



Book of the Short Papers

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



UNIVERSITÀ
POLITECNICA
DELLE MARCHE



LIUC | BUSINESS
ANALYTICS AND
DATA SCIENCE HUB
Università Cattaneo



CHAIRS

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

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

PROGRAM COMMITTEE

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

LOCAL ORGANIZING COMMITTEE

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

ORGANIZERS OF INVITED SESSIONS

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

FURTHER PEOPLE OF LOCAL ORGANIZING COMMITTEE

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

Technical support: Matteo Mercuri, Maila Ragni, Daniele Ripanti

Copyright © 2023

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891935618AAVV

Contents

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

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

Statistical Machine Learning for environmental applications

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

Statistical Process Monitoring for Complex Data in Industry 4.0

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

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

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

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

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

Population Dynamics, Climate Change and Sustainability

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

Statistical Learning for health research and omics data

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

The Economic behaviour of Sustainability

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

Advances in statistical methods for complex problems

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

Explainable machine learning models

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

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

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

- Analysis of University Grades: An IRT Model for Responses and Response Times with Censoring 186
Michela Battauz
- Predicting high schools' students performances with registry's data: a machine learning approach 191
Lidia Rossi, Marta Cannistrà and Tommaso Agasisti
- Using response times to identify cheaters in CAT: A simulation study 195
Luca Bungaro, Bernard P. Veldkamp and Mariagiulia Matteucci

Spatial and Spatio-Temporal Modeling: Theory and Applications

- A geostatistical investigation of the ammonia-livestock relationship in the Po Valley, Italy 200
Paolo Maranzano, Kelly McConville , Philipp Otto and Felicetta Carillo
- Bayesian multi-species N-mixture models for large scale spatial data in community ecology 206
Michele Peruzzi
- Minimum contrast for point processes' first-order intensity estimation 212
Nicoletta D'Angelo and Giada Adelfio

Statistical Framework for Measuring the Sustainability of Tourism

- Data validity and statistical conformity with Benford's Law: the case of tourism in Sicily 217
Roy Cerqueti and Davide Provenzano
- Exploring the level of digitalization of the Italian museums through a multilevel ordered logit model 228
Claudia Cappello, Sabrina Maggio and Sandra De Iaco
- Functional Partial Least-Squares via Regression Splines. An application on Italian Sustainable Development Goals data 232
Ida Camminatiello, Rosaria Lombardo, Jean-Francois Durand and Leonardo S. Alaimo

Statistical learning for well-being analysis

- Assessing multidimensional poverty of the Italian provinces during Covid-19: a small area estimation approach 238
Mariateresa Ciommi, Chiara Gigliarano, Francesca Mariani and Gloria Polinesi
- The fuzzy set approach as statistical learning for the analysis of multidimensional well-being 244
Gianni Betti, Federico Crescenzi, Antonella D'Agostino and Laura Neri
- What Makes a Satisfying Life? Prediction and Interpretation with Machine-Learning Algorithms n.a.
Conchita D'Ambrosio

Bayesian contributions to Statistical Learning

A Bayesian framework for early cancer screening Sally Paganin and Jeff Miller	249
Imputing Synthetic Pseudo Data from Aggregate Data: Development and Validation for Precision Medicine Cecilia Balocchi	n.a.
Linear models with assumptions-free residuals: a Bayesian Nonparametric approach Filippo Ascolani and Valentina Ghidini	254

Data Visualization for Smart Insights and Advanced Predictive Analytics

Applications of data visualization for industry Martina Dossi, Stefano Sangaletti, Marilena Di Bari and Federica Bruschini	259
Some Notes on the Use of the Circular Boxplot Giovanni Camillo Porzio and Davide Buttarazzi	n.a.
TERRA: a smart visualization tool for international trade in goods statistics Francesco Amato, Mauro Bruno and Maria Serena Causo	265

Methods for the analysis of distributional data

Clustering of Distributional Data based on LDQ transformation Gianmarco Borrata and Rosanna Verde	271
Dynamic learning from data streams through the combined use of probability density functions and simplicial functional principal component analysis Francesca Fortuna, Fabrizio Maturo and Tonio Di Battista	276
Multivariate Parametric Analysis of Distributional Data Paula Brito	n.a.

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

Labor Market Return to Refugees' Human Capital Investment: A Natural Experiment in Sweden Eleonora Mussino	n.a.
Social networks and loneliness among older migrants in Italy Viviana Amati, Eralba Cela and Elisa Barbiano di Belgiojoso	282
The Italian Decree on Security: An Analysis of the Impact on Asylum Applications Giorgio Piccitto	287

Modelling and Forecasting High-dimensional time series

Adaptive combinations of tail-risk forecasts Alessandra Amendola, Vincenzo Candila, Antonio Naimoli and Giuseppe Storti	293
Are Monetary Policy Announcements related to Volatility Jumps? Giampiero Gallo, Demetrio Lacava and Edoardo Otranto	299

Regularized Estimation and Prediction of the El Nino/Southern Oscillation Cycle n.a.
 Alessandro Giovannelli and Tommaso Proietti

3 Contributed Sessions 305

Bayesian nonparametric methods

Bayesian density estimation for modeling age-at-death distribution 306
 Davide Agnoletto, Tommaso Rigon and Bruno Scarpa

Bayesian mixing distribution estimation in the Gaussian-smoothed 1-Wasserstein distance 311
 Catia Scricciolo

Bayesian nonparametric estimation of heterogeneous intrinsic dimension via product partition models 316
 Francesco Denti, Antonio Di Noia and Antonietta Mira

Bayesian nonparametric multiple change point detection for time series of compositional data 322
 Edoardo Marchionni and Riccardo Corradin

Galton-Watson process: a non parametric prior for the offspring distribution 328
 Massimo Cannas, Michele Guindani and Nicola Piras

Hierarchical processes in survival analysis 333
 Riccardo Cogo, Federico Camerlenghi and Tommaso Rigon

Economics and Statistics

A regression analysis for count data to investigate the effectiveness of incentives on the adoption of 4.0 technologies 339
 Stefano Bonnini and Michela Borghesi

Statistical analysis on SDGs indicators related to environmental sustainability 344
 Najada Firza, Anisa Bakiu and Dante Mazzitelli

Empowering futures adopting a spatial convergence of opinions: a Real-Time Spatial Delphi approach 349
 Yuri Calleo, Simone Di Zio and Francesco Pilla

Stocks price forecasts using Stochastic Differential Equations: an empirical assessment 355
 Dario Frisardi and Matteo Spuri

The Added-Worker Effect within Italian Households 361
 Donata Favaro and Anna Giraldo

Health statistics 1

A model for the natural history of breast cancer: application to a Norwegian screening dataset 365
 Laura Bondi, Marco Bonetti and Solveig Hofvind

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

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

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

Mixture Models

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

Sampling methods and analysis of survey data

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

Space-time statistics

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

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

Environmental learning and indicators

Assessing the performance of nuclear norm-based matrix completion methods on CO ₂ emissions data	776
Rodolfo Metulini, Francesco Biancalani, Giorgio Gnecco and Massimo Riccaboni	
Deep Learning for smart and sustainable agriculture	782
Amalia Vanacore, Armando Ciardiello, Annalisa Izzo, Pierdomenico Zaffino, Carolina Vecchio, Gennaro Pio Auricchio and Luigi Uccelli	
Do green transition, environmental taxes and renew-able energy promote ecological sustainability in G7 countries? Evidence from panel quantile regression	788
Aamir Javed, Agnese Rapposelli and Asif Javed	
Doubly Robust DID for National Parks evaluation: “just” environmental benefits, or socioeconomic impacts as well?	795
Riccardo D’Alberto, Francesco Pagliacci and Matteo Zavalloni	
On the gap between emitted and absorbed carbon dioxide. Are trees enough to save us?	801
Lorenzo Mori and Maria Rosaria Ferrante	
Small scale analysis of energy vulnerability in the municipality of Palermo	806
Giuliana La Mantia	

Health statistics 2

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

Migrants in Italy and return migration

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

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

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

The determinants of leaving the parental home in Italy: 2012-18 1003
Ilaria Rocco and Gianpiero Dalla Zuanna

Environmental modeling

A Bayesian weather-driven spatio-temporal model for PM10 in Lombardy 1109
Michela Frigeri, Alessandra Guglielmi and Giovanni Lonati

A preliminary study on shape descriptors for the characterization of microplastics
ingested by fish 1015
Greta Panunzi, Tommaso Valente, Marco Matiddi and Giovanna Jona Lasinio

Artificial neural network in predicting odour concentrations: a case study 1021
Veronica Distefano and Gideon Mazuruse

Bayesian analysis of PM10 concentration by spatio-temporal ARIMA and STS
models 1026
Michela Frigeri and Ilenia Epifani

Functional ANOVA to monitor yearly Adriatic sea temperature variations 1032
Annalina Sarra, Adelia Evangelista, Tonio Di Battista and Nicola Di Deo

New perspectives in the measurement of biodiversity 1038
Linda Altieri, Daniela Cocchi and Massimo Ventrucci

Multivariate data analysis 2

Feature Selection via anomaly detection autoencoders in radiogenomics studies 1044
Alessia Mapelli, Michela Carlotta Massi, Nicola Rares Franco, Francesca Ieva,
Catharine West, Petra Seibold, Jenny Chang-Claude and the REQUITE and RADprecise
Consortia

Further considerations on the Spectral Information Criterion 1050
Luca Martino

How to increase the power of the test in sparse contingency tables: a simulation
study 1057
Federica Nicolussi and Manuela Cazzaro

Latent event history models for quasi-reaction systems 1063
Matteo Framba, Veronica Vinciotti and Ernst Wit

Quantile-based graphical models for continuous and discrete variables 1069
Luca Merlo, Marco Geraci and Lea Petrella

The logratio Student t distribution 1075
Gianna Monti and Gloria Mateu-Figueras

Statistics in Society 2

A decomposition of the changes in tourism demand in Tuscany over the 2019-2021
period 1079
Mauro Mussini

Bayesian networks as a territorial gender impact assessment tool 1084
Flaminia Musella, Lorenzo Giammei, Fulvia Mecatti and Paola Vicar

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

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

Sequential marginal likelihood selection for the estimation of sparse correlation matrices	1246
Claudia Di Caterina and Davide Ferrari	

Nonparametric statistical methods

A Comparison of Distribution-Free Control Charts	1252
Michele Scagliarini	
Characterizing Heterogeneity of Causal Effects in Air Pollution in Florida	1257
Dafne Zorzetto	
Comparing three robust procedures for CANDECOMP/PARAFAC estimation	1262
Valentin Todorov, Violetta Simonacci, Michele Gallo and Nikolay Trendafilov	
How active is a genetic pathway? Comparative analysis of post-hoc permutation-based methods	1268
Anna Vesely and Angela Andreella	
Non Parametric Combination methodology: a literature review on recent developments	1274
Elena Barzizza, Nicolò Biasetton and Riccardo Ceccato	

Regression modeling

A Quantile Regression Model to Evaluate the Performance of the Italian Courts of Law	1280
Carlo Cusatelli, Massimiliano Giacalone and Eugenia Nissi	
A variable selection procedure based on predictive ability: a preliminary study on logistic regression	1285
Rosaria Simone and Mariarosaria Coppola	
Comparison of binary regressions with asymmetric link function for imbalanced data	1291
Michele La Rocca, Marcella Niglio and Marialuisa Restaino	
New advances in Regression Forests	1297
Mila Andreani, Lea Petrella and Nicola Salvati	
On the Optimal Non-Convexity of Penalty in Sparse Regression Models	1303
Daniele Cuntrera, Vito M.R. Muggeo and Luigi Augugliaro	
Using expectile regression with latent variables for digital assets	1309
Beatrice Foroni, Luca Merlo and Lea Petrella	

4 Program 1315

Preface

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

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

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

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

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

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

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

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

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

Using expectile regression with latent variables for digital assets

Beatrice Foroni^a, Luca Merlo^b, and Lea Petrella^a

^aMEMOTEF Department, Sapienza University of Rome; beatrice.foroni@uniroma1.it,
lea.petrella@uniroma1.it

^bDepartment of Human Sciences, European University of Rome; luca.merlo@unier.it

Abstract

In this paper we introduce a linear expectile hidden Markov model with the goal of modeling the entire conditional distribution of asset returns and, at the same time, to grasp unobserved serial heterogeneity and rapid volatility jumps typical of financial time series. The temporal evolution of asset returns is captured by introducing time-dependent coefficients evolving according to a latent discrete homogeneous Markov chain. To implement the procedure, we consider the Asymmetric Normal distribution as a working likelihood for the estimation of model parameters and the estimation procedure is carried out using an efficient EM algorithm. The empirical application investigates the relationship between daily Bitcoin returns and major world market indices.

Keywords: Bitcoin, financial time series, hidden Markov models, tail risk

1. Introduction

The recent exploit of the cryptocurrency market have attracted investors and risk managers as never before. The enormous price jumps and levels of high volatility of these digital assets, and Bitcoin in particular, caused by speculative behaviors have threatened the stability of financial markets (19), and research and practitioners have recently started to investigate the peculiar characteristics of cryptocurrencies, as well as their relationship with tradition financial markets. Many contributions to this strand of literature rely on well-known econometric techniques such as GARCH models (9), variance decomposition (6; 21) and Granger causality test (4). Rather than investigating measures of conditional central tendencies, fewer works have focused on the tails of returns distribution. For instance, (15) focused on the tail connectedness among major cryptocurrencies in extreme downward and upward market conditions using LASSO penalized quantile regressions, while (22) apply a risk spillover approach based on generalized quantiles, showing the existence of a downside risk spillover between Bitcoin and traditional assets. From a risk management perspective, it is of extreme importance to be able to investigate the dynamics of extreme occurrences. Since the seminal work of (10), quantile regression has represented one of the most used approaches for modeling the entire distribution of returns while accounting for the well-known stylized facts, i.e., high kurtosis, skewness and serial correlation, that typically characterize financial assets. Since then, several generalizations of the concept of quantiles have been presented, among which we find expectile regression (14), which, similar to quantile regression, allows to describe the entire conditional distribution of a response variable based on an asymmetric squared loss function. Despite having a more difficult interpretation, expectiles possess several advantages, both from an informative and a computational point of view. In particular, the asymmetric squared loss is continuously

differentiable, which makes the estimators and their covariance matrix easier to compute using fast and efficient algorithms. In the context of risk management, expectiles have gained an important role as potential competitors to the Value at Risk (VaR) and the Expected Shortfall measures, and indeed possess several interesting properties in terms of risk measures, being the only risk measure that is both coherent and elicitable (11; 23). However, homogeneous regression models are not able to capture the volatility clustering behavior that often financial time series exhibit. In this context, hidden Markov models (HMMs) have been intensively employed to characterize temporal evolution of returns distribution, modeling volatility regime shifting through a latent Markov chain. Since (8), different works have combined the quantile framework with HMMs by introducing in the model parameters that vary according to the outcome of a latent Markov process (20; 12; 13). To the best of our knowledge, however, an expectile hidden Markov regression model has not been explored yet. In this paper we introduce an expectile regression model to analyze the entire conditional distribution of Bitcoin returns where the dynamics of returns over time is described by state-specific regression coefficients which follow a latent discrete homogeneous Markov chain. The proposed model contributes to the existing literature regarding the relations between cryptocurrencies and traditional asset class to control for potential inherent risks related to the participation in crypto exchanges. As usual for latent variable models, inference is carried out in a Maximum Likelihood (ML) approach using an Expectation-Maximization (EM) algorithm based on the asymmetric normal distribution of (17) as working likelihood. In the empirical analysis, we model daily Bitcoin log-returns as a function of major stock and global market indices, including Crude Oil, Standard & Poor's 500 (S&P500), Gold COMEX daily closing prices and the Volatility Index (VIX) from September 2014 until October 2022.

The rest of the paper is organized as follows. In Sect. 2. we specify the proposed model with the EM algorithm for estimating the model parameters and the computational aspects. Sect. 3. discusses the results obtained and concludes.

2. Model Specification and Inference

In this section we describe the proposed expectile hidden Markov regression model. Formally, let $\{S_t\}_{t=1}^T$ be a latent, homogeneous, first-order Markov chain defined on the discrete state space $\{1, \dots, K\}$. Let $\pi_k = Pr(S_1 = k)$ be the initial probability of state k , $k = 1, \dots, K$, and $\pi_{k|j} = Pr(S_{t+1} = k | S_t = j)$, with $\sum_{k=1}^K \pi_{k|j} = 1$ and $\pi_{k|j} \geq 0$, denote the transition probability between states j and k , that is, the probability to visit state k at time $t + 1$ from state j at time t , $j, k = 1, \dots, K$ and $t = 1, \dots, T$. More concisely, we collect the initial and transition probabilities in the K -dimensional vector $\boldsymbol{\pi}$ and in the $K \times K$ matrix $\boldsymbol{\Pi}$, respectively. To build the proposed model, let Y_t denote a continuous observable response variable and $\mathbf{X}_t = (1, X_{t2}, \dots, X_{tP})'$ be a vector of P exogenous covariates, with the first element being the intercept, at time $t = 1, \dots, T$.

For a given expectile level $\tau \in (0, 1)$, the proposed linear expectile hidden Markov model is defined as follows:

$$Y_t = \mathbf{X}_t' \boldsymbol{\beta}_k(\tau) + \epsilon_{tk}(\tau), \quad (1)$$

where $\mu_{tk} = \mathbf{X}_t' \boldsymbol{\beta}_k(\tau)$ defines the linear expectile model, $\boldsymbol{\beta}_k(\tau) = (\beta_{1k}(\tau), \dots, \beta_{Pk}(\tau))' \in \mathbb{R}^P$ is the state-specific coefficient vector that assumes one of the values $\{\boldsymbol{\beta}_1(\tau), \dots, \boldsymbol{\beta}_K(\tau)\}$ depending on the outcome of the unobservable Markov chain S_t and $\epsilon_{tk}(\tau)$ is the error term whose conditional τ -th expectile is assumed to be zero. When $\tau = \frac{1}{2}$, expectile regression reduces to the standard mean regression while, when $\tau \neq \frac{1}{2}$, the regression targets the entire conditional distribution of the response given the covariates. The estimation of the model parameters is carried on through a Maximum Likelihood approach. We employ the Asymmetric Normal (AN) distribution, originally introduced by (16), to describe the conditional distribution of the response given covariates and the state occupied by the latent process at time t , whose probability density function is given by

$$f_Y(y_t | \mathbf{X}_t = \mathbf{x}_t, S_t = k) = \frac{2\sqrt{\tau(1-\tau)}}{\sqrt{\pi\sigma_k^2(\sqrt{\tau} + \sqrt{1-\tau})}} \exp \left[-\omega_\tau \left(\frac{y_t - \mu_{tk}}{\sigma_k} \right) \right], \quad (2)$$

where $\omega_\tau(\cdot)$ is the expectile loss function defined as $\omega_\tau(u) = u^2|\tau - \mathbb{I}(u < 0)|$, which assigns weights τ and $1 - \tau$ to positive and negative deviations, respectively, and $\mathbb{I}(\cdot)$ denotes the indicator function. The location parameter μ_{tk} is defined by the linear model $\mu_{tk} = \mathbf{x}'_t \boldsymbol{\beta}_k(\tau)$ and corresponds to the τ -th expectile, $\sigma_k > 0$ is a scale parameter and $\tau \in (0, 1)$ determines the asymmetry of the distribution. Particularly, when $\tau = \frac{1}{2}$ the density in eq. (2) reduces to the well-known normal distribution, and μ_{tk} and σ_k coincide with its mean and standard deviation, respectively. The use of this distribution is deemed to be as a likelihood inferential tool for estimating the model parameters in a regression framework rather a parametric assumption. As common for latent variable models, and HMMs in particular, inference on model parameters is made through the development of an EM algorithm (1). To ease the notation, unless specified otherwise, hereinafter we omit the expectile level τ , yet all model parameters are allowed to depend on it. The complete log-likelihood of the proposed model is defined as follows for a given number of hidden states K :

$$\begin{aligned} \ell_c(\boldsymbol{\theta}_\tau) = & \sum_{k=1}^K \gamma_1(k) \log \pi_k + \sum_{t=1}^T \sum_{k=1}^K \sum_{j=1}^K \xi_t(j, k) \log \pi_{k|j} \\ & + \sum_{t=1}^T \sum_{k=1}^K \gamma_t(k) \log f_Y(y_t | \mathbf{x}_t, S_t = k), \end{aligned} \quad (3)$$

where $\boldsymbol{\theta}_\tau = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K, \sigma_1, \dots, \sigma_K, \boldsymbol{\pi}, \boldsymbol{\Pi})$ represents the vector of all model parameters, $\gamma_t(k)$ denotes a dummy variable equal to 1 if the latent process is in state k at occasion t and 0 otherwise, and $\xi_t(j, k)$ is a dummy variable equal to 1 if the process is in state j in $t - 1$ and in state k at time t and 0 otherwise.

To estimate $\boldsymbol{\theta}_\tau$, the algorithm iterates between the E- and M-steps until convergence, as briefly showed in what follows. In the E-step, at the generic $(h + 1)$ -th iteration, the unobservable indicator variables $\gamma_t(k)$ and $\xi_t(j, k)$ in eq. (3) are replaced by their conditional expectations given the observed data and the current parameter estimates $\boldsymbol{\theta}_\tau^{(h)}$. To compute such quantities one can use the Forward-Backward algorithm of (18). Then, we use these to calculate the conditional expectation of the complete log-likelihood function in eq. (3) given the observed data and the current estimates. In particular, in the M-step update of the regression coefficients is obtained by using Iteratively Reweighted Least Squares for cross-sectional data with appropriate weights.

The EM algorithm is initialized by assigning the initial states partition, $\{S_t^{(0)}\}_{t=1}^T$, to a Multinomial distribution with probabilities $1/K$. From the generated partition, the elements of $\boldsymbol{\Pi}^{(0)}$ are computed as proportions of transition, while we obtain $\boldsymbol{\beta}_k^{(0)}$ and $\sigma_k^{(0)}$ by fitting mean regressions on the observations within state k . A multiple random starts strategy is adapted to deal with the possibility of multiple roots. Once we computed the ML estimate of the model parameters, to estimate the standard errors we employ a parametric bootstrap scheme, refitting the model to R bootstrap samples and approximating the standard error of each model parameter with the corresponding standard deviation of the bootstrap estimates.

3. Main Results and Conclusions

The empirical analysis is based on the log-returns of Bitcoin, Crude Oil, S&P500, Gold COMEX daily closing prices and the VIX from September 2014 to October 2022. We consider the following model with the idea of providing insights into the temporal evolution of Bitcoin returns and its relationship with traditional global financial assets

$$\mu_{tk}^{Bitcoin} = \beta_{1k}(\tau) + \beta_{2k}(\tau)r_t^{Crude\ Oil} + \beta_{3k}(\tau)r_t^{S\&P500} + \beta_{4k}(\tau)r_t^{Gold} + \beta_{5k}(\tau)r_t^{VIX}, \quad (4)$$

with $\mu_{tk}^{Bitcoin}$ corresponding to the τ -th conditional expectile of Bitcoin return at time t in state k , while $r_t^{Crude\ Oil}$ denotes the return of the same date for Crude Oil, and similarly for the other indices. We fit the proposed model for two values of K , representing high and low volatility market conditions, at three expectile levels $\tau = \{0.10, 0.50, 0.90\}$, which allow us to focus on both downside and upside

risks. For the selected models, we report the clustering results in Figure 1 at $\tau = 0.50$ expectile level. The plot shows the time series of Bitcoin daily returns colored according to the estimated posterior probability of class membership, $\max_k \gamma_t(k)$, with the vertical dashed lines representing globally relevant events such as the Chinese stock market crash in 2015, the cryptocurrencies crash at the beginning of 2018, the COVID-19 market crash in March 2020, Biden’s election at the USA presidency in November 2020 and the Russian invasion of Ukraine at the beginning of 2022. Here we clearly see that the latent components can be associated to specific market regimes characterized by low and high volatility periods. Specifically, light-blue points (State 1) tend to identify low returns, while dark-blue ones (State 2) correspond to periods of extreme positive and negative returns. Table 1 shows the parameter estimates along with the standard errors (in brackets) computed by using the parametric bootstrap technique over $R = 1000$ resamples, as illustrated in Sec. 2. As it happens in the quantile regression framework, the state-specific intercepts are increasing somewhat with τ , with State 1 having lower values than State 2 for all τ 's. Moving forward with the analysis, at $\tau = 0.50$ we observe none or few interactions among Bitcoin and financial assets, during low and high volatility states, respectively. In particular, S&P500 and Gold significantly influence the mean of Bitcoin only in the second state, highlighting a weak hedge behavior of the crypto-asset during tranquil periods and confirming results founded in (2; 5). If we move to the tails of return distributions, in the not-at-risk state (State 1) at the extreme left-tail ($\tau = 0.10$) the S&P500, Gold and the VIX index positively influence Bitcoin returns, while only S&P500 and Gold significantly influence the right-tail ($\tau = 0.90$) expectiles of the cryptocurrency, exposing a connection during high volatility periods between traditional financial markets and Bitcoin both for negative and positive returns. In the at-risk state (State 2) we observe a positive influence of the S&P500 and Gold across the conditional distribution of returns. Also, one can see that Crude Oil is negatively associated with the crypto returns at the 10-th expectile. This finding is in line with (3) but it is contrary to the works of (7) and (6), which may be due to the events and crises occurred in the last years.

In conclusion, we developed a linear expectile hidden Markov model for the analysis of time series where temporal behaviors of the data are captured via time-dependent coefficients following an unobservable discrete homogeneous Markov chain. The proposed method enables us to model the entire conditional distribution of asset returns and, at the same time, to grasp unobserved serial heterogeneity and rapid volatility jumps that would otherwise go undetected. With this model we strengthen the existing literature in this field, contributing towards a deeper understanding of the interrelations between Bitcoin and traditional financial markets.

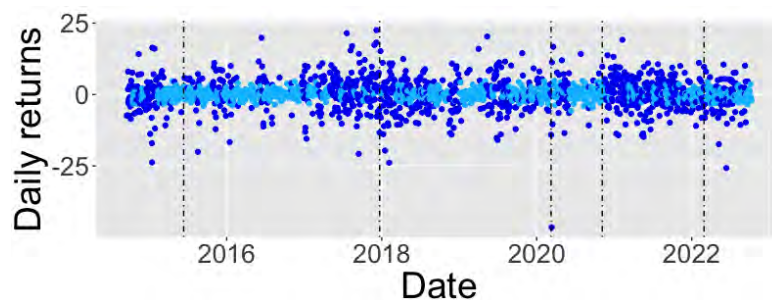


Figure 1: Bitcoin returns series classified according to the estimated posterior probability of class membership at $\tau = 0.50$. Vertical dashed lines indicate globally relevant events in the financial markets that occurred in 2015,06; 2017,12; 2020,03; 2020,11; and 2022,02.

	Intercept	Crude Oil	S&P500	Gold	VIX	σ_k
State 1						
$\tau = 0.10$	-1.036 (0.280)	0.024 (0.021)	0.595 (0.096)	0.189 (0.072)	0.029 (0.012)	1.433 (0.040)
$\tau = 0.50$	0.122 (0.158)	0.031 (0.072)	0.409 (0.383)	0.263 (0.249)	0.009 (0.036)	1.695 (0.062)
$\tau = 0.90$	1.297 (0.061)	-0.009 (0.020)	0.589 (0.088)	0.134 (0.065)	0.014 (0.011)	1.335 (0.041)
State 2						
$\tau = 0.10$	-6.52 (0.060)	-0.256 (0.096)	2.072 (0.476)	1.032 (0.320)	-0.055 (0.058)	4.964 (0.157)
$\tau = 0.50$	0.242 (0.092)	-0.056 (0.055)	1.087 (0.357)	0.613 (0.214)	-0.025 (0.026)	6.164 (0.169)
$\tau = 0.90$	6.244 (0.229)	0.017 (0.079)	0.948 (0.291)	0.835 (0.249)	-0.002 (0.041)	4.692 (0.128)

Table 1: State-specific parameter estimates for three expectile levels, with bootstrapped standard errors (in brackets) obtained over 1000 replications. Point estimates are displayed in boldface when significant at the standard 5% level.

References

- [1] Baum, L.E., Petrie, T., Soules, G., Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. of Math. Stat.* **41**(1), 164–171 (1970)
- [2] Bouri, E., Jalkh, N., Molnár, P., Roubaud, D.: Bitcoin for energy commodities before and after the december 2013 crash: diversifier, hedge or safe haven? *Appl. Econ.* **49**(50), 5063–5073 (2017)
- [3] Bouri, E., Lucey, B., Roubaud, D.: Cryptocurrencies and the downside risk in equity investments. *Fin. Res. Lett.* **33**, 101,211 (2020)
- [4] Bouri, E., Lucey, B., Roubaud, D.: The volatility surprise of leading cryptocurrencies: Transitory and permanent linkages. *Fin. Res. Lett.* **33**, 101,188 (2020)
- [5] Bouri, E., Molnár, P., Azzi, G., Roubaud, D., Hagfors, L.I.: On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? *Fin. Res. Lett.* **20**, 192–198 (2017)
- [6] Corbet, S., Meegan, A., Larkin, C., Lucey, B., Yarovaya, L.: Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Econ. Lett.* **165**, 28–34 (2018)
- [7] Dyhrberg, A.H.: Hedging capabilities of Bitcoin. Is it the virtual gold? *Fin. Res. Lett.* **16**, 139–144 (2016)
- [8] Farcomeni, A.: Quantile regression for longitudinal data based on latent markov subject-specific parameters. *Stat. and Comput.* **22**(1), 141–152 (2012)
- [9] Guesmi, K., Saadi, S., Abid, I., Ftiti, Z.: Portfolio diversification with virtual currency: Evidence from Bitcoin. *Int. Rev. of Fin. Anal.* **63**, 431–437 (2019)
- [10] Koenker, R., Bassett, G.: Regression quantiles. *Econ.: J. of the Econ. Soc.* **46**(1), 33–50 (1978)
- [11] Lambert, N.S., Pennock, D.M., Shoham, Y.: Eliciting properties of probability distributions. In: *Proc. of the 9th ACM Conf. on Electron. Commer.*, pp. 129–138. ACM (2008)
- [12] Maruotti, A., Petrella, L., Sposito, L.: Hidden semi-Markov-switching quantile regression for time series. *Comp. Stat. & Data Anal.* **159**, 107,208 (2021)
- [13] Merlo, L., Maruotti, A., Petrella, L., Punzo, A.: Quantile hidden semi-Markov models for multivariate time series. *Stat. and Comp.* **32**(4), 1–22 (2022)
- [14] Newey, W.K., Powell, J.L.: Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society* pp. 819–847 (1987)
- [15] Shahzad, S.J.H., Bouri, E., Ahmad, T., Naeem, M.A.: Extreme tail network analysis of cryptocurrencies and trading strategies. *Fin. Res. Lett.* **44**, 102,106 (2022)
- [16] Waldmann, E., Sobotka, F., Kneib, T.: Bayesian geoadditive expectile regression. *arXiv preprint arXiv:1312.5054* (2013)
- [17] Waldmann, E., Sobotka, F., Kneib, T.: Bayesian regularisation in geoadditive expectile regression. *Stat. and Comp.* **27**(6), 1539–1553 (2017)
- [18] Welch, L.R.: Hidden Markov models and the Baum-Welch algorithm. *IEEE Inf. Theory Soc. Newsl.* **53**(4), 10–13 (2003)
- [19] Yarovaya, L., Brzeszczyński, J., Lau, C.K.M.: Intra-and inter-regional return and volatility spillovers across emerging and developed markets: Evidence from stock indices and stock index

- futures. *Intern. Rev. of Fin. Anal.* **43**, 96–114 (2016)
- [20] Ye, W., Zhu, Y., Wu, Y., Miao, B.: Markov regime-switching quantile regression models and financial contagion detection. *Insur.: Math. and Econ.* **67**, 21–26 (2016)
- [21] Yi, S., Xu, Z., Wang, G.J.: Volatility connectedness in the cryptocurrency market: Is Bitcoin a dominant cryptocurrency? *Intern. Rev. of Fin. Anal.* **60**, 98–114 (2018)
- [22] Zhang, Y.J., Bouri, E., Gupta, R., Ma, S.J.: Risk spillover between bitcoin and conventional financial markets: An expectile-based approach. *The N. Am. J. of Econ. and Fin.* **55**, 101,296 (2021)
- [23] Ziegel, J.F.: Coherence and elicibility. *Math. Fin.* **26**(4), 901–918 (2016)