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. (Co-author(s): S. Freitag-Wolf, A. Dempfle, W. Lieb, M. Krawczak)

(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): B. Cavalletti, 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. (Co-author(s): É. Lesage)

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, G. Biagi, 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. W. 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. Böhning)

(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)

Comparing Prediction Intervals in Quantile and OLS Regression

Un'analisi comparativa degli intervalli di previsione nella regressione quantile e OLS

Cristina Davino and Domenico Vistocco

Abstract In the regression framework, prediction intervals are a valuable tool to estimate the value of the response variable. Such prediction intervals can be formulated in terms of the expected value of the response variable as well as for a single specific value. Both the type of intervals suffer of violations of the assumptions of the classical regression models, resulting in empirical coverage levels not consistent with the nominal levels. Among the several possibilities proposed in literature to face this problem, we consider the estimations provided by quantile regression at two different quantiles to obtain prediction intervals. Exploiting the non parametric nature of quantile regression, such intervals are useful in situations characterised by heteroscedasticity or when the response variable is skewed.

Abstract *Abstract Una delle applicazioni più utilizzate dei modelli di regressione è rappresentata dagli intervalli di previsione. Tali intervalli possono riguardare la media condizionata della variabile di risposta o un valore puntuale della variabile dipendente. In entrambi i casi eventuali violazioni delle assunzioni del classico modello di regressione possono comportare livelli di copertura empirica non consistenti rispetto al livello nominale. Tra i diversi contributi presenti in letteratura per affrontare tali situazioni, il lavoro si concentra sulle stime ottenute dalla regressione quantile per il valore di previsione della distribuzione condizionata della variabile di risposta. La natura non parametrica della regressione quantile, consente di affrontare gli intervalli di previsione anche in situazioni di eteroschedasticità o di variabili di risposta asimmetriche.*

Key words: Prediction intervals, OLS regression, quantile regression

---

Cristina Davino
Dip.to di Scienze Politiche, della Comunicazione e delle Relazioni Internazionali, Università di Macerata e-mail: cristina.davino@unimc.it

Domenico Vistocco
Dip.to di Economia e Giurisprudenza, Università di Cassino e del Lazio Meridionale e-mail: vistocco@unicas.it

1 Methodological Framework

The need for robust statistics alternative to least squares estimation dates back to the nineteenth century as discussed in several important surveys [9] [4]. In this framework, quantile regression (QR), introduced by Koenker and Basset [6] [5] represents a turning point. QR can be considered the extension of ordinary least squares (OLS) to the estimation of a set of conditional quantile functions. It allows to verify if the effect played by the regressors varies on the low, middle and upper parts of the dependent variable thus suggesting different interpretation paths and revealing a scale and/or shape effect.

The QR model for a given conditional quantile θ can be formulated as follows:

$$Q_\theta\left(\hat{\mathbf{y}}|\mathbf{X}\right) = \mathbf{X}\hat{\beta}\left(\theta\right) \tag{1}$$

where $\mathbf{y}$ is the response variable observed on n individuals, $\mathbf{X}$ is a matrix with as many columns as the number of regressors plus a vector of ones for the intercept estimation, $0 < \theta < 1$ is the θ^{th} quantile and $Q_\theta(.|.)$ denotes the conditional quantile function for the θ^{th} quantile.

Although different functional forms can be used, this paper deals only with linear regression models.

QR estimators are asymptotically normally distributed with a covariance matrix that depends on the model assumptions (independent and identically distributed errors or non-identically distributed errors) [7] [1]. Resampling methods can represent a valid alternative to the asymptotic inference (among many see [8]) because they allow the estimation of parameter standard errors without requiring any assumption in relation to the error distribution.

In this paper we present a comparison between OLS and QR prediction intervals. In regression, it is possible to distinguish prediction intervals for the expected value of the response variable (the so called confidence intervals for $\hat{y}$) and prediction intervals for a single response $\hat{y}$. It is well known that prediction intervals are much wider than confidence intervals because there is more uncertainty when predicting the dependent variable than when predicting its mean: an individual observation is more variable than a summary statistic such as the mean computed from several observations. In the following we refer to the first type of intervals, i.e. prediction intervals for the expected value of the response variable.

Moreover, it is worthwhile to mention that prediction intervals can be computed both in a parametric and in a non parametric setting. The former requires strong assumptions for the error distribution and they cannot benefit of large sample approximations. Prediction intervals are then constructed to be symmetric around the sample mean, and account for uncertainty in both the estimated mean and standard deviation. On the other hand, non parametric prediction intervals are based on the percentile method and they are obtained using empirical quantiles of the vector of parameter estimates obtained though several replications of the data set (e.g. using resampling method such as the bootstrap or Monte Carlo simulations).

In the OLS framework, the construction of prediction intervals becomes critical in case of violations against the normal distribution as often occurs in real data (for example in case of outliers or skewness dependent variables). Such features have a strong impact on the construction of prediction intervals.

Notwithstanding several contributes have been proposed in the framework of robust alternatives to least squares [3], QR represents an even more attractive alternative solution as it is able to handle regression models with heteroscedastic relationships, skewed dependent variables and outliers. The interval obtained considering two distinct quantile estimates $\hat{q}_y\ (\theta_1, X = x)$ and $\hat{q}_y\ (\theta_2, X = x)$, at any specific value of the regressor X, can be indeed interpreted as a $(\theta_2 - \theta_1)\%$ prediction interval. Exploiting the QR features, such and interval does not suffer departures from the classical assumptions of the normal linear model.

2 A comparative study

The aim of the contribution is to compare OLS and QR prediction intervals. A comparative study takes into account models with different distributions of the error term. The comparison is achieved in terms of empirical coverage and interval widths.

A set of 10 artificial models is used to compare prediction intervals in QR and OLS regression:

$model_1 \rightarrow \mathbf{y}^{(1)} = 1 + 2\mathbf{x} + \mathbf{e}_N$ $\quad$ $model_2 \rightarrow \mathbf{y}^{(2)} = 1 + 2\mathbf{x} + \mathbf{e}_{LN_1}$

$model_3 \rightarrow \mathbf{y}^{(3)} = 1 + 2\mathbf{x} + \mathbf{e}_{LN_2}$ $\quad$ $model_4 \rightarrow \mathbf{y}^{(4)} = 1 + 2\mathbf{x} + \mathbf{e}_{LN_3}$

$model_5 \rightarrow \mathbf{y}^{(5)} = 1 + 2\mathbf{x} - \mathbf{e}_{LN_3}$ $\quad$ $model_6 \rightarrow \mathbf{y}^{(6)} = 1 + 2\mathbf{x} + (1 + \mathbf{x})\mathbf{e}_N$

$model_7 \rightarrow \mathbf{y}^{(7)} = 1 + 2\mathbf{x} + \mathbf{e}_{AR(1)[\rho=-0.2]}$ $\quad$ $model_8 \rightarrow \mathbf{y}^{(8)} = 1 + 2\mathbf{x} + \mathbf{e}_{AR(1)[\rho=+0.2]}$

$model_9 \rightarrow \mathbf{y}^{(9)} = 1 + 2\mathbf{x} + \mathbf{e}_{AR(1)[\rho=-0.5]}$ $\quad$ $model_{10} \rightarrow \mathbf{y}^{(10)} = 1 + 2\mathbf{x} + \mathbf{e}_{AR(1)[\rho=+0.5]}$

Data differ with respect to the error term which can be homogeneous (models from 1 to 5), heterogeneous (model 7) or related to dependent data (models from 8 to 10). In particular the error terms are defines as follows:

$\mathbf{e}_N \sim N(\mu = 0, \sigma = 1)$; $\quad$ $\mathbf{e}_{LN_1} \sim LN(\mu = 0, \sigma = 0.25)$;

$\mathbf{e}_{LN_2} \sim LN(\mu = 0, \sigma = 0.5)$; $\quad$ $\mathbf{e}_{LN_3} \sim LN(\mu = 0, \sigma = 1.25)$;

$\mathbf{e}_{AR(1)_{[rho]}} \rightarrow e_i = \rho e_{i-1} + a_i$.

where N is a standard normal variable, LN a log-normal variable with location parameter $\mu = 0$ and $\mathbf{e}_{AR(1)_{[rho]}} \rightarrow e_i = \rho e_{i-1} + a_i$.

In the group of the homogeneous models, $model_1$ respects the assumptions of the classical normal regression framework [2], while the other four models are characterized by a violation of the classical assumptions with respect to the presence of a symmetric not normal error, ($model_2$), or asymmetric errors with different degrees of skewness ($model_3$ and $model_4$). Finally, $model_5$ has the same error term of $model_4$, but it enters in the model with a different sign. $model_6$ represents a classical case of increasing variance of the error term and thus of violation of the

homoskedastic assumption. The four dependent data models are based on an autoregressive error term different according to the ρ values (-0.2, +0.2, -0.5, +0.5).

For each model, 1000 random samples were generated and prediction intervals were computed using both OLS and QR method.

A comparison between nominal and empirical coverage levels for the 10 models is carried out using a confidence level (1-α)=0.8 and five distinct values of X, (8, 9, 10, 11, 12), spanning along the range of the regressor variable. For obtaining QR prediction intervals, $\hat{q}_y$ ($\theta_2 = 0.9, X = x$) and $\hat{q}_y$ ($\theta_1 = 0.2, X = x$) have been considered.

Results are shown in Figure 1 which is organized in 10 panels, one for each model. In each panel, the solid line depicts the OLS empirical coverage level, while the QR level is shown using a dashed line. The horizontal line at $y = 0.8$ denotes the nominal coverage level. The graph immediately illustrates the best performance of QR prediction intervals: the empirical coverage lines are very close to the nominal levels for almost all 10 models in the QR panel, minor differences are present in case of dependent error models. OLS empirical level is close to the nominal level only for $model_1$, it becomes worse as the skewness of the error term increases. With respect to $model_6$ OLS coverage shows an increasing trend moving from lower to higher X values. Finally, the two methods provide closer results for the error dependent models, i.e. from $model_7$ up to $model_{10}$, although also for these models, QR outperforms OLS.

The average interval widths, shown in Table 1, indicate how QR intervals offer better results also with respect to the interval precision, providing narrower, and therefore, more informative, intervals. Also, for the interval width, the strong differences between the OLS and QR methods in the case of the models with skew error terms ($model_3$, $model_4$ and $model_5$) and for the heterogeneous error model ($model_6$) are staring us in the face.

Further simulations will be carried out to examine the effect played by the sample size, different empirical coverages and presence of outliers. Finally QR results will be compared to OLS results corrected for facing violations of the classical assumptions.

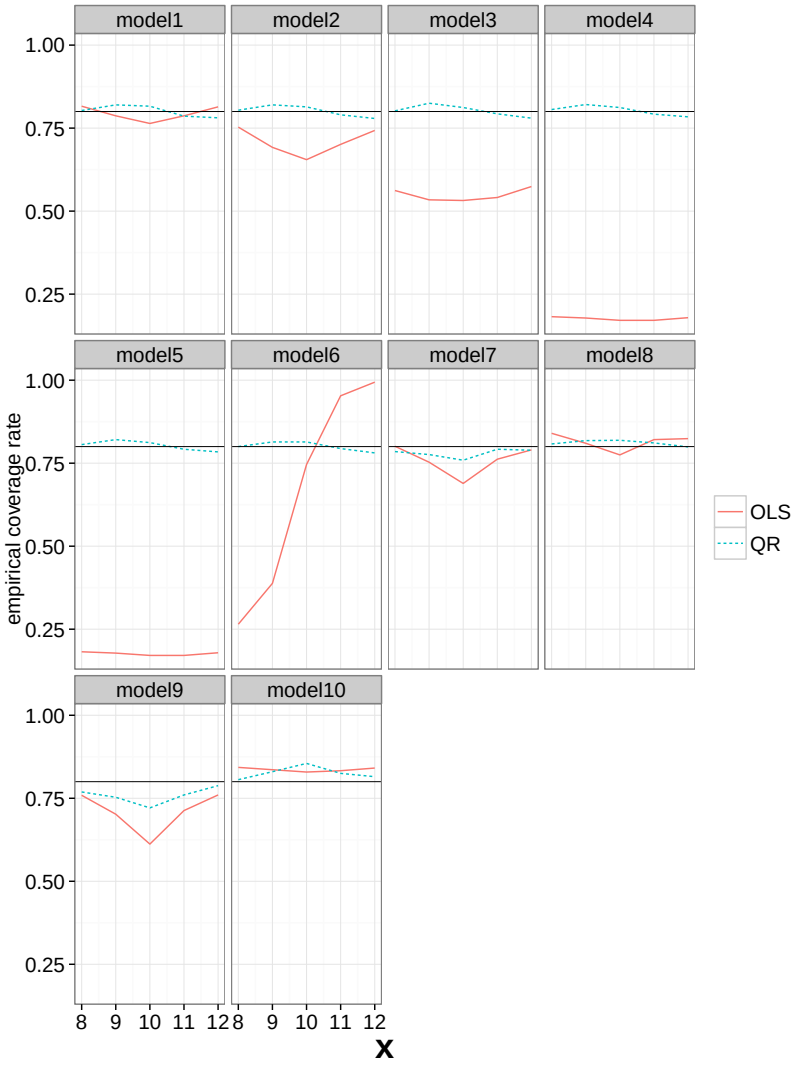

Fig. 1 Empirical coverage levels computed on 1000 simulated datasets from the 10 different models.

Table 1 Average interval width for OLS and QR prediction intervals computed using 1000 random samples extracted from each of the 10 considered models.

		Average interval width [Nominal coverage level $(1-\alpha) = 80\%$]				
		$x = 8$	$x = 9$	$x = 10$	$x = 11$	$x = 12$
$model_1$	LS	2.63	2.59	2.57	2.59	2.63
	QR	2.51	2.52	2.53	2.55	2.56
$model_2$	LS	0.69	0.68	0.67	0.68	0.69
	QR	0.64	0.64	0.64	0.65	0.65
$model_3$	LS	1.57	1.55	1.54	1.55	1.57
	QR	1.35	1.36	1.36	1.37	1.37
$model_4$	LS	9.71	9.57	9.52	9.57	9.71
	QR	4.84	4.85	4.86	4.87	4.87
$model_5$	LS	9.71	9.57	9.52	9.57	9.71
	QR	4.84	4.85	4.86	4.87	4.87
$model_6$	LS	29.03	28.60	28.45	28.60	29.02
	QR	22.70	25.30	27.90	30.50	33.09
$model_7$	LS	2.69	2.65	2.64	2.65	2.69
	QR	2.61	2.61	2.60	2.60	2.59
$model_8$	LS	2.68	2.64	2.62	2.64	2.68
	QR	2.60	2.59	2.58	2.58	2.57
$model_9$	LS	3.06	3.01	3.00	3.01	3.06
	QR	2.98	2.97	2.96	2.95	2.94
$model_{10}$	LS	3.02	2.97	2.96	2.97	3.02
	QR	2.92	2.92	2.91	2.91	2.90

References

1. Davino, C., Furno, M., Vistocco, D.: Quantile Regression: Theory and Applications. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Inc. (2013)
2. Gujarati, D.N.: Basic Econometrics. McGrawHill, International Edition (2003)
3. Horn, P.S.: Robust quantile estimators for skewed populations. Biometrika **77**, 631–636 (1990)
4. Huber, P.J.: Robust Statistics: A Review. Annals of Mathematical Statistics **43**, 1041–67 (1972)
5. Koenker, R.: Quantile Regression. Cambridge University Press (2005)
6. Koenker, R.W., Basset, G.: Regression Quantiles. Econometrica, **46**(1) (1978)
7. Koenker, R.W., Basset, G.: Robust tests for heteroscedasticity based on regression quantiles. Econometrica **50**(1) (1982)
8. Kocherginsky, M., He, X., Mu, Y.: Practical Confidence Intervals for Regression Quantiles. Journal of Computational and Graphical Statistics **14**(1), 41–55 (2005)
9. Stigler, S.M.: Simon Newcomb, Percy Daniell, and the History of Robust Estimation 1885-1920. Journal of the American Statistical Association **68**(344), 872–879 (1973)

Comparing Prediction Intervals in Quantile and OLS Regression

Un'analisi comparativa degli intervalli di previsione nella regressione quantile e OLS

Cristina Davino and Domenico Vistocco

Abstract In the regression framework, prediction intervals are a valuable tool to estimate the value of the response variable. Such prediction intervals can be formulated in terms of the expected value of the response variable as well as for a single specific value. Both the type of intervals suffer of violations of the assumptions of the classical regression models, resulting in empirical coverage levels not consistent with the nominal levels. Among the several possibilities proposed in literature to face this problem, we consider the estimations provided by quantile regression at two different quantiles to obtain prediction intervals. Exploiting the non parametric nature of quantile regression, such intervals are useful in situations characterised by heteroscedasticity or when the response variable is skewed.

Abstract *Abstract Una delle applicazioni più utilizzate dei modelli di regressione è rappresentata dagli intervalli di previsione. Tali intervalli possono riguardare la media condizionata della variabile di risposta o un valore puntuale della variabile dipendente. In entrambi i casi eventuali violazioni delle assunzioni del classico modello di regressione possono comportare livelli di copertura empirica non consistenti rispetto al livello nominale. Tra i diversi contributi presenti in letteratura per affrontare tali situazioni, il lavoro si concentra sulle stime ottenute dalla regressione quantile per il valore di previsione della distribuzione condizionata della variabile di risposta. La natura non parametrica della regressione quantile, consente di affrontare gli intervalli di previsione anche in situazioni di eteroschedasticità o di variabili di risposta asimmetriche.*

Key words: Prediction intervals, OLS regression, quantile regression

---

Cristina Davino
Dip.to di Scienze Politiche, della Comunicazione e delle Relazioni Internazionali, Università di Macerata e-mail: cristina.davino@unimc.it

Domenico Vistocco
Dip.to di Economia e Giurisprudenza, Università di Cassino e del Lazio Meridionale e-mail: vistocco@unicas.it

1 Methodological Framework

The need for robust statistics alternative to least squares estimation dates back to the nineteenth century as discussed in several important surveys [9] [4]. In this framework, quantile regression (QR), introduced by Koenker and Basset [6] [5] represents a turning point. QR can be considered the extension of ordinary least squares (OLS) to the estimation of a set of conditional quantile functions. It allows to verify if the effect played by the regressors varies on the low, middle and upper parts of the dependent variable thus suggesting different interpretation paths and revealing a scale and/or shape effect.

The QR model for a given conditional quantile θ can be formulated as follows:

$$Q_\theta\left(\hat{\mathbf{y}}|\mathbf{X}\right) = \mathbf{X}\hat{\beta}\left(\theta\right) \tag{1}$$

where $\mathbf{y}$ is the response variable observed on n individuals, $\mathbf{X}$ is a matrix with as many columns as the number of regressors plus a vector of ones for the intercept estimation, $0 < \theta < 1$ is the θ^{th} quantile and $Q_\theta(.|.)$ denotes the conditional quantile function for the θ^{th} quantile.

Although different functional forms can be used, this paper deals only with linear regression models.

QR estimators are asymptotically normally distributed with a covariance matrix that depends on the model assumptions (independent and identically distributed errors or non-identically distributed errors) [7] [1]. Resampling methods can represent a valid alternative to the asymptotic inference (among many see [8]) because they allow the estimation of parameter standard errors without requiring any assumption in relation to the error distribution.

In this paper we present a comparison between OLS and QR prediction intervals. In regression, it is possible to distinguish prediction intervals for the expected value of the response variable (the so called confidence intervals for $\hat{y}$) and prediction intervals for a single response $\hat{y}$. It is well known that prediction intervals are much wider than confidence intervals because there is more uncertainty when predicting the dependent variable than when predicting its mean: an individual observation is more variable than a summary statistic such as the mean computed from several observations. In the following we refer to the first type of intervals, i.e. prediction intervals for the expected value of the response variable.

Moreover, it is worthwhile to mention that prediction intervals can be computed both in a parametric and in a non parametric setting. The former requires strong assumptions for the error distribution and they cannot benefit of large sample approximations. Prediction intervals are then constructed to be symmetric around the sample mean, and account for uncertainty in both the estimated mean and standard deviation. On the other hand, non parametric prediction intervals are based on the percentile method and they are obtained using empirical quantiles of the vector of parameter estimates obtained though several replications of the data set (e.g. using resampling method such as the bootstrap or Monte Carlo simulations).

In the OLS framework, the construction of prediction intervals becomes critical in case of violations against the normal distribution as often occurs in real data (for example in case of outliers or skewness dependent variables). Such features have a strong impact on the construction of prediction intervals.

Notwithstanding several contributes have been proposed in the framework of robust alternatives to least squares [3], QR represents an even more attractive alternative solution as it is able to handle regression models with heteroscedastic relationships, skewed dependent variables and outliers. The interval obtained considering two distinct quantile estimates $\hat{q}_y\ (\theta_1, X = x)$ and $\hat{q}_y\ (\theta_2, X = x)$, at any specific value of the regressor X, can be indeed interpreted as a $(\theta_2 - \theta_1)\%$ prediction interval. Exploiting the QR features, such and interval does not suffer departures from the classical assumptions of the normal linear model.

2 A comparative study

The aim of the contribution is to compare OLS and QR prediction intervals. A comparative study takes into account models with different distributions of the error term. The comparison is achieved in terms of empirical coverage and interval widths.

A set of 10 artificial models is used to compare prediction intervals in QR and OLS regression:

$model_1 \rightarrow \mathbf{y}^{(1)} = 1 + 2\mathbf{x} + \mathbf{e}_N$ $\quad$ $model_2 \rightarrow \mathbf{y}^{(2)} = 1 + 2\mathbf{x} + \mathbf{e}_{LN_1}$

$model_3 \rightarrow \mathbf{y}^{(3)} = 1 + 2\mathbf{x} + \mathbf{e}_{LN_2}$ $\quad$ $model_4 \rightarrow \mathbf{y}^{(4)} = 1 + 2\mathbf{x} + \mathbf{e}_{LN_3}$

$model_5 \rightarrow \mathbf{y}^{(5)} = 1 + 2\mathbf{x} - \mathbf{e}_{LN_3}$ $\quad$ $model_6 \rightarrow \mathbf{y}^{(6)} = 1 + 2\mathbf{x} + (1 + \mathbf{x})\mathbf{e}_N$

$model_7 \rightarrow \mathbf{y}^{(7)} = 1 + 2\mathbf{x} + \mathbf{e}_{AR(1)[\rho=-0.2]}$ $\quad$ $model_8 \rightarrow \mathbf{y}^{(8)} = 1 + 2\mathbf{x} + \mathbf{e}_{AR(1)[\rho=+0.2]}$

$model_9 \rightarrow \mathbf{y}^{(9)} = 1 + 2\mathbf{x} + \mathbf{e}_{AR(1)[\rho=-0.5]}$ $\quad$ $model_{10} \rightarrow \mathbf{y}^{(10)} = 1 + 2\mathbf{x} + \mathbf{e}_{AR(1)[\rho=+0.5]}$

Data differ with respect to the error term which can be homogeneous (models from 1 to 5), heterogeneous (model 7) or related to dependent data (models from 8 to 10). In particular the error terms are defines as follows:

$\mathbf{e}_N \sim N(\mu = 0, \sigma = 1)$; $\quad$ $\mathbf{e}_{LN_1} \sim LN(\mu = 0, \sigma = 0.25)$;

$\mathbf{e}_{LN_2} \sim LN(\mu = 0, \sigma = 0.5)$; $\quad$ $\mathbf{e}_{LN_3} \sim LN(\mu = 0, \sigma = 1.25)$;

$\mathbf{e}_{AR(1)_{[rho]}} \rightarrow e_i = \rho e_{i-1} + a_i$.

where N is a standard normal variable, LN a log-normal variable with location parameter $\mu = 0$ and $\mathbf{e}_{AR(1)_{[rho]}} \rightarrow e_i = \rho e_{i-1} + a_i$.

In the group of the homogeneous models, $model_1$ respects the assumptions of the classical normal regression framework [2], while the other four models are characterized by a violation of the classical assumptions with respect to the presence of a symmetric not normal error, ($model_2$), or asymmetric errors with different degrees of skewness ($model_3$ and $model_4$). Finally, $model_5$ has the same error term of $model_4$, but it enters in the model with a different sign. $model_6$ represents a classical case of increasing variance of the error term and thus of violation of the

homoskedastic assumption. The four dependent data models are based on an autoregressive error term different according to the ρ values (-0.2, +0.2, -0.5, +0.5).

For each model, 1000 random samples were generated and prediction intervals were computed using both OLS and QR method.

A comparison between nominal and empirical coverage levels for the 10 models is carried out using a confidence level (1-α)=0.8 and five distinct values of X, (8, 9, 10, 11, 12), spanning along the range of the regressor variable. For obtaining QR prediction intervals, $\hat{q}_y$ ($\theta_2 = 0.9, X = x$) and $\hat{q}_y$ ($\theta_1 = 0.2, X = x$) have been considered.

Results are shown in Figure 1 which is organized in 10 panels, one for each model. In each panel, the solid line depicts the OLS empirical coverage level, while the QR level is shown using a dashed line. The horizontal line at $y = 0.8$ denotes the nominal coverage level. The graph immediately illustrates the best performance of QR prediction intervals: the empirical coverage lines are very close to the nominal levels for almost all 10 models in the QR panel, minor differences are present in case of dependent error models. OLS empirical level is close to the nominal level only for $model_1$, it becomes worse as the skewness of the error term increases. With respect to $model_6$ OLS coverage shows an increasing trend moving from lower to higher X values. Finally, the two methods provide closer results for the error dependent models, i.e. from $model_7$ up to $model_{10}$, although also for these models, QR outperforms OLS.

The average interval widths, shown in Table 1, indicate how QR intervals offer better results also with respect to the interval precision, providing narrower, and therefore, more informative, intervals. Also, for the interval width, the strong differences between the OLS and QR methods in the case of the models with skew error terms ($model_3$, $model_4$ and $model_5$) and for the heterogeneous error model ($model_6$) are staring us in the face.

Further simulations will be carried out to examine the effect played by the sample size, different empirical coverages and presence of outliers. Finally QR results will be compared to OLS results corrected for facing violations of the classical assumptions.

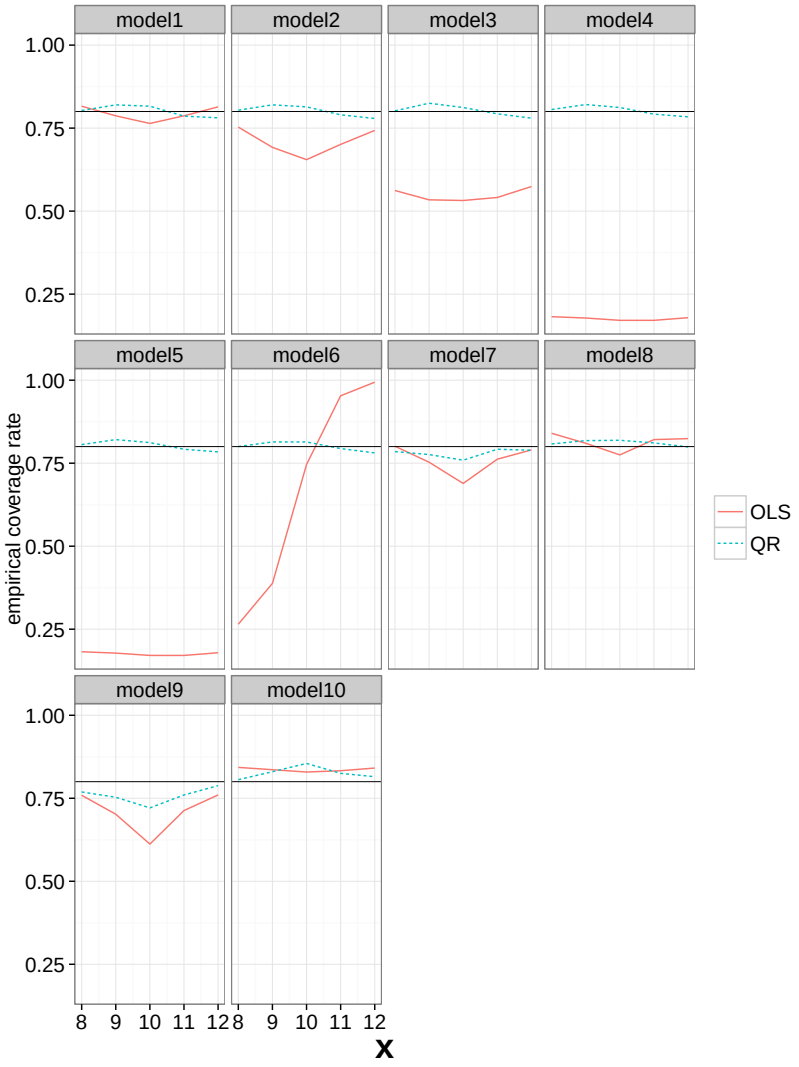

Fig. 1 Empirical coverage levels computed on 1000 simulated datasets from the 10 different models.

Table 1 Average interval width for OLS and QR prediction intervals computed using 1000 random samples extracted from each of the 10 considered models.

		Average interval width [Nominal coverage level $(1-\alpha) = 80\%$]				
		$x = 8$	$x = 9$	$x = 10$	$x = 11$	$x = 12$
$model_1$	LS	2.63	2.59	2.57	2.59	2.63
	QR	2.51	2.52	2.53	2.55	2.56
$model_2$	LS	0.69	0.68	0.67	0.68	0.69
	QR	0.64	0.64	0.64	0.65	0.65
$model_3$	LS	1.57	1.55	1.54	1.55	1.57
	QR	1.35	1.36	1.36	1.37	1.37
$model_4$	LS	9.71	9.57	9.52	9.57	9.71
	QR	4.84	4.85	4.86	4.87	4.87
$model_5$	LS	9.71	9.57	9.52	9.57	9.71
	QR	4.84	4.85	4.86	4.87	4.87
$model_6$	LS	29.03	28.60	28.45	28.60	29.02
	QR	22.70	25.30	27.90	30.50	33.09
$model_7$	LS	2.69	2.65	2.64	2.65	2.69
	QR	2.61	2.61	2.60	2.60	2.59
$model_8$	LS	2.68	2.64	2.62	2.64	2.68
	QR	2.60	2.59	2.58	2.58	2.57
$model_9$	LS	3.06	3.01	3.00	3.01	3.06
	QR	2.98	2.97	2.96	2.95	2.94
$model_{10}$	LS	3.02	2.97	2.96	2.97	3.02
	QR	2.92	2.92	2.91	2.91	2.90

References

1. Davino, C., Furno, M., Vistocco, D.: Quantile Regression: Theory and Applications. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Inc. (2013)
2. Gujarati, D.N.: Basic Econometrics. McGrawHill, International Edition (2003)
3. Horn, P.S.: Robust quantile estimators for skewed populations. Biometrika **77**, 631–636 (1990)
4. Huber, P.J.: Robust Statistics: A Review. Annals of Mathematical Statistics **43**, 1041–67 (1972)
5. Koenker, R.: Quantile Regression. Cambridge University Press (2005)
6. Koenker, R.W., Basset, G.: Regression Quantiles. Econometrica, **46**(1) (1978)
7. Koenker, R.W., Basset, G.: Robust tests for heteroscedasticity based on regression quantiles. Econometrica **50**(1) (1982)
8. Kocherginsky, M., He, X., Mu, Y.: Practical Confidence Intervals for Regression Quantiles. Journal of Computational and Graphical Statistics **14**(1), 41–55 (2005)
9. Stigler, S.M.: Simon Newcomb, Percy Daniell, and the History of Robust Estimation 1885-1920. Journal of the American Statistical Association **68**(344), 872–879 (1973)