SIS2016

Università degli Studi di Salerno

June 8th – June 10th, 2016

# PROCEEDINGS

of the 48th scientific meeting of the
Italian Statistical Society

Editors: Monica Pratesi and Cira Pena

ISBN: 9788861970618

# PLENARY SESSIONS

**(A) E. Baldacci** — Financial Crises and their Impacts: Data Gaps and Innovation in Statistical Production.

**(B) D. Dunson** — Probabilistic inference from big and complex data.

**(C) S. Strozza** — Foreign immigration in Italy: a forty-year-old history.

# SPECIALIZED SESSION (SPE)

### (SPE-01) Inference, sampling and survey design

**P. Conti** — Resampling from finite populations under complex designs: the pseudo-population approach. (Co-author(s): F. Andreis, D. Marella, F. Mecatti)

**P. Righi** — A joint use of model based and design based frameworks for defining optimal sampling designs. (Co-author(s): P. D. Falorsi)

**A. Ruiz-Gazen** — A unified approach for robustness in survey sampling. (Co-author(s): J. Beaumont, D. Haziza)

### (SPE-02) Multivariate models for risk assessment

**M. Billio** — A Bayesian nonparametric approach to macroeconomic risk. (Co-author(s): R. Casarin, M. Costola, M Guindani)

**P. Cerchiello** — Bank risk contagion:an analysis through big data. (Co-author(s): P. Giudici, G. Nicola)

**L. De Angelis** — A Markov-switching regression model with non-Gaussian innovations for systemic risk measurement. (Co-author(s): C. Viroli)

### (SPE-03) Bayesian nonparametrics

**D. Durante** — Bayesian Nonparametric Modeling of Dynamic International Relations. (Co-author(s): D. Dunson)

**A. Guglielmi** — Bayesian autoregressive semiparametric models for gap times of recurrent events. (Co-author(s): G. Paulon, M. De Iorio)

**A. Rodriguez** — Restricted Nonparametric Mixtures models for Disease Clustering. (Co-author(s): T. Xifara)

### (SPE-04) Statistical methods for the analysis of gene-environment interaction in the study of complex pathologies

**C. Angelini** — An introduction to next generation sequencing for studying omic-environment interactions.

**L. Calciano** — Statistical approaches for the evaluation of genetic associations in complex diseases: the heterogeneity of asthma phenotypes. (Co-author(s): L. Portas, S. Accordini)

**Y. Pankaj** — Improved case-only approach to study genome-wide gene-environment interaction.

### (SPE-05) Nonlinear time series

**M. Niglio** — Probabilistic properties of Self Exciting Threshold Autoregressive processes. (Co-author(s): F. Giordano, C. D. Vitale)

**T. Proietti** — Optimal prediction of stochastic trends. (Co-author(s): A. Giovannelli)

**H. Tong** — On model selection from a finite family of possibly misspecified models. (Co-author(s): H. Hsu, C. Ing)

### (SPE-06) Spatial analyses in demography

**F. Heins** — Measuring residential segregation with spatial indices: an appraisal and applications for the metropolitan area of Rome. (Co-author(s): F. Benassi, F. Lipizzi, E. Paluzzi)

**A. Mazza** — Immigrants' settlement patterns in the city of Naples. (Co-author(s): G. Gabrielli, S. Strozza)

**L. Natale** — Native Immigration and Pull Factor Evolution in Italy: a Spatial Approach. (Co-author(s): A. Santacroce, F. G. Truglia)

### (SPE-07) Recent developments in Volatility modeling

**R. Casarin** — Dynamic Model Averaging for Quantile Regression. (Co-author(s): M. Bernardi, B. Maillet, L. Petrella)

**A. Rahbek** — Testing volatility: consistency of bootstrap testing for a parameter on the boundary of the parameter space.

**E. Ruiz** — Asymmetric Stochastic Volatility Models: Properties and Estimation. (Co-author(s): V. Czellar, X. Mao, H. Veiga)

### (SPE-08) Advances in ordinal contingency table analysis

**L. D'Ambra** — Dimensionality reduction methods for contingency tables with ordinal variables. (Co-author(s): P. Amenta, A. D'Ambra)

**R. Lombardo** — Modelling Trends in Ordered Three-Way Non-Symmetrical Correspondence Analysis. (Co-author(s): P. Kroonenberg, E. Beh)

**M. Riani** — Using Collapsing and Multiple Comparisons to Detect Association in Two Way Contingency Tables. (Co-author(s): S. Arsenis)

### (SPE-09) Statistical models for directional and circular data

**C. Ley** — The WeiSSVM: a tractable, parsimonious and flexible model for cylindrical data.

**G. Mastrantonio** — The multivariate projected-skew normal distribution: Bayesian estimation and a hidden Markov model application.

**A. Panzera** — Circular density estimation via matching local trigonometric moments. (Co-author(s): M. Di Marzio, S. Fensore, C. C. Taylor)

### (SPE-10) The interplay between frequentist and bayesian inference

**C. Grazian** — Classical inference for intractable likelihoods.

**J. Hannig** — Fusion learning for Interlaboratory Comparison. (Co-author(s): Q. Feng, H. Iyer, C. Wang, X. Liu)

**F. Pauli** — p-value in science: a review of issues and proposed solutions.

### (SPE-11) Société Française de Statistique

**B.H. Avner** — Stochastic Block Model for Multiplex network: an application to a multilevel network of researchers..

**Y. Bennani** — Nonnegative Matrix Factorization for Transfer Learning. (Co-author(s): I. Redko)

**T. Laloe** — Detection of dependence patterns with delay.

**J. Poggi** — Disaggregated Electricity Forecasting using Wavelet-Based Clustering of Individual Consumers. (Co-author(s): J. Cugliari, Y. Goude)

### (SPE-12) National accounts

**A. Coli** — The European Welfare State in times of crisis according to macroeconomic official statistics. (Co-author(s): E. Micheletti, B. Pacini)

**C. Martelli** — National Account and Open Data: a new semantic approach.

**G. Oneto** — New information contents of the National Accounts for the monitoring of the economic situation.

### (SPE-13) Statistical tools for monitoring the educational system and assessing students' performances

**L. Grilli** — Evaluation of university students' performance through a multidimensional finite mixture IRT model. (Co-author(s): S. Bacci, F. Bartolucci, C. Rampichini)

**G. Leckie** — Monitoring school performance using value-added and value-table models: Lessons from the UK.

**P. Sarnacchiaro** — A statistical model to assess teacher performance. (Co-author(s): I. Camminatiello, R. Palma)

### (SPE-14) Robust inference by bounded estimating functions

**A.C. Monti** — M Estimation based Inference for Ordinal Response Model.

**E. Ruli** — Approximate Robust Bayesian Inference with an Application to Linear Mixed Models. (Co-author(s): N. Sartori, L. Ventura)

**J. Valeinis** — Some robust methods using empirical likelihood for two samples. (Co-author(s): M. Velina, E. Cers, G. Luta)

# SOLICITED SESSION (SOL)

### (SOL-01) Subjective wellbeing and demographic events over the life course

**G. Fuochi** — Cultural and institutional drivers of basic psychological needs satisfaction. (Co-author(s): P. Conzo, A. Aassve, L. Mencarini)

**L. Mencarini** — Five reasons to be happy about childbearing. (Co-author(s): A. Aassve, F. Luppi)

**B. Nowok** — Migration motivations and migrants' satisfaction in the life course: A sequence analysis of geographical mobility trajectories in the United Kingdom.

**A. Pirralha** — Does becoming a parent change the meaning of happiness and life satisfaction? Evidence from the European Social Survey. (Co-author(s): H. Dobewall)

### (SOL-02) Statistics for equitable and sustainable development

**E. di Bella** — Wellbeing and sustainable development: a multi-indicator approach to evaluate urban waste management systems. (Co-author(s): C. Barbara, M. Corsi)

**C. Giusti** — Small Area Estimation for Local Welfare Indicators in Italy. (Co-author(s): S. Marchetti, L. Faustini, L. Porciani)

**T. Laureti** — Does socio-economic variables influence the Italians' adherence towards a sustainable diet?. (Co-author(s): L. Secondi)

**F. Riccardini** — Sustainability of wellbeing: an analysis of resilience and vulnerability through subjective indicators. (Co-author(s): M. Bachelet, F. Maggino)

### (SOL-03) New approaches to treat undercoverage and nonresponse

**F. Andreis** — Methodological perspectives for surveying rare and clustered population: towards a sequentially adaptive approach.

**E. Furfaro** — Dealing with under-coverage bias via Dual/Multiple Frame designs: a simulation study for telephone surveys.

**D. Haziza** Weight adjustment procedures for the treatment of unit nonresponse in surveys.

**E. Kabzinska** Empirical likelihood multiplicity adjusted estimator for multiple frame surveys. (Co-author(s): Y. G. Berger)

### (SOL-04) Statistical models and methods for network data

**M. Cugmas** Measuring stability of co-authorship structures in time. (Co-author(s): A. Ferligoj)

**J. Koskinen** A dynamic discrete-choice model for movement flows. (Co-author(s): T. Mueller, T. Grund)

**G. Ragozini** Prototyping and Comparing Networks through Archetypal Analysis. (Co-author(s): D. De Stefano, M.R. D'Esposito)

**S. Zaccarin** Modeling network dynamics: evidence from policy-driven innovation networks. (Co-author(s): A. Caloffi, D. De Stefano, F. Rossi, M. Russo)

### (SOL-05) Recent developments in computational statistics

**R. Argiento** A conditional algorithm for Bayesian finite mixture models via normalized point process.

**S. Favaro** Thompson sampling for species discovery. (Co-author(s): M. Battiston, Y. Teh)

**A. Mira** An application of Reinforced Urn Process to advice network data. (Co-author(s): S. Peluso, P. Muliere, F. Pallotti, A. Loni)

**N. Sartori** Bootstrap prepivoting in the presence of many nuisance parameters. (Co-author(s): R. Bellio, I. Kosmidis, A. Salvan)

### (SOL-06) Statisticians meet naturalists: issues on ecological and environmental statistics

**F. Ferretti** Estimating the abundance of wildlife ungulate populations in Mediterranean areas: methods, problems and findings. (Co-author(s): A. Sforzi)

**M. Ferretti** The monitoring of forests in Europe: methods, problems and proposals.

**D. Rocchini** The power of generalized entropy for biodiversity assessment by remote sensing: an open source approach. (Co-author(s): L. Delucchi, G. Bacaro)

### (SOL-07) From survey data to new data sources and big data in official statistics

**G. Barcaroli** Machine learning and statistical inference: the case of Istat survey on ICT. (Co-author(s): G. Bianchi, R. Bruni, A. Nurra, S. Salamone, M. Scarnò)

**S. Falorsi** Forecasting Italian Youth Unemployment Rate Using Online Search Data. (Co-author(s): S. Loriga, A. Naccarato, A. Pierini)

**B. Liseo** Bayesian nonparametric methods for record linkage. (Co-author(s): A. Tancredi)

**T. Tuoto** Exploring solutions for linking Big Data in Official Statistics. (Co-author(s): L. Di Consiglio, D. Fusco)

### (SOL-08) Symbolic data analysis methods and applications

**E. Diday** Explanatory and discriminatory power of variables in Symbolic Data Analysis.

**M.B. Ferraro** Fuzzy and possibilistic approach to clustering of imprecise data. (Co-author(s): P. Giordani)

**L. Grassini** Symbolic data analysis approach for monitoring the stability of monuments.. (Co-author(s): B. Bertaccini, A. Giusti)

**M. Ichino** Similarity and Dissimilarity Measures for Mixed Feature-type Symbolic Data. (Co-author(s): K. Umbleja)

### (SOL-09) Compositional analysis

**L. Crosato** Forecasting CPI weights through compositional VARIMA: an application to Italian data.. (Co-author(s): F. Lovisolo, B. Zavanella)

**J. A. Martín-Fernández** Understanding association rules from a compositional data approach. (Co-author(s): M. Vives-Mestres, R. Kenett)

**A. Menafoglio** Object Oriented Geostatistical Simulation of Functional Compositions via Dimensionality Reduction in Bayes spaces. (Co-author(s): A. Guadagnini, P. Secchi)

**V. Simonacci** Fitting CANDECOMP-PARAFAC model for compositional data: a combined SWATLD-ALS algorithm. (Co-author(s): M. Di Palma, V. Todorov)

### (SOL-10) Sustainable development: theory, measures and applications

**F. Riccardini** Measuring sustainable development goals from now to 2030.

**F. Riccardini** How the nexus of food/water/energy can be seen with the perspective on well-being of people and the Italian BES framework. (Co-author(s): D. De Rosa)

**T. Rondinella** An innovative methodology for the analysis of sustainability, inclusion and smartness of growth through Europe2020 indicators.. (Co-author(s): E. Grimaccia)

**P. Ungaro** The Italian population behaviours toward environmental sustainability: a study from Istat surveys. (Co-author(s): I. Mingo, V. Talucci)

### (SOL-11) Detecting heterogeneity in ordinal data surveys

**E. Di Nardo** CUB models: a preliminary Fuzzy approach to heterogeneity. (Co-author(s): R. Simone)

**S. Giordano** Modelling uncertainty in bivariate models for ordinal responses. (Co-author(s): R. Colombi, A. Gottard, M. Iannario)

**M. Manisera** — Treatment of "don't know" responses in rating data: effects on the heterogeneity of the CUB distribution. (Co-author(s): P. Zuccolotto)

**F. Pennoni** — Modelling a multivariate hidden Markov process on survey data.

### (SOL-12) Active ageing: age management and lifelong learning strategies

**P. E. Cardone** — Age management in Italian companies. Findings of two Isfol surveys. (Co-author(s): M. Aversa, L. D'Agostino)

**A. Lorenti** — Working after Retirement in Europe.

**C. Polli** — Older low-skilled workers and economic crisis in Italy. (Co-author(s): R. Angotti)

**G. Rivellini** — Population ageing and human resources management. A chance for Applied Demography. (Co-author(s): F. Marcaletti, F. Racioppi)

### (SOL-13) Statistical models for evaluating policy impact

**M. Bia** — Evaluation of Training Programs by exploiting secondary outcomes in Principal Stratification frameworks: the case of Luxembourg. (Co-author(s): F. Li, A. Mercatanti)

**G. Cerulli** — Testing Stability of Regression Discontinuity Models. (Co-author(s): Y. Dongz, A. Lewbel, A. Poulsen)

**R. P. Mamede** — Counterfactual Impact Evaluation of Vocational Education in Portugal. (Co-author(s): D. Cruz, T. Fernandes)

**G. Pellegrini** — Italian public guarantees to SME: the impact on regional growth. (Co-author(s): M. De Castris)

### (SOL-14) Usage of geocoded micro data in the economic analysis

**M. Dickson** — Spatial sampling methods with locational errors. (Co-author(s): D. Filipponi)

**D. Giuliani** — Spatial Micro-Econometrics Models with Locational Errors. (Co-author(s): S. Cozzi, G. Espa)

**F. Santi** — Three-Year Survival Probability of Italian Start-up Businesses in Healthcare Industry: an Empirical Investigation through Logistic Multilevel Modelling. (Co-author(s): M. M. Dickson, D. Giuliani, D. Piacentino)

### (SOL-15) Statistical models in functional data analysis

**G. Adelfio** — Space-time FPCA Algorithm for clustering ofmultidimensional curves. (Co-author(s): F. Di Salvo, M. Chiodi)

**C. Miller** — Functional data analysis approaches for satellite remote sensing applications. (Co-author(s): R. O'Donnell, M. Gong, M. Scott)

**E. Romano** — Order statistics for spatially dependent functional data. (Co-author(s): A. Balzanella, R. Verde)

**L. M. Sangalli** — A penalized regression model for functional data with spatial dependence. (Co-author(s): M. S. Bernardi, G. Mazza, J. O. Ramsay)

### (SOL-16) Forecasting economic and financial time series

**G. Goracci** — Asymptotics and power of entropy based tests of dependence for categorical data. (Co-author(s): S. Giannerini)

**M. M. Pelagatti** — Forecasting electricity load and price: a comparison of different approaches. (Co-author(s): F. Lisi)

**G. Storti** — Flexible Realized GARCH Models. (Co-author(s): R. Gerlach)

### (SOL-17) Immigrations and integration in Italy

**O. Casacchia** — Minorities internal migration in Italy: an analysis based on gravity models. (Co-author(s): C. Reynaud, S. Strozza, E. Tucci)

**C. Conti** — Growing generations and new models of integration.

**N. Tedesco** — Measurement of segregation in the labour market. An alternative approach. (Co-author(s): L. Salaris)

**L. Terzera** — Family behaviours among first generation migrants. (Co-author(s): E. Barbiano di Belgiojoso)

### (SOL-18) Open data, linked data and big data in public administration and official statistics

**G. Di Bella** — Linked Administrative Data in Official Statistics: a Positive Feedback for the Quality?. (Co-author(s): G. Garofalo)

**C. Martelli** — Generating high quality administrative data: new technologies in a national statistical reuse perspective. (Co-author(s): M. Calzaroni, A. Samaritani)

**V. Santarcangelo** — An innovative approach about the analysis of quality and efficiency in Italian law. (Co-author(s): A. Buondonno, A. Romano, M. Giacalone, C. Cusatelli)

**B. Squittieri** — Prato municipality experience towards a high integration between administrative and statistical data.

### (SOL-19) Evaluation of prognostic biomarkers

**F. Ambrogi** — Combining Clinical and Omics data: hope or illusion?. (Co-author(s): P. Boracchi)

**L. Antolini** — Graphical representations and summary indicators to assess the performance of risk predictors. (Co-author(s): D. Bernasconi)

**P. Chiodini** — Multivariable prognostic model: external validation and model recalibration with application to non-metastatic renal cell carcinoma. (Co-author(s): L. Cindolo)

### (SOL-20) Models for studying the mobility of students

**S. Balia** — Modelling inter-regional patient mobility: evidence from the Italian NHS. (Co-author(s): R. Brau, E. Marrocu)

**A. D'Agostino** — University mobility at enrollment: geographical disparities in Italy. (Co-author(s): G. Ghellini, S. Longobardi)

**M. Enea** — From South to North? Mobility of Southern Italian students at the transition from the first to the second level university degree.

**F. Giambona** — Measuring territory student-attractiveness in Italy. Longitudinal evidence.

# CONTRIBUTED SESSION (CON)

### (CON-01) Bayesian statistics (1)

**F. Giummolè** — Reference priors based on composite likelihoods. (Co-author(s): V. Mameli, L. Ventura)

**B. Nipoti** — On Bayesian nonparametric inference for discovery probabilites. (Co-author(s): J. Arbel, S. Favaro, Y. Teh)

**R. Pappadà** — Relabelling in Bayesian mixture models by pivotal units. (Co-author(s): L . Egidi, F. Pauli, N. Torelli)

**C. Scricciolo** — On Deconvolution of Dirichlet-Laplace Mixtures.

### (CON-02) Statistical modeling

**P. Faroughi** — A New Bivariate Regression Model for Count Data with Excess Zeros. (Co-author(s): N. Ismail)

**B. Francis** — Dynamic latent class profiles in cross-sectional surveys: some preliminary results. (Co-author(s): V. Hoti)

**P. M. Kroonenberg** — The use of deviance plots for non-nested model selection in loglinear models, structural equations, three-mode analysis.

**A. Lucadamo** — Variable selection through Multinomial LASSO for PCMR. (Co-author(s): L. Greco)

**O. Paccagnella** — Integrating CUB Models and Vignette Approaches. (Co-author(s): S. Pavan, M. Iannario)

### (CON-03) Demographics and social statistics (1)

**D. Bellani** — Gender egalitarianism, education and life-long singlehood: A multilevel analysis. (Co-author(s): G. Esping-Andersen, L. Nedoluzhko)

**L. Colangelo** — Fear of Crime and Victimization among Sexual Harassed Women: Evidence from Italy. (Co-author(s): P. Mancini)

**S. De Cantis** — A survival approach for the analysis of cruise passengers' behavior at the destination. (Co-author(s): M. Ferrante, A Parroco, N Shoval)

**A. Di Pino** — Retirement of the Male Partner and the Housework Division in the Italian Couples: Estimation of the Causal Effects. (Co-author(s): M. Campolo)

**F. Lariccia** — Many women start, but few continue: determinants of breastfeeding in Italy. (Co-author(s): A. Pinnelli)

### (CON-04) Environmental statistics

**F. Bono** — Measuring sustainable economic development through a multidimensional Gini index. (Co-author(s): M. Giacomarra, R. Giaimo)

**C. Calculli** — Modeling multi-site individual corals growth. (Co-author(s): B. Cafarelli, D. Cocchi, E. Pignotti)

**F. Di Salvo** — GAMs and functional kriging for air quality data. (Co-author(s): A. Plaia, M. Ruggieri)

**F. Durante** — The Kendall distribution and multivariate risks.

### (CON-05) Health statistics

**E. di Bella** — Dental care systems across Europe: the case of Switzerland. (Co-author(s): L. Leporatti, I. Krejci, S. Ardu)

**F. Gasperoni** — Multi-state models for hospitalizations of heart failure patients in Trieste. (Co-author(s): F. Ieva, G. Barbati)

**F. Grossetti** — Multi-state Approach to Administrative Data on Patients affected by Chronic Heart Failure. (Co-author(s): F. Ieva, S. Scalvini, A. M. Paganoni)

**G. Montanari** — Evaluation of health care services through a latent Markov model with covariates. (Co-author(s): S. Pandolfi)

### (CON-06) Labor market statistics

**A. Bianchi** — Multifactor Partitioning: an analysis of employment and firm size. (Co-author(s): S. Biffignandi)

**G. Busetta** — Ugly Betty looks for a job. Will she ever find it in Italy?. (Co-author(s): F. Fiorillo)

**G. Busetta** — No country for foreigners: an analysis of hiring process in Italian labor market. (Co-author(s): M. Campolo, D. Panarello)

**F. Crippa** — Know your audience. Towards a partnership between employers and university. (Co-author(s): M. Zenga)

**I. Vannini** — Online Job Vacancies: a big data analysis. (Co-author(s): D . Rotolone, C. Di Stefano, A. P. Paliotta, D. F. Iezzi)

### (CON-07) Robust statistics

**F. Greselin** — Robust estimation of mixtures of skew-normal distributions. (Co-author(s): L. García-Escudero, A. Mayo-Iscar, G. McLachlan)

**M. Musio** — Renyi's Scoring Rules. (Co-author(s): A. F. Dawid)

**A. Paganoni** — Robust classification of multivariate functional data. (Co-author(s): F. Ieva)

**G. C. Porzio** — A robust estimator for the mean direction of the von Mises-Fisher distribution. (Co-author(s): T. Kirschstein, S. Liebscher, G. Pandolfo, G. Ragozini)

**F. Palumbo** — Robust Partial Possibilistic Regression Path Modeling. (Co-author(s): R. Romano)

### (CON-08) Sampling methods

**A. Ghiglietti** — Adaptive Randomly Reinforced Urn design and its asymptotic properties.

**D. Marella** — PC algorithm from complex sample data. (Co-author(s): P. Vicard)

**S. Missiroli** — Optimal Adaptive Group Sequential Procedure for Finite Populations in the Presence of a Cost Function. (Co-author(s): E. Carfagna)

**E. Pelle** — The Rao regression-type estimator in ranked set sampling. (Co-author(s): P. Perri)

**M. Ruggiero** — Modelling stationary varying-size populations via Polya sampling. (Co-author(s): P. De Blasi, S. Walker)

### (CON-09) Economic data analysis

**M. Brunetti** — Getting older and riskier: the effect of Medicare on household portfolio choices. (Co-author(s): M. Angrisani, V. Atella)

**E. Ciavolino** — Modelling the Public Opinion on the European Economy with the HO-MIMIC Model. (Co-author(s): M. Carpita)

**G. D'Epifanio** — Indexing the Worthiness of Social Agents. To norm index on conventional specifications.

**G. Guagnano** — An econometric model for undeclared work. (Co-author(s): M. Arezzo)

**M. Mussini** — A spatial shift-share decomposition of energy consumption variation. (Co-author(s): L. Grossi)

### (CON-10) Quantile methods

**M. Bernardi** — Bayesian inference for $L_p$–quantile regression models. (Co-author(s): V. Bignozzi, L. Petrella)

**V. Bignozzi** — On the $L_p$-quantiles and the Student t distribution. (Co-author(s): M. Bernardi, L. Petrella)

**M. Marino** — M-quantile regression for multivariate longitudinal data. (Co-author(s): M. Alfò, M. Ranalli, N. Salvati)

**D. Vistocco** Comparing Prediction Intervals in Quantile and OLS Regression. (Co-author(s): C. Davino)

### (CON-11) Statistical algorithms

**N. Loperfido** An Algorithm for Finding Projections with Extreme Kurtosis. (Co-author(s): C. Franceschini)

**L. Scrucca** Poisson change-point models estimated by Genetic Algorithms.

**A. Stamm** Maximum Likelihood Estimators of Brain White Matter Microstructure. (Co-author(s): O. Commowick, S. Vantini, S. K. Warfield)

### (CON-12) Statistics for medicine

**G. Barbati** Competing risks between mortality and heart failure hospital re-admissions: a community-based investigation from the Trieste area. (Co-author(s): F. Ieva, A. Scagnetto, G. Sinagra, A. Di Lenarda)

**C. Brombin** Evaluating association between emotion recognition and Heart Rate Variability indices. (Co-author(s): F. Cugnata, R. M. Martoni, M. Ferrario, C. Di Serio)

**M. Ferrante** Socio-economic deprivation, territorial inequalities and mortality for cardiovascular diseases in Sicily. (Co-author(s): A. Milito, A. Parroco)

**M. Giacalone** The use of Permutation Tests on Large-Sized Datasets. (Co-author(s): A. Alibrandi, A. Zirilli)

### (CON-13) Statistics for the education system

**G. Boscaino** Further considerations on a new indicator for higher education student performance. (Co-author(s): G. Adelfio, V. Capursi)

**C. Masci** Analysis of pupils' INVALSI achievements by means of bivariate multilevel models. (Co-author(s): A. Paganoni, F. Ieva, T. Agasisti)

**A. Valentini** Promoting statistical literacy to university students: a new approach adopted by Istat. (Co-author(s): G. De Candia, M . Carbonara)

### (CON-14) Testing procedures

**E. Cascini** A Reliability Problem: Censored Tests.

**G. De Santis** Testing the Gamma-Gompertz-Makeham model. (Co-author(s): G. Salinari)

**M. M. Pelagatti** A nonparametric test of independence.

**A. Pini** Functional Data Analysis of Tongue Profiles. (Co-author(s): L. Spreafico, S. Vantini, A. Vietti)

**A. Vagheggini** On the asymptotic power of the statistical test under Response-Adaptive randomization. (Co-author(s): A. Baldi Antognini, M. Zagoraiou)

### (CON-15) Time series analysis

**C. Cappelli** Robust Atheoretical Regression Tree to detect structural breaks in financial time series. (Co-author(s): P. D'Urso, F. Di Iorio)

**P. Chirico** Prediction intervals for heteroscedastic series by Holt-Winters methods.

**M. Costa** Inequality decomposition for financial variables evaluation.

**G. De Luca** Three-stage estimation for a copula-based VAR model. (Co-author(s): G. Rivieccio)

### (CON-16) Forecasting methods

**M. Andreano** Forecasting with Mixed Data Sampling Models (MIDAS) and Google trends data: the case of car sales in Italy. (Co-author(s): R. Benedetti, P. Postiglione)

**V. Candila** Probability forecasts in the market of tennis betting: the CaSco normalization. (Co-author(s): A. Scognamillo)

**S. Vantini** Daily Prediction of Demand and Supply Curves. (Co-author(s): A. Canale)

### (CON-17) Bayesian statistics (2)

**G. Marchese** Bayesian hierarchical models for analyzing and forecasting football results. (Co-author(s): P. Brutti, S. Gubbiotti)

**L. Paci** Bayesian modeling of spatio-temporal point patterns in residential property sales. (Co-author(s): A. E. Gelfand, M. Beamonte, P. Gargallo, M. Salvador)

**V. Vitale** Non-parametric Bayesian Networks for Managing an Energy Market. (Co-author(s): V. Guizzi, F. Musella, P. Vicard)

### (CON-18) Business statistics

**E. Bartoloni** How do firms perceive their competitiveness? Measurement and determinants.

**C. Bocci** An evaluation of export promotion programmes with repeated multiple treatments. (Co-author(s): M. Mariani)

**A. Righi** The inter-enterprise relations in Italy. (Co-author(s): A. Nuccitelli, G. Barbieri)

### (CON-19) Clustering and classification

**C. Drago** Dendrograms Stability Analysis of Sub-periods Time Series Clustering. (Co-author(s): R. Ricciuti)

**G. Menardi** Stability-based model selection in nonparametric clustering.

**T. Padellini** Topological signatures for classification. (Co-author(s): P. Brutti)

### (CON-20) Demographics and social statistics (2)

**M. Antonicelli** — Ecolabels: informin or confusing customers? Evidences form the agrifood sector. (Co-author(s): D. Calace, D. Morrone, A. Russo, V. Vastola)

**B. Arpino** — What makes you feeling old? An analysis of the factors influencing perceptions of ageing. (Co-author(s): V. Bordone, A. Rosina)

**G. De Santis** — A (partial) solution to the intractability of APC models. (Co-author(s): M. Mucciardi)

**G. Gabrielli** — Partner reunification of first generation immigrants in Lombardy. (Co-author(s): A. Paterno, L. Terzera)

### (CON-21) Statistical inference

**E. Kenne Pagui** — Median bias reduction of maximum likelihood estimates in binary regression models. (Co-author(s): A. Salvan, N. Sartori)

**N. Lunardon** — On penalized likelihood and bias reduction. (Co-author(s): G. Adimari)

**A. Maruotti** — Population size estimation and heterogeneity in capture-recapture count data. (Co-author(s): O. Anan, D. Bohning)

### (CON-22) Survey methods

**A. Pinto** — Italian consumers' food risks perception: an approach based on the correspondence analysis. (Co-author(s): E. Ruli, S. Crovato, L. Ventura, L. Ravarotto)

**R. Salvatore** — Spatial-temporal multivariate small area estimation. (Co-author(s): F. Cappuccio)

**D. Toninelli** — Is the Smartphone Participation Affecting the Web Survey Experience?. (Co-author(s): M. Revilla)

# POSTER SESSION (POS)

**M. Bernardi** — Non-conjugate Variational Bayes Approximation. (Co-author(s): E. Ruli)

**M. Bernardi** — The Multivariate Fuzzy Skew Student–t distribution.

**M. Bini** — Quality of Educational Services, Institutional Image, Students' Satisfaction and Loyalty in Higher Education. (Co-author(s): L. Masserini, M. Pratesi)

**L. Bisaglia** — Estimation of INAR(p) models using bootstrap. (Co-author(s): M. Gerolimetto)

**D. Bossoli** — Effect of internet-based cognitive therapy on children anxiety disorders: results from a marginal logistic quantile regression.

**C. Calì** Some mathematical properties of the ROC curve. (Co-author(s): M. Longobardi)

**M. Cannas** Machine learning for the estimation of the propensity score: a simulation study. (Co-author(s): B. Arpino)

**M. Cannas** An R package for propensity score matching with clustered data. (Co-author(s): B. Arpino, C. Conversano)

**A. Coli** Mapping local social protection data in Italy. (Co-author(s): B. Pacini, A. Valentini, S. Venturi)

**A. Cosma** Indirect inference for nonlinear panel data.

**I. L. Danesi** Cluster Analysis of Transactional Data in the Frequency Domain. (Co-author(s): F. M. Pons, C. Rea)

**L. Gabrielli** Using purchase market behavior to estimate collective well-being. (Co-author(s): G. Riccardi, L. Pappalardo)

**F. Giambona** The Bifactor Item Response Theory Model for the analysis of repeated measurements. An application to the measurement of Italian households' well-being. (Co-author(s): M. Porcu, I. Sulis)

**A. Lepore** A Bayesian short-term strategy for site specific wind potential assessment. (Co-author(s): P. Erto, B. Palumbo, M. Lepore)

**A. Magrini** Distributed-Lag Structural Equation Modelling: An Application to Impact Assessment of Research Activity on European Agriculture. (Co-author(s): F. Bartolini, A. Coli, B. Pacini)

**G. Mastrantonio** A multivariate circular-linear hidden Markov model for site-specific assessment of wind predictions by an atmospheric simulation system. (Co-author(s): A. Pollice, F. Fedele)

**F. Musella** Bayesian networks for supporting the digitization process in Italian schools. (Co-author(s): S. Capogna, M.C. De Angelis)

**B. Palumbo** Statistical approach in aerospace industry innovation. (Co-author(s): P. Erto, F. Tagliaferri, G. De Chiara, R. Marrone, C. Leone, S. Genna)

**B. Palumbo** Ship fuel consumption control and engineering approach to fault-detection. (Co-author(s): P. Erto, A. Lepore, L. Vitiello, C. Capezza, D. Bocchetti, A. D'Ambra, B. Antonelli)

**A. Petrucci** Small area model-based direct estimator for spatial data. (Co-author(s): C. Bocci, E. Rocco)

**F. Poggioni** Dynamic Quantile Lasso Regression. (Co-author(s): L. Petrella, M. Bernardi)

**A. Pramov** Estimating dependence within neuropsychological models for designing risk profiles of decision-makers..

**A. Pramov** Confidence intervals for a partially identified parameter with bounds estimated by the minimum and the maximum of two correlated and normally distributed statistics.

| | |
|---|---|
| **I. Primerano** | Semantic Knowledge Detection in Open-ended Questionnaire. (Co-author(s): G. Giordano) |
| **G. Ragozini** | A joint approach to the analysis of time-varying affiliation networks. (Co-author(s): D. D'Ambrosio, M. Serino) |
| **M. Restaino** | Non-parametric estimators for estimating bivariate survival function under randomly censored and truncated data. (Co-author(s): H. Dai, H. Wang) |
| **G. Riccardi** | Bayesian M-quantile regression in Small Area Estimation. |
| **E. Ruli** | Optimal B-robust posterior distributions for operational risk. (Co-author(s): I. Danesi, F. Piacenza, L. Ventura) |
| **F. Schirripa Spagnolo** | Estimating income of immigrant communities in Italy using small area estimation methods. (Co-author(s): N. Salvati, A. D'Agostino) |
| **M. Soscia** | The Switching Skew-GARCH Model. (Co-author(s): M. Bernardi, L. Petrella) |
| **S. Spina** | Inference on a non-homogeneous Gompertz process with jumps as model of tumor dynamics. (Co-author(s): V. Giorno, P. Román-Román, F. Torres-Ruiz) |
| **G. Storti** | Combining multiple frequencies in multivariate volatility forecasting. (Co-author(s): A. Amendola, V. Candila) |
| **D. Toninelli** | An Enhanced Measure of Well-being through Structural Equation Modeling: a Cross-Country Approach. (Co-author(s): M. Cameletti) |
| **A. Vanacore** | Statistics for knowledge improvement of an innovative manufacturing process and quality cost management. (Co-author(s): B. Palumbo, F. Del Re, P. Corrado, M. Lanza, G. La Sala, M. Mastrovita) |
| **A. Vanacore** | Statistics for Safety and Ergonomics in Design. (Co-author(s): A. Lanzotti, C. Percuoco, A. Capasso, F. Liccardo, B. Vitolo) |
| **L. Zanin** | Modelling transition probabilities in a flexible hierarchical logit framework: evidence from the Italian labour market. (Co-author(s): R. Calabrese) |

# Probability forecasts in the market of tennis betting: the CaSco normalization

## *Prevedere le probabilitá nel mercato delle scommesse sul tennis: La normalizzazione CaSco*

Vincenzo Candila and Antonio Scognamillo

**Abstract** The probability odds, obtained as the inverse of the betting odds, represent the most accurate proxy of the bookmaker forecasts associated to each player/team victory. However, these probability odds cannot be defined as actual forecasts because they incorporate the bookmaker margin and the longshot bias such that their summation is greater than one. In literature, different normalization methods have been proposed in order to close the gap between probability odds and the underlying forecasts. This paper proposes a new method of normalization, named CaSco (Candila-Scognamillo) normalization, for the fixed odds offered on sports with only two possible outcomes. The performances of the new procedure are evaluated using the betting odds provided by one of the most important on-line professional bookmaker on over 27.000 male tennis matches. The results show that the CaSco normalization has a better forecasting ability than that of the other approaches presented in literature.

**Abstract** Le probability odds, definite come l'inverso delle quote pubblicate dai bookmakers, sono considerate la miglior proxy della probabilità di vittoria prevista ed associata ad ogni giocatore/team. Tuttavia, tali probabilità non corrispondono alle reali previsioni dei bookmakers in quanto incorporano il margine di profitto dell'allibratore e il cosiddetto "longshot bias". Di conseguenza, la loro somma è maggiore di uno. In letteratura sono stati proposti differenti metodi di normalizzazione con lo scopo di ricondurre le probability odds alle previsioni sottostanti. Questo paper propone un nuovo metodo di normalizzazione per le quote fisse, chiamato CaSco (Candila-Scognamillo), applicabile in sport con soli due esiti possibili. Le performances della procedura sono state valutate utilizzando le quote su oltre 27.000 partite di tennis maschile fornite da uno dei maggiori bookmaker online. I risultati mostrano che la normalizzazione CaSco fornisce previsioni migliori rispetto agli altri metodi tradizionalmente utilizzati in letteratura.

**Key words:** Forecasting, Betting, Normalization

Vincenzo Candila,
University of Salerno, Dipartimento Scienze Economiche e Statistiche (DISES), Via Giovanni Paolo II, 132 - 84084 - Fisciano (SA), e-mail: vcandila@unisa.it

Antonio Scognamillo,
University of Florence, Dipartimento di Scienze per l'Economia e l'Impresa (DISEI), Via delle Pandette, 32 - 50127 - Florence, e-mail: antonio.scognamillo@unifi.it

## 1 Introduction

The widespread use of betting odds is due to the empirical evidence that they are the most accurate publicly available source of probability forecasts for sports [10]. This means that the sets of odds issued by bookmakers may be interpreted as implicit probabilistic forecasts of sporting events [2]. However, since the raw betting odds include the bookmaker margin and in several case a positive bias for the longshot [3] - that is, financially superior returns (i.e. smaller losses) accrue to a strategy of wagering on short-odds rather than longodds players - their inverse do not represent the actual bookmakers predictions on the event. As a result, the sum of the probabilities associated to each possible match outcome (hereafter, booksum) is typically greater than one. Thus, in order to obtain the probabilistic forecasts a normalization is needed. So far, three methods have been proposed in literature to normalize the (inverse of the) betting odds: the basic, the regression and the Shin [9, 5] normalization methods. The basic normalization consists of dividing the inverted odds by their booksum. This is the simplest and the most widespread normalization technique. However, it does not take into account the longshot bias. The normalization by regression allows to infer the outcome probabilities regressing the observed outcomes on historical betting odds. However, it requires an historical set of betting odds and match outcomes. The Shin normalization is based on the assumption that bookmakers formulate the betting odds maximizing their expected profit in a market where both uninformed bettors and a small portion of insider traders are present. [5], starting from the solution of the game proposed by Shin, reverse the problem and derive the probabilistic beliefs (i.e. the Shin probabilities) taking into account the longshot bias. However, this approach relies on a set of *a priori* assumptions about the proportion of insider traders, based on the observed spread between the odds of each possible outcome.

This paper proposes an alternative normalization method, the CaSco normalization, preserving the intuition of taking into account the longshot bias as in the Shin method. Using a dataset of over 27.000 male tennis betting quotes provided by Bet365 from 2005 to 2015, we compare the forecast accuracy of the CaSco normalization against the other normalization methods using different evaluation approaches from an in-sample perspective. Our findings show that CaSco normalization improves the forecasting ability with respect to all the other normalization approaches.

The rest of the paper proceeds as follows: Section 2 presents the methodology of the CaSco normalization. Section 3 describes the data used in this work. Section 4 is devoted to compare the forecast ability of the CaSco normalization methods against the other normalization approaches from an in-sample perspective. Section 5 concludes.

## 2 CaSco normalization

The CaSco normalization can be applied in betting markets with only two outcomes such as tennis, basketball, volleyball, and so forth. In this work, we consider the betting odds in the market of tennis because of the easy availability of these data.

Let the "margin" be the summation of the probability odds for the two players minus one, that is: $m = \pi_1 + \pi_2 - 1$. Moreover, let $d = \pi_1 - \pi_2$ be the distance between the two probability odds. Formally, the CaSco normalization calculates the probability of winning of player 1, for the $j-th$ match $p_{1,j}$, under the circumstance that player 1 is the favourite (that is, $d > 0$) as follows:

$$p_{1,j} = \begin{cases} \pi_{1,j} - 0.5 \cdot m & \text{if } d \leq \psi \\ \pi_{1,j} - \delta \cdot m & \text{if } d > \psi \end{cases}. \tag{1}$$

In (1), the key quantities are $\psi$ and $\delta$, respectively defined as the threshold and the ratio, that is the amount of the margin to subtract to the probability odd. The threshold is bounded in the interval of minimum and maximum observed distances between the probability odds entering the analysis. The ratio is bounded in the interval $]0 - 0.5]$. The sense of (1) is that the probability odds of player 1 is diminished of a quantity varying from $\delta \cdot m$ to, at most, $0.5 \cdot m$. The greater the distance $d$ is, the more the player 1 is the favourite, the less the Casco normalization subtracts the margin to the inverted quote. If, instead, $d < 0$ but however $|d| > \psi$, meaning that player 1 is the longshot, the inverted quote is diminished of a quantity greater than $0.5 \cdot m$, that is: $p_{i,j} = \pi_{1,j} - (1 - \delta) \cdot m$. Note that once calculated the probability for player 1, the other is obtained as its complementary. Moreover, the CaSco normalization does not care of which player enters (1), because it will give always consistent results: the probability that would be obtained through (1) for player 1 is the same that would be obtained if it was firstly calculated $p_{2,j}$ and then $p_{1,j}$ was obtained by $1 - p_{2,j}$. Last but not least, this kind of normalization is asymmetric. Even if the distance is smaller than the threshold $\psi$, the operation $\pi_i - 0.5 \cdot m$ alters the proportion between the original inverted quotes and the resulting normalized probabilities.

The estimation of $\psi$ and $\delta$ is carried out by the minimum mean square error (MMSE) estimator. For ease of notation, from now on, we will consider only the player 1 such that $p_{1,j}$ becomes $p_j$. We obtain $\psi$ and $\delta$ by minimizing the squared distance between the observed outcome $Y_j$ and $p_j$, with $j = 1, \cdots, J$:

$$\begin{aligned} \underset{\psi,\delta}{\arg\min} \quad & \sum_{j=1}^{J} \left(Y_j - p_{j,(\psi,\delta)}\right)^2 \\ \text{subject to} \quad & 0 \leq \delta \leq 0.5, \\ & \min(d) \leq \psi \leq \max(d). \end{aligned} \tag{2}$$

In (2), $Y_j$ denotes the observed outcome of the match $j$ for the player 1. If $Y_j = 0$, then player 1 has been defeated. Otherwise, if $Y_j = 1$, then player 1 has defeated player 2 in the match $j$. Practically speaking, the estimation of $\psi$ and $\delta$ is carried out by numerically solving equation (2).

## 3 Data and descriptive analysis

We focus on betting quotes offered from the professional bookmaker Bet365 in the market of male tennis. More precisely, we collect over 27.000 male tennis matches from Tennis Data provider[1]. The matches cover the period 2005-2015 and consider all the four Grand Slams (Australian Open, Roland Garros, Wimbledon, U.S. Open) per each year, all the ATP would tour tournaments, (ATP 250, ATP 500 and ATP 1000) plus the ATP Finals. We consider the period 2005-2012 for the in-sample evaluation and the remaining period 2013-2015 for the out-of-sample evaluation[2].

Table 1 presents the number of matches as well as the median of the margin $m$ and distance $d$ per tournament typologies and rounds.

**Table 1: Summary statistics**

| | All rounds | | | 1st rounds | | | Semifinals | | |
|---|---|---|---|---|---|---|---|---|---|
| | J | Margin | Distance | J | Margin | Distance | J | Margin | Distance |
| ATP 250 | 12249 | 0.076 | 0.356 | 5692 | 0.076 | 0.351 | 836 | 0.069 | 0.331 |
| ATP 500 | 4105 | 0.073 | 0.402 | 1807 | 0.076 | 0.376 | 229 | 0.067 | 0.428 |
| ATP 1000 | 5806 | 0.070 | 0.402 | 2278 | 0.076 | 0.351 | 181 | 0.057 | 0.484 |
| ATP Finals | 159 | 0.058 | 0.458 | - | - | - | 22 | 0.057 | 0.484 |
| Grand Slam | 5234 | 0.068 | 0.533 | 2613 | 0.069 | 0.533 | 85 | 0.056 | 0.550 |
| Tot./Median | 27553 | 0.070 | 0.402 | 12390 | 0.076 | 0.376 | 1353 | 0.067 | 0.376 |

**Notes**: Columns "J" show the number of matches, per type of tournament. Columns "Margin" show the median of the margins obtained from the summation of the inverted odds offered by Bet365 minus one, per type of tournament. Columns "Distance" show the median of the distance between the inverted odds, per type of tournament.

In the table it is clear that, as long as the tournament has a greater importance, the median of the margin decreases while the median of the distance increases. Looking at their correlation, we find a strong negative linear relationship (-0.57). That is, the greater the distance is, the clear a favourite player is, the smaller the margin is and vice-versa. This holds independently of the round or the type of the tournament considered.

## 4 Evaluating the forecasting ability of the CaSco normalization

The first column of Table 2 presents the estimated values as resulting by the MMSE estimation. Because of unknown distribution of $\hat{\delta}$ and $\hat{\psi}$, we adopt a bootstrap procedure in order to find the confidence intervals (CI). Repeating the estimation of (2) 200.000 times, we obtain the 95% CI, as highlighted in column 2 and 3 of the same table.

[1] `www.tennis-data.co.uk`

[2] Due to the limited space, the out-of-sample evaluation is not report here but it is available upon request.

**Table 2: $\delta$ and $\psi$ estimated values and 95% bootstrap confidence intervals**

| | $\hat{\theta}$ | $CI_{LB}$ | $CI_{UB}$ |
|---|---|---|---|
| $\delta$ | 0.165 | 0.022 | 0.300 |
| $\psi$ | 0.800 | 0.730 | 0.820 |

**Notes**: $CI_{LB}$ and $CI_{UB}$ denote the lower and upper bound of the 95% bootstrap confidence interval, respectively.

In order to compare the CaSco normalization method with respect to the other existing approaches, we consider the Brier score [1], the Kullback−Leibler (K-L) score ([7], eq. (2.1)), the Skill score ([6], eq. (40)) and the Nagelkerke's $R^2$ [8]. The smaller the first two scores are, the better the relative method is while the opposite holds for the last two scores.

The results of the evaluation are in Table 3. It results that the CaSco normalization has the best forecasting accuracy, regardless of the score function employed. In fact, it has the smallest distance from the true outcome, on average, and the greatest $R^2$. Furthermore, accordingly to the previous literature, we find that the normalization by regression has the worst performance [10].

**Table 3: In-sample normalization methods evaluation**

| | Brier | K-L | Skill score | $R^2$ |
|---|---|---|---|---|
| Basic | 0.1871*** | 0.5531*** | 0.2505 | 0.3246 |
| Regression | 0.1868*** | 0.5516*** | 0.2518 | 0.3277 |
| Shin | 0.1884*** | 0.5587*** | 0.2456 | 0.3133 |
| CaSco | **0.1868***** | **0.5512***** | **0.2519** | **0.3284** |

**Notes**: *** denote significance at the 1% levels.

The improvement of the forecasting performances guaranteed by the CaSco normalization is also confirmed by a Diebold-Mariano (DM) test, in the version proposed for binary outcomes by [4] (eq. (11)). In particular, we consider the two-tailed test, comparing the performance of the CaSco with respect to the other three existing methods, taking into account both the Brier and the K-L loss functions. Table 4 shows that the null hypothesis of equal predictive ability between the CaSco and each of the other alternative method is always rejected. Moreover, being the DM statistics negative, the CaSco procedure is always preferred to the Basic, the regression and the Shin normalization methods.

## 5 Conclusions

In this paper, we have proposed a new normalization approach to derive the probabilistic forecasts from the betting odds offered by the professional bookmakers in

**Table 4: Differences evaluation**

| Loss | Basic | Regression | Shin |
|---|---|---|---|
| Brier | -4.0487*** | -2.0767** | -6.5408*** |
| K-L | -4.6480*** | -2.0735** | -5.6388*** |

**Notes**: The table reports the DM statistics of the two tailed test of equal predictive accuracy between the CaSco and each model in column, using the loss function in row.
** and *** denote significance at the 5% and 1% levels. respectively.

sports with only two outcomes. Applying the CaSco normalization to the market of tennis male matches has shown that the new approach has a superior forecasts ability with respect to other methods regardless of the score function employed. The improvement of the forecasting performances guaranteed by the CaSco normalization is also confirmed by a Diebold-Mariano (DM) two-tailed test which always rejects the null hypothesis of equal predictive ability between the CaSco and each of the other alternative methods.

## References

1. Glenn W Brier. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.
2. David Forrest, John Goddard, and Robert Simmons. Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, 21(3):551–564, 2005.
3. David Forrest and Ian McHale. Anyone for tennis? *British Journal of Sports Medicine*, 37(4):283, 2003.
4. Tilmann Gneiting and Matthias Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151, 2014.
5. Bruno Jullien and Bernard Salanie. Measuring the Incidence of Insider Trading: A Comment on Shin. *The Economic Journal*, 104(427):1418, November 1994.
6. Kajal Lahiri and Liu Yang. Forecasting binary outcomes. *Handbook of economic forecasting*, 2, 2012.
7. Tze Leung Lai, Shulamith T Gross, David Bo Shen, et al. Evaluating probability forecasts. *The Annals of Statistics*, 39(5):2356–2382, 2011.
8. Nico JD Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991.
9. Hyun Song Shin. Optimal Betting Odds Against Insider Traders. *The Economic Journal*, 101(408):1179–1185, 1991.
10. Erik Štrumbelj. On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30(4):934–943, 2014.