PROGRAMME AND ABSTRACTS

16th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computational and Methodological Statistics (CMStatistics 2023)

http://www.cmstatistics.org/CMStatistics2023

and

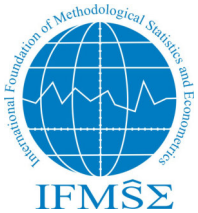
17th International Conference on

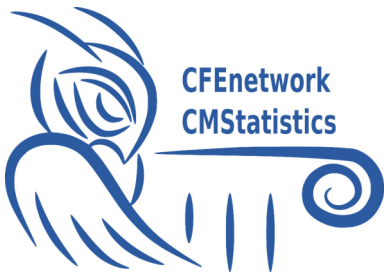
Computational and Financial Econometrics (CFE 2023)

http://www.cfenetwork.org/CFE2023

HTW Berlin, University of Applied Sciences, Germany

16–18 December 2023





ISBN 978-9925-7812-7-0

International Organizing Committee:
Ana Colubi, Erricos Kontoghiorghes and Manfred Deistler.

CFE 2023 Co-chairs:
Christina Erlwein-Sayer, Joshua Chan, Joann Jasiak and Carsten H.Y. Chong.

CFE 2023 Programme Committee:

Alessandra Amendola, Anindya Banerjee, Massimiliano Caporin, Gianluca Cubadda, Luca De Angelis, Timo Dimitriadis, Christian Francq, Deborah Gefang, Domenico Giannone, Niko Hauzenberger, Alain Hecq, Rustam Ibragimov, Florian Ielpo, Juan-Angel Jimenez-Martin, Gary Koop, Robinson Kruse-Becher, Emese Lazar, Degui Li, Laura Liu, Johan Lyhagen, Svetlana Makarova, Alexander Meyer-Gohde, Antonio Montanes, Carlos Montes-Galdon, Markus Pelger, Alla Petukhina, Tommaso Proietti, Artem Prokhorov, Anders Rahbek, Anindya Roy, Esther Ruiz, Willi Semmler, Etsuro Shioji, Katja Smetanina, Andrej Srakar, Genaro Sucarrat, Anastasija Tetereva, Martin Wagner, Toshiaki Watanabe, Ralf Wilke and Saeed Zaman.

CMStatistics 2023 Co-chairs:
Yanrong Yang, Armelle Guillou, Zhaoyuan Li and Christopher Hans.

CMStatistics 2023 Programme Committee:

Jesus Arroyo, Andriette Bekker, Veronica Berrocal, Annamaria Bianchi, Ana Bianco, Pier Giovanni Bissiri, Natalia Bochkina, Scott Bruce, Maddalena Cavicchioli, Peter Craigmile, Marzia Cremona, Rosa Crujeiras, Michael Daniels, Miguel De Carvalho, Thorsten Dickhaus, Dennis Dobler, Charles Doss, M. Brigida Ferraro, Stathis Gennatas, Sabrina Giordano, Stephane Girard, Rajarshi Guhaniyogi, Roman Hornung, Xianzheng Huang, Piotr Jaworski, Nadja Klein, John Kornak, Marie Kratz, Johannes Lederer, Christophe Ley, Tianxi Li, Monia Lupparelli, Jinchi Lv, Vince Lyzinski, Maria Francesca Marino, Cristina Mollica, Erica Moodie, Kalliopi Mylona, Thomas Opitz, Philipp Otto, Elisa Perrone, Eugen Pircalabelu, Yumou Qiu, Emanuela Raffinetti, David Ruegamer, Marta Nai Ruscone, Russell Shinohara, Jonathan Stewart, Gilles Stupfler, Garth Tarr, Sara Taskinen, Antoine Usseglio-Carleve, Thomas Verdebout, Asaf Weinstein, Wenbo Wu, Liyan Xie, Keisuke Yano and Qianqian Zhu.

Local Organizer:
HTW Berlin, University of Applied Sciences.

Dear Friends and Colleagues,

We warmly welcome you to Berlin for the 17th International Conference on Computational and Financial Econometrics (CFE 2023) and the 16th International Conference of the ERCIM Working Group on Computational and Methodological Statistics (CMStatistics 2023).

The primary objective of this conference is to convene researchers and practitioners to explore recent advancements in computational methods within economics, finance, and statistics. The extensive CFE-CMStatistics 2023 programme comprises approximately 440 sessions, featuring five plenary talks and over 1730 presentations. More than 1900 participants make the conference a cornerstone in our series, marked by substantial size and qualitative growth. Undeniably, it stands as one of the most prominent international scientific events in our field.

The co-chairs have diligently curated a balanced and stimulating programme, designed to cater to the diverse interests of our participants. We trust that the hybrid nature of this conference will foster an optimal environment for effective communication.

The success of this conference is a testament to the collective efforts of numerous individuals and organizations, including the Scientific Programme Committee, Session Organizers, supporting universities, and various agents. We express our heartfelt appreciation for their substantial contributions to the meticulous organization of this event, and we extend our gratitude for the unwavering support of our networks.

We also extend our sincere gratitute to HTW Berlin for providing outstanding facilities and creating a superb networking environment. The local hosts have played a valuable role in ensuring the seamless organization of this conference, and we are profoundly grateful for their invaluable support.

We are pleased to announce that the official journal of CFEnetwork and CMStatistics, EcoSta, has been honored with its inaugural impact factor of 1.9 in 2022, as announced in June 2023. Concurrently, Computational Statistics & Data Analysis (CSDA) continues to uphold its commendable and consistent performance, with an impact factor of 1.8 for the year 2022.

Econometrics and Statistics, EcoSta, is an Elsevier journal publishing research papers across all facets of econometrics and statistics, comprising two sections, namely, Part A: Econometrics and Part B: Statistics. We strongly encourage participants to submit their papers to special or regular peer-reviewed issues of EcoSta and its supplement, Annals of Computational and Financial Econometrics.

CMStatistics also publishes The Annals of Statistical Data Science (SDS) as a supplement to the Elsevier journal CSDA, an official journal of CMStatistics. Authors are warmly encouraged to submit their papers to The Annals of Statistical Data Science or regular peer-reviewed issues of CSDA.

We are excited to announce that CFE-CMStatistics 2024 will be hosted at King's College London, UK, from Saturday, December 14th, to Monday, December 16th, 2024, with tutorials scheduled prior to the conference. We extend a heartfelt invitation and enthusiastic encouragement for your active participation in these forthcoming events.

We wish you a highly productive and inspiring conference.

Warm regards,

Ana Colubi, Erricos J. Kontoghiorghes and Manfred Deistler

Coordinators of CMStatistics & CFEnetwork and EcoSta.

CMStatistics: ERCIM Working Group on COMPUTATIONAL AND METHODOLOGICAL STATISTICS

http://www.cmstatistics.org

The working group (WG) CMStatistics comprises a number of specialized teams in various research areas of computational and methodological statistics. The teams act autonomously within the framework of the WG in order to promote their own research agenda. Their activities are endorsed by the WG. They submit research proposals, organize sessions, tracks and tutorials during the annual WG meetings and edit journal special issues. The Econometrics and Statistics (EcoSta) and Computational Statistics & Data Analysis (CSDA) are the official journals of the CMStatistics.

Specialized teams

Currently, the ERCIM WG has over 1950 members and the following specialized teams

BIO: Biostatistics
BS: Bayesian Statistics
DMC: Dependence Models and Copulas
DOE: Design Of Experiments
FDA: Functional Data Analysis
HDS: High-Dimensional Statistics
IS: Imprecision in Statistics
LVSEM: Latent Variable and Structural Equation Models
MM: Mixture Models
NPS: Non-Parametric Statistics
RS: Robust Statistics
SA: Survival Analysis
SAE: Small Area Estimation
SDS: Statistical Data Science
SEA: Statistics of Extremes and Applications
SL: Statistical Learning
TSMC: Times Series: Methods and Computations

You are encouraged to become a member of the WG. For further information, please contact the Chairs of the specialized groups (see the WG's website) or email at info@cmstatistics.org.

CFEnetwork COMPUTATIONAL AND FINANCIAL ECONOMETRICS

http://www.CFEnetwork.org

The Computational and Financial Econometrics (CFEnetwork) comprises a number of specialized teams in various research areas of theoretical and applied econometrics, financial econometrics and computation, and empirical finance. The teams contribute to the network's activities by organizing sessions, tracks and tutorials during the annual CFEnetwork meetings, and by submitting research proposals. Furthermore, the teams edit special issues currently published under the Annals of CFE. The Econometrics and Statistics (EcoSta) is the official journal of the CFEnetwork. Currently, the CFEnetwork has over 1100 members.

You are encouraged to become a member of the CFEnetwork. For further information, please see the website or contact by email at info@cfenetwork.org.

of candidate SNPs. A Bayesian framework is proposed, adapting ideas from group Lasso regression, that seeks to detect groups of correlated SNPs associated with the trait more accurately. In this model, priors are informed by biological assumptions about the sparsity of associated groups to improve the precision of association detection; signals from causative SNPs and SNPs correlated with causative ones are accumulated to make the detection easier; and the total number of variables that need to be tested is vastly reduced. A population-based MCMC method is used for efficient posterior sampling. Results from a variety of contexts show that the proposed method improves on a variety of existing methods at association detection, especially when signals are weak.

E0307: **Bayesian graph-structured variable selection**
Presenter: **Mahlet Tadesse**, Georgetown University, United States
Co-authors: Marie Denis

A graph structure is commonly used to characterize the dependence between variables, which may be induced by time, space, biological networks or other factors. Incorporating this dependence structure into the variable selection process can increase the power to detect subtle effects without increasing the probability of false discoveries and can improve predictive performance. Methods presented are proposed to accomplish this in the context of spike-and-slab priors as well as global-local shrinkage priors. For the former, a binary Markov random field prior is specified that leverages evidence from correlated outcomes on the variable selection indicators to identify outcome-specific covariates. For the latter, a Gaussian Markov random field prior is combined with a horseshoe prior to performing selection on graph-structured variables. The methods using epigenomic are illustrated, genomic and transcriptomic data.

E0512: **Unsupervised learning approaches for multi-OMICS data**
Presenter: **Marina Evangelou**, Imperial College London, United Kingdom

It is increasingly common these days for biomedical studies to generate multiple OMICS datasets for the same individuals. The conventional approaches for understanding the relationships between the OMICS datasets and the complex traits of interest (e.g. diseases) would be through the analysis of each dataset separately from the rest. Similarly, if researchers are interested in understanding the relationships between the OMICS datasets, they will perform pairwise tests with the features of the two OMICS datasets. It is illustrated that integrating multiple OMICS datasets improves understanding of their in-between relationships and improves their predictive performance. Two alternative data integration approaches will be presented: an extension of sparse canonical correlation analysis (sCCA) for the integration of multiple (more than 2) OMICS datasets. Although sCCA is an unsupervised learning approach, it is illustrated that by including the response variable as one of the datasets the predictive performance is increased. The second approach presented, named multi-SNE, is an extension of the well-known t-SNE approach for dimensionality reduction and visualisation of multi-view data. By incorporating the obtained low-dimensional embeddings of multi-SNE into the K-means clustering algorithm, it is shown that sample clusters are accurately identified.

E0522: **Harnessing public genomics big data to gain functional insights on complex diseases**
Presenter: **Zhaohui Qin**, Emory University, United States

Understanding the biological mechanisms underlying complex human diseases remains a fundamental challenge in biomedical research. In recent years, rapid development and dissemination of high throughput technologies have resulted in massive amounts of genomics data produced and publicly available, which gives researchers new opportunities to spark new hypotheses and uncover fresh insights. Many researchers including the group have developed powerful computational tools to enable researchers to better utilize the massive genomics data to gain insights on complex diseases they are interested in. The statistical and machine learning methods are reviewed that played key roles in these computational methods. The hope is to present a big picture of how genomics big data can potentially make its way into the clinics and help improve health care.

EO190 Room 404 ADVANCES IN KERNEL METHODS AND GAUSSIAN PROCESSES Chair: Meng Li

E0298: **Kernel cumulants**
Presenter: **Zoltan Szabo**, LSE, United Kingdom
Co-authors: Patric Bonnier, Harald Oberhauser

Maximum mean discrepancy (MMD, also called energy distance) and Hilbert-Schmidt independence criterion (HSIC, a.k.a. distance covariance) rely on the mean embedding of probability distributions and are among the most successful approaches in machine learning and statistics to quantify the difference and the independence of random variables, respectively. Higher-order variants of MMD and HSIC are presented by extending the notion of cumulants to reproducing kernel Hilbert spaces. The resulting kernelized cumulants have various benefits: (i) they are able to characterize the equality of distributions and independence under very mild conditions, (ii) they are easy to estimate with minimal computational overhead compared to their degree one (MMD and HSIC) counterparts, (iii) they achieve improved power when applied in two-sample and independence testing for environmental and traffic data analysis.

E1056: **Spectral regularized kernel two-sample test**
Presenter: **Bharath Sriperumbudur**, Pennsylvania State University, United States
Co-authors: Omar Hagrass, Bing Li

Over the last decade, an approach that has gained a lot of popularity to tackle non-parametric testing problems on general (i.e., non-Euclidean) domains is based on the notion of reproducing kernel Hilbert space (RKHS) embedding of probability distributions. The main goal is to understand the optimality of two-sample tests constructed based on this approach. First, it is shown that the popular MMD (maximum mean discrepancy) two-sample test is not optimal in terms of the separation boundary measured in Hellinger distance. Second, a modification to the MMD test is proposed based on spectral regularization by taking into account the covariance information (which is not captured by the MMD test) and the proposed test is proven to be minimax optimal with a smaller separation boundary than that achieved by the MMD test. Third, an adaptive version of the above test is proposed which involves a data-driven strategy to choose the regularization parameter and show the adaptive test to be almost minimax optimal up to a logarithmic factor. Moreover, the results hold for the permutation variant of the test where the test threshold is chosen elegantly through the permutation of the samples. Through numerical experiments on synthetic and real-world data, the superior performance of the proposed test in comparison to the MMD test is demonstrated.

E1669: **Optimal plug-in Gaussian processes for inferring functional derivatives and equivalence with kernel methods**
Presenter: **Meng Li**, Rice University, United States

Functional derivatives are key nonparametric functionals in wide-ranging applications that require the analysis of the rate of change in unknown functions. In the Bayesian paradigm, Gaussian processes (GPs) are routinely used as flexible priors for unknown functions but lack a comprehensive theoretical and methodological basis for derivative estimation. A plug-in strategy is presented by differentiating the posterior distribution with GP priors for derivatives of any order. Contrary to existing perceptions of sub-optimality, it is demonstrated that plug-in GPs offer adaptive and optimal posterior contraction rates. An empirical Bayes approach for data-driven hyperparameter tuning is also introduced. The approach satisfies optimal rate conditions while maintaining computational efficiency. To the knowledge, this constitutes the first positive result for plug-in GPs in the context of inferring derivative functionals and leads to a practically simple nonparametric Bayesian method with optimal and adaptive hyperparameter tuning for simultaneously estimating the regression function and its derivatives. Time permitting, an equivalence connection between GPs and kernel ridge regression will be introduced for function derivatives, which serve as a mathematical foundation.

