time Analysis of Movements in Basketball using Sensor Data</i>	701
Giorgio E. Montanari, Marco Doretto, Francesco Bartolucci <i>An ordinal Latent Markov model for the evaluation of health care services</i>	707
Isabella Morlini, Maristella Scorza <i>New fuzzy composite indicators for dyslexia</i>	713
Fionn Murtagh <i>Big Textual Data: Lessons and Challenges for Statistics</i>	719
Gaetano Musella, Gennaro Punzo <i>Workers' skills and wage inequality: A time-space comparison across European Mediterranean countries</i>	731

Marta Nai Ruscone <i>Exploratory factor analysis of ordinal variables: a copula approach</i>	737
Fausta Ongaro, Silvana Salvini <i>IPUMS Data for describing family and household structures in the world</i>	743
Tullia Padellini, Pierpaolo Brutti <i>Topological Summaries for Time-Varying Data</i>	747
Sally Paganin <i>Modeling of Complex Network Data for Targeted Marketing</i>	753
Francesco Palumbo, Giancarlo Ragozini <i>Statistical categorization through archetypal analysis</i>	759
Michela Eugenia Pasetto, Umberto Noè, Alessandra Luati, Dirk Husmeier <i>Inference with the Unscented Kalman Filter and optimization of sigma points</i>	767
Xanthi Pedeli, Cristiano Varin <i>Pairwise Likelihood Inference for Parameter-Driven Models</i>	773
Felicia Pelagalli, Francesca Greco, Enrico De Santis <i>Social emotional data analysis. The map of Europe</i>	779
Alessia Pini, Lorenzo Spreafico, Simone Vantini, Alessandro Vietti <i>Differential Interval-Wise Testing for the Inferential Analysis of Tongue Profiles</i>	785
Alessia Pini, Aymeric Stamm, Simone Vantini <i>Hotelling meets Hilbert: inference on the mean in functional Hilbert spaces</i>	791

Silvia Polettini, Serena Arima

Accounting for measurement error in small area models: a study on generosity
795

Gennaro Punzo, Mariateresa Ciommi

Structural changes in the employment composition and wage inequality: A comparison across European countries
801

Walter J. Radermacher

Official Statistics 4.0 – learning from history for the challenges of the future
809

Fabio Rapallo

Comparison of contingency tables under quasi-symmetry
821

Valentina Raponi, Cesare Robotti, Paolo Zaffaroni

Testing Beta-Pricing Models Using Large Cross-Sections
827

Marco Seabra dos Reis, Biagio Palumbo, Antonio Lepore, Ricardo Rendall, Christian Capezza

On the use of predictive methods for ship fuel consumption analysis from massive on-board operational data
833

Alessandra Righi, Mauro Mario Gentile

Twitter as a Statistical Data Source: an Attempt of Profiling Italian Users Background Characteristics
841

Paolo Righi, Giulio Barcaroli, Natalia Golini

Quality issues when using Big Data in Official Statistics
847

Emilia Rocco

Indicators for the representativeness of survey response as well as convenience samples
855

- Emilia Rocco, Bruno Bertaccini, Giulia Biagi, Andrea Giommi
A sampling design for the evaluation of earthquakes vulnerability of the residential buildings in Florence
861
- Elvira Romano, Jorge Mateu
A local regression technique for spatially dependent functional data: an heteroskedastic GWR model
867
- Eduardo Rossi, Paolo Santucci de Magistris
Models for jumps in trading volume
873
- Renata Rotondi, Elisa Varini
On a failure process driven by a self-correcting model in seismic hazard assessment
879
- M. Ruggieri, F. Di Salvo and A. Plaia
Functional principal component analysis of quantile curves
887
- Massimiliano Russo
Detecting group differences in multivariate categorical data
893
- Michele Scagliarini
A Sequential Test for the C_{pk} Index
899
- Steven L. Scott
Industrial Applications of Bayesian Structural Time Series
905
- Catia Scricciolo
Asymptotically Efficient Estimation in Measurement Error Models
913

Index	XXIII
Angela Serra, Pietro Coretto, Roberto Tagliaferri <i>On the noisy high-dimensional gene expression data analysis</i>	919
Mirko Signorelli <i>Variable selection for (realistic) stochastic blockmodels</i>	927
Marianna Siino, Francisco J. Rodriguez-Cortés, Jorge Mateu, Giada Adelfio <i>Detection of spatio-temporal local structure on seismic data</i>	935
A. Sottosanti, D. Bastieri, A. R. Brazzale <i>Bayesian Mixture Models for the Detection of High-Energy Astronomical Sources</i>	943
Federico Mattia Stefanini <i>Causal analysis of Cell Transformation Assays</i>	949
Paola Stolfi, Mauro Bernardi, Lea Petrella <i>Estimation and Inference of SkewStable distributions using the Multivariate Method of Simulated Quantiles</i>	955
Paola Stolfi, Mauro Bernardi, Lea Petrella <i>Sparse Indirect Inference</i>	961
Peter Struijs, Anke Consten, Piet Daas, Marc Debusschere, Maiki Ilves, Boro Nikic, Anna Nowicka, David Salgado, Monica Scannapieco, Nigel Swier <i>The ESSnet Big Data: Experimental Results</i>	969
Jérémie Sublime <i>Smart view selection in multi-view clustering</i>	977

- Emilio Sulis
Social Sensing and Official Statistics: call data records and social media sentiment analysis
985
- Matilde Trevisani, Arjuna Tuzzi
Knowledge mapping by a functional data analysis of scientific articles databases
993
- Amalia Vanacore, Maria Sole Pellegrino
Characterizing the extent of rater agreement via a non-parametric benchmarking procedure
999
- Maarten Vanhoof, Stephanie Combes, Marie-Pierre de Bellefon
Mining Mobile Phone Data to Detect Urban Areas
1005
- Viktoriya Voytsekhovska, Olivier Butzbach
Statistical methods in assessing the equality of income distribution, case study of Poland
1013
- Ernst C. Wit
Network inference in Genomics
1019
- Dilek Yildiz, Jo Munson, Agnese Vitali, Ramine Tinati, Jennifer Holland
Using Twitter data for Population Estimates
1025
- Marco Seabra dos Rei
Structured Approaches for High-Dimensional Predictive Modeling
1033

Measuring tourism from demand side

La misura del turismo dal lato della domanda

Stefano De Cantis, Mauro Ferrante, Anna Maria Parroco

Abstract This Paper proposes an analysis of tourism from the demand side, taking into account for both the total level of tourism demand produced by some European countries (domestic and outgoing) and its general tendency, and for the seasonal fluctuations which characterize many tourism-related aggregates. Tourist flows from the demand side at the European level are analyzed in the last decade, and a special focus on Italian tourism demand is provided, jointly with an analysis of its seasonal fluctuations. The analysis of general tendency of tourism demand and of the impacts of seasonality is a fundamental pre-requisite for the implementation of tourism policies.

Abstract *Il presente articolo propone un'analisi dei flussi turistici dal lato della domanda, prendendo in esame sia il livello complessivo e la tendenza generale della domanda turistica, che il suo andamento stagionale. A tal fine vengono analizzati dati sulla domanda turistica a livello Europeo nell'ultimo decennio e viene proposto un focus specifico sulla domanda turistica in Italia, ponendo particolare enfasi sull'andamento stagionale. Una maggiore comprensione delle dinamiche della domanda turistica nonché degli impatti della stagionalità rappresentano un pre-requisito essenziale per l'implementazione delle politiche turistiche.*

Key words: Tourism Statistics, Tourist behavior, Seasonal pattern, Seasonal amplitude

Stefano De Cantis

Dipartimento di Scienze Economiche Aziendali e Statistiche, University of Palermo, Viale delle Scienze, ed.13 90128, Palermo, e-mail: stefano.decantis@unipa.it

Mauro Ferrante

Dipartimento di Culture e Società, University of Palermo, Viale delle Scienze, ed.15 90128, Palermo, e-mail: mauro.ferrante@unipa.it

Anna Maria Parroco

Dipartimento di Psicologia, University of Palermo, Viale delle Scienze, ed.15 90128, Palermo, e-mail: annamaria.parroco@unipa.it

1 Introduction

Country-specific measurement and the analysis of tourism activity of residents (both in terms of physical and economic volumes) allow us to describe the determinants of consumer behavior, to investigate motivations regarding the choices of making or not making tourism, and finally to estimate new tendencies in tourism related behaviors. In order to describe the main features of tourism demand at the country level, it is important to consider: a) the total level of tourism demand produced by each country (domestic and outgoing); b) the general tendency which, in the medium run, characterizes the demand; c) the seasonal fluctuations of tourism demand. In particular, seasonality plays an important role, since the same level of tourism demand could determine very different impacts according to its distribution over time. Starting from these premises, the present work aims at analyzing tourist flows from the demand side at the European level. In the European context, the European Regulation (EU) No 692/2011 [3] concerning European statistics on tourism aims at establishing a common framework in the European Union, concerning the collection of statistical information on tourism. The Regulation states that data to be transmitted by the Member States concerns: a) the participation in tourism and the characteristics of tourism trips and visitors, and b) the characteristics of same-day visits. Collected and integrated data is provided by Eurostat in the tourism section of the database available at Eurostat's website. Moreover, a special focus is given for the Italian case; thanks to the availability of micro-data on trips made by residents, the seasonal component of tourism demand is analyzed in detail and some synthetic measures of seasonal concentration are provided.

2 Analysis of tourism demand in Europe

In the European context, the level of tourism demand of residents in the European Union is largely determined by a relatively limited number of countries. Germany, France, UK, Italy, and Spain represent a relevant share of the EU's population, and, beyond Scandinavian countries, also have among the highest rates of tourism propensity. Germany presents the highest tourism demand with more than 300 millions trips made by residents in 2011, followed by France, the UK, Italy and Spain.

In order to offer a wider perspective of tourism demand at the European Level, in Tab. 2, Net Travel Propensity Index is reported for a set of European Countries for 2011, along with their number of trips and resident population values. Scandinavian countries (Norway, Sweden, and Finland) are those which present the highest Net Travel Propensity, but also Germany, with a value of about 60% in the third quarter, demonstrate a high travel propensity. On the contrary, countries such as Bulgaria, Poland, Portugal, and Romania are those in which the highest value of net travel propensity index does not exceed 25%. Moreover, although almost similar in terms of population, the five selected countries (Germany, France, UK, Italy, and Spain) are very different in terms of travel propensity (Tab. 1). First, European Countries

Table 1 Quarterly Net Travel Propensity Index for selected European countries (*data in thousands*), 2011

Country	Population	Quarterly number of resident travelers				Net Travel Propensity Index			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Bulgaria	7,369	369	668	1,170	624	5.0%	9.1%	15.9%	8.5%
Czech Republic	10,487	3,627	4,487	4,596	2,802	34.6%	42.8%	43.8%	26.7%
Denmark	5,561	2,560	2,736	3,181	2,726	46.0%	49.2%	57.2%	49.0%
Germany	81,752	32,897	42,750	47,208	39,569	40.2%	52.3%	57.7%	48.4%
Greece	11,123	1,019	1,824	3,253	1,287	9.2%	16.4%	29.2%	11.6%
Spain	46,667	7,902	11,478	15,624	9,035	16.9%	24.6%	33.5%	19.4%
France	64,979	21,012	26,550	34,169	23,560	32.3%	40.9%	52.6%	36.3%
Croatia	4,290	755	1,013	1,521	955	17.6%	23.6%	35.5%	22.3%
Italy	59,365	9,575	10,140	21,095	8,113	16.1%	17.1%	35.5%	13.7%
Latvia	2,075	317	484	623	353	15.3%	23.3%	30.0%	17.0%
Lithuania	3,053	573	911	1,210	761	18.8%	29.8%	39.6%	24.9%
Luxembourg	512	214	267	318	228	41.9%	52.2%	62.1%	44.5%
Hungary	9,986	1,584	2,067	2,787	1,886	15.9%	20.7%	27.9%	18.9%
Malta	415	72	78	117	56	17.4%	18.8%	28.3%	13.5%
Austria	8,375	2,528	3,342	4,652	2,854	30.2%	39.9%	55.5%	34.1%
Poland	38,530	4,610	5,830	9,560	4,807	12.0%	15.1%	24.8%	12.5%
Portugal	10,573	1,017	1,423	2,508	1,641	9.6%	13.5%	23.7%	15.5%
Romania	20,199	2,123	2,918	3,348	3,212	10.5%	14.4%	16.6%	15.9%
Slovenia	2,050	447	650	1,079	476	21.8%	31.7%	52.6%	23.2%
Slovakia	5,392	1,374	1,450	2,375	1,423	25.5%	26.9%	44.0%	26.4%
Finland	5,375	3,026	3,202	3,583	3,090	56.3%	59.6%	66.7%	57.5%
Sweden	9,416	5,411	6,360	6,871	5,883	57.5%	67.6%	73.0%	62.5%
United Kingdom	63,023	13,791	22,868	29,910	18,152	21.9%	36.3%	47.5%	28.8%
Norway	4,920	2,672	3,016	3,257	2,562	54.3%	61.3%	66.2%	52.1%

exhibit different levels of tourism demand, both in absolute and in relative terms. The causes of these differences in tourism propensity should be more deeply investigated, and can be related to economic and sociocultural differences, and with the country specific uses and habits related to tourist behaviors, which only partially have been discussed in academic literature [1, 2]. Second, the way in which tourism demand by residents is distributed during the year is also very different from one country to another with different seasonal variations both in terms of pattern and amplitude [4].

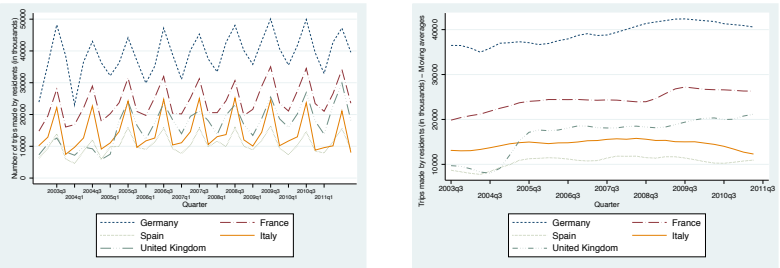


Fig. 1 Quarterly number of trips made by Residents in top five travel generating European countries, actual and moving averages series (data in thousands), 2003-2011.

3 Tourism demand of Italian residents from Istat 'Trips and holidays' survey

Istat, until 2014, annually presented the estimates of the main aggregates of tourism demand in Italy, based on CATI survey 'Trips and Holidays' which had been conducted on a quarterly basis from 1997 until 2013. Starting from 2014 a new survey on 'Consumption of Italian Families' has been introduced, which replaces and updates the survey on 'Trips and Holidays'. In Tab. 2, annual data on trips (by purpose and destination) made by Italian residents are reported, from 2005 to 2013. Also the number of nights spent, average length of trip, and the gross travel propensity index are reported. This data highlights the decrease in the number of trips and nights registered in the last years: from about 123 million trips in 2008 to less than 65 million in 2013, with a loss of about 60 million trips (-48.6%). Also in terms of nights, a considerable decrease can be observed: from more than 706 million nights in 2008 to about 417 million in 2013, with a loss of about 289 million nights (-41.0%). In other words, if in 2008 there were about 210 trips per 100 residents, in 2013 there were only 106. From this preliminary analysis, the strong effect of the economic crisis is evident on tourism, representing a relatively less-investigated phenomenon.

In order to isolate the seasonal component from other potential sources of variability in the series several methods could be used. For our application, we implemented the TRAMO-SEATS procedure, which allows seasonal factors to be derived. A summary of SARIMA models estimated for each series is reported in Tab. 3. Additive seasonal factors were produced for the series related to propensity indices related to both holiday and business trips in Italy, whereas multiplicative seasonal factors were derived for the series related to holiday and business trips abroad.

Once seasonal factors are derived, the cycle plot is a useful tool to synthesize the seasonal behavior of the series over all the considered periods. The series of seasonal factors related to travel propensity in Italy for holiday purposes seems to be one that presents a very high degree of stability of pattern of seasonality. Both the series related to holiday trips, in Italy and abroad, present almost the same pattern with a peak in August, and values of seasonal factors above the trend-cycle component only in summer months (from June to September). Several measures can be used in order to summarize the amplitude of seasonal fluctuations, some of which are reported in Tab. 4. All the amplitude measures indicate a relative stability of seasonal amplitude with a slight decrease of inequality in the last considered years.

Tab. 4 reveals that the overall level of seasonality in each series is relatively stable during all the years considered, with a slight decrease in seasonal amplitude for the series related to holiday trips in Italy, holiday trips abroad, and business trips in Italy, and a strong irregular behavior for the series related to holiday trips abroad.

Table 2 Number of trips, nights and related index of trips characteristics made by Italian residents in Italy and abroad, 2005-2013

Variable	2005	2006	2007	2008	2009	2010	2011	2012	2013
<i>Data in thousands</i>									
Holiday trips in Italy	77,860	78,606	80,972	90,463	82,265	71,926	59,807	54,733	46,062
Holiday trips abroad	14,268	15,284	16,201	16,347	16,412	15,524	12,751	13,967	11,389
Total trips (<i>all purposes</i>)	107,100	107,895	112,369	122,938	114,099	100,040	83,417	78,703	63,154
Nights spent in Italy (<i>holiday purposes</i>)	493,775	534,672	480,724	508,722	497,230	467,796	393,015	357,772	300,709
Nights spent abroad (<i>holiday purposes</i>)	123,003	133,119	146,267	135,374	125,351	118,250	101,757	113,828	100,946
Total nights (<i>all purposes</i>)	676,243	719,763	689,313	706,650	680,215	626,990	527,811	501,059	417,126
Population (1 st Jan)	57,875	58,064	58,224	58,653	59,001	59,190	59,365	59,394	59,685
Avg. duration of trips in Italy	6.34	6.80	5.94	5.62	6.04	6.50	6.57	6.54	6.53
Avg. duration of trips abroad	8.62	8.71	9.03	8.28	7.64	7.62	7.98	8.15	8.86
Avg. duration of trips	6.31	6.67	6.13	5.75	5.96	6.27	6.33	6.37	6.60
Gross travel propensity index in Italy (<i>holiday purposes</i>)	1.35	1.35	1.39	1.54	1.39	1.22	1.01	0.92	0.77
Gross travel propensity index abroad (<i>holiday purposes</i>)	0.25	0.26	0.28	0.28	0.28	0.26	0.21	0.24	0.19
Travel propensity index (<i>all purposes</i>)	1.85	1.86	1.93	2.10	1.93	1.69	1.41	1.33	1.06

Table 3 Summary of TRAMO-SEATS procedure on monthly travel propensity index series in Italy, by purposes and destination. Years 2002-2013

	Holiday trips in Italy	Business trips in Italy	Holiday trips abroad	Business trips abroad
Transformation	none	none	Log-transformation	Log-transformation
SARIMA Model	(0,1,1)(0,1,1)	(1,0,0)(0,1,1)	(0,1,3)(0,1,1)	(1,1,1)(1,0,0)
Parameter estimates	$\theta_1 = -0.7240$ $\theta_{12} = -0.8283$	$\phi_1 = -0.3629$ $\theta_{12} = -0.8435$	$\theta_1 = -0.8648$ $\theta_2 = 0.1459$ $\theta_3 = -0.1378$ $\theta_{12} = -0.7447$	$\phi_1 = -0.4643$ $\theta_1 = -0.9900$ $\phi_{12} = -0.4883$

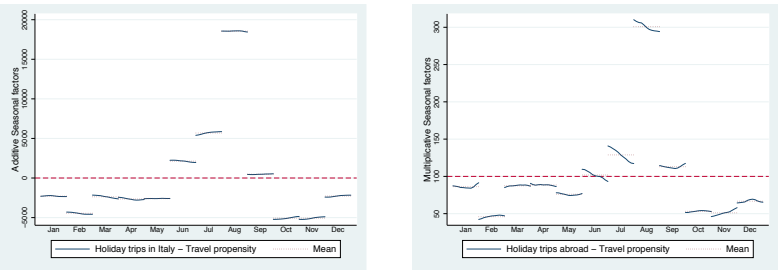


Fig. 2 Cycle plots of seasonal factors for travel propensity index, holiday trips in Italy and abroad, 2002-2013.

Table 4 Measures of amplitude derived from seasonal factors of Travel Propensity Index, by purpose and destination of trips, Italian residents 2002-2013

Year	Holiday trips in Italy	Business trips in Italy	Holiday trips abroad		Business trips abroad	
	Seasonal Range	Seasonal Range	Coefficient of Seasonal Variation	Gini Index	Coefficient of Seasonal Variation	Gini Index
2002	23,787.7	1,269.9	0.68	0.12	0.21	0.34
2003	23,792.1	1,268.3	0.67	0.11	0.20	0.33
2004	23,773.5	1,262.2	0.67	0.11	0.19	0.33
2005	23,741.6	1,249.6	0.67	0.16	0.26	0.32
2006	23,725.9	1,233.9	0.66	0.20	0.32	0.32
2007	23,699.0	1,221.4	0.65	0.19	0.33	0.31
2008	23,643.4	1,207.0	0.64	0.13	0.20	0.31
2009	23,584.8	1,187.5	0.64	0.10	0.17	0.30
2010	23,495.8	1,176.3	0.64	0.15	0.25	0.30
2011	23,397.0	1,178.0	0.64	0.35	0.61	0.30
2012	23,335.6	1,181.2	0.63	0.20	0.32	0.30
2013	20,708.7	958.8	0.71	0.30	0.38	0.24

4 Conclusion

European residents have very different tourism behaviour in terms of overall travel propensity and very different seasonal travel demands. Thanks to microdata obtained from the survey on tourism made by Italian residents, a detailed analysis of seasonality has been possible. Despite of the relevant decrease in tourism trips following the recent economic crisis, a strong and persistent seasonal behaviour characterizes the pattern of tourism trips in Italy. In social phenomena it is common to observe changes in human behaviours and attitudes, but, quite surprisingly, this is not the case. From the methodological perspective, this work used the Gini index, as well as other indices for synthesizing the seasonal burden, which, as pointed out by De Cantis et al. [4], do not take into account for the natural ordering of months. Subsequently, a deep analysis on the causes that determine the persistent seasonal behaviour, as well as critical reflections on the appropriate measurement of seasonal fluctuations merits further investigations from a variety of perspectives.

References

1. Alegre J, Mateo S, Pou L (2009) Participation in tourism consumption and the intensity of participation: an analysis of their socio-demographic and economic determinants. *Tourism Economics*, 15(3): 531–546.
2. Bernini C, Cracolici MF (2015) Demographic change, tourism expenditure and life cycle behaviour. *Tourism Management*, 47: 191–205.
3. European Parliament (2011) Regulation (EU) No 692/2011. *Official Journal of the European Union*, L192(54), 17–32.
4. De Cantis S, Ferrante M, Vaccina F (2011) Seasonal Pattern and amplitude – a logical framework to analyse seasonality in tourism: an application to bed occupancy in Sicilian hotels. *Tourism Economics*, 17(3): 655–675.