Preface

1. Introduction to the Special Issue on the New Techniques and Technologies

The fast technological progress society has been observing for several years has led to substantial changes in every aspect of work and daily life. Obviously, scientific research has also been strongly affected. One of the main aspects of this process is the availability of huge amounts of data suitable to be collected and analysed by both established and frontier methods. Many disciplines have benefited from this recent availability of data, sometimes leading to new fields of study, especially in the economic and social sciences. Studies on time-use exploiting data from mobile devices, social network analyses, consumer/user profiling, and new classification methodologies are just a few examples. Thanks to these advances, production of official statistics has already gained and is likely to gain more benefits in the future.

One of the immediate consequences of the development of these new technologies is the possibility to reduce the update delay of archives and official registers and to enhance their integration. National Statistical Offices (NSOs) – which have always used archives and registers for their surveys – have seized the opportunity offered by new technologies to make the best use of constantly updated information.

Techniques for data matching are developing rapidly, specifically in order to obtain variables and indicators based on more than one source. Subsequently, NSOs have been required to redesign their surveys to take into account new integration opportunities.

Some of the consequent advantages are obvious: the possibility to provide prompt official information, as in the case of flash forecasts; a reduced need for revision; an exploration of new data previously difficult to obtain; the introduction of new indicators; greater accuracy in data collection; and lower costs. Such innovations as disclosure issues, dissemination of new information, definition of new evaluation procedures and guidelines, and measurement and communication of uncertainty in official statistics have also entailed new prominent challenges.

The New Techniques and Technologies (NTTS) conferences, organised by Eurostat every three years from 1992 to 2001, and then every two years since 2003, have witnessed this evolution. The issues of Big Data in all related aspects, archive linking, and new source integration have been addressed in an increasing number of contributions, available in two NTTS special issues of Journal of Official Statistics (2015, 31:2; 2018, 34:4).

The program of the 2019 Conference, to which nearly 150 papers were submitted, fully reflects the extent of the pervasive entry of new technologies into official statistics. Twelve out of 32 parallel sessions refer to topics directly or indirectly related to big data and machine learning, acquisition and processing of mobile data, statistical matching, data integration, multisource statistics, administrative data, and registers.

Following the conference call for papers, the present JOS special issue embodies the positive outcome of NTTS 2019. The articles published in this volume well represent the great variety of topics and operational and methodological proposals discussed during the conference. These contributions are basically of two types: (1) articles that aim to present methodological approaches and results from several (European) projects, and (2) methodological proposals for official statistics production.

1.1. Dissemination

With regard to dissemination, the contribution by Máténé Bella and Ritzlné Kazimir, “*A Structural Equation Model for Measuring Relative Development of Hungarian Counties in the Years 1994–2016*”, presents an investigation on the economic development of the Hungarian regions at NUTS3 level. The model assumes that the development depends on observable economic variables, as well as on other non-economic, social, demographic, cultural, and infrastructural measures, evaluated based on regional indicators published by the Hungarian Central Statistical Office. Main findings constitute a starting point to investigate how the regional development has changed in Hungary after the regime change, and how these variables are influenced by the country’s integration into the global value chain.

The article by Mazzi, Mitchell and Carausu, “*Measuring and Communicating the Uncertainty in Official Economic Statistics*”, illustrates the results achieved so far in the context of the Eurostat Comunikos Project (COMmunicating UNCertainty In Key Official Statistics). The article reviews different categorisations of data uncertainty, distinguishes sampling from non-sampling errors, and introduces a novel typology. The importance of measuring and communicating data uncertainty is emphasised as it is difficult to measure some sources of uncertainty itself, especially those relevant for administrative and big data sets.

The study by Bacchini, Baldazzi, De Carli, Di Biagio, Savioli, Sorvillo and Tinto, “*The Evolution of the Italian Framework to Measure Well-Being*”, points out some of the outcomes of a joint project between the Italian Statistical Institute (Istat) and the National Council for Economics and Labor (CNEL) that aims to establish a multi-dimensional framework for measuring “equitable and sustainable well-being” (Bes). Since 2013, Istat has published an annual report on well-being, disseminating constantly updated Bes indicators, thereby allowing more effective communication. Bes indicators are taken into account in Italian government budget documents, as established by law. The Italian Bes experience may be considered to be one of the most relevant practices at the European level, and displays the potential to become a benchmark for other countries.

The article “*Improving Time Use Measurement with Personal Big Data Collection – The Experience of the European Big Data Hackathon 2019*”, by Zeni, Bison, Reis, Gauckler and Giunchiglia, describes the main contents of the Big Data satellite event of the NTTS Conference 2019. The article illustrates a pilot study on a system, called *i-Log*, which allows to collect personal big data from smartphones’ internal sensors to be used for time-use measurement. The *i-Log* system permits to gather heterogeneous types of data, thereby creating new possibilities for urban studies. Responses are used to train machine-learning algorithms, and allow the system to learn from the user’s habits and generate new so called “time-diary answers”. This new information could be used to assess the quality of existing answers, or to fill the gaps when the user does not provide any response.

1.2. Methodological Proposals

In “*A Diagnostic for Seasonality Based Upon Polynomial Roots of ARMA Models*”, McElroy presents a new methodological approach for seasonal adjustment. In particular, the study suggests connecting the concept of seasonality to a mathematical definition regarding the oscillatory character of the moving average (MA) representation coefficients, and defines a new seasonality diagnostic based on autoregressive (AR) roots. This procedure may help to assess dynamic versus stable forms of seasonality, and to deal with arbitrary seasonal periods, for both raw and seasonally adjusted data. Uncovering over-adjustment can be achieved using the proposed extension of the AR diagnostic to an MA diagnostic. The method is illustrated through simulation studies and empirical evidence.

The two articles by Roberson and by Chessa are both examples of how the new acquisition and data processing methods may be usefully applied in the context of official statistics.

“*Applying Machine Learning for Automatic Product Categorization*” by Roberson shows how machine-learning techniques have rapidly become one of the most used tools for classification. The U.S. Bureau of Census has developed a procedure based on the proposed methodology to address the challenges of high respondent burden and low survey response rates. The presented strategy automatically classifies goods and services – with an accuracy rate of over 90% – based on product information and text descriptions associated for each sold item to the Universal Product Code. Similarly, the same classification strategy may be applied to barcode (GTIN, Global Trade Item Number) product identification. GTINs is affected by the relaunch problem, when disappearing and re-introduced items have to be linked in order to capture possible price changes.

The study by Chessa, “*A Product Match Adjusted R Squared Method for Defining Products with Transaction Data*”, is focused on a method that groups GTINs into strata (‘products’). The method balances the need of “homogeneity” of GTINs within products with the “degree” to which products can be “matched” each month with respect to a base comparison period. The proposed technique combines explained variance in product prices with product match over time, so that different stratification schemes may be ranked according to the combined measure.

Two articles are concerned with the multiple imputation of missing data and new methods of administering questionnaires in social surveys.

Scholtus and Dalmaans, in “*Variance Estimation after Mass Imputation Based on Combined Administrative and Survey Data*”, present a simulation study on artificial data and an application to the Dutch census of 2011, in order to evaluate the variance of estimated frequency tables based on mass imputation, when data may be available from both administrative sources and a sample survey. The motivating application for this article is the Dutch virtual population census, for which it has been proposed to use mass imputation to estimate tables involving educational attainment. The article also discusses a more general bootstrap method for variance estimation.

In “*Assessing and Adjusting Bias Due to Mixed-Mode in Aspect of Daily Life Survey*”, De Vitiis, Guandalini, Inglese and Terribili verify the usefulness of a mixed mode (MM) approach, that is, the use of different collection techniques in the same survey, to both contrast declining response and coverage rates and reduce surveys cost. The mode effect

can be addressed at different levels: at the design phase the best collection instruments to contain the measurement error must be defined, and at the estimation phase the mode effect must be measured and corrected. The article presents the evaluation of the mode effect in the ISTAT mixed-mode survey “Aspects of Daily Life – 2017”, in an experimental context.

The article by Alleva, Falorsi, Petrarca and Righi “*Measuring the Accuracy of Aggregates Computed from a Statistical Register*” proposes the global mean squared error (GMSE) as a measure of accuracy in the production of official statistics, based on a massive integration of administrative archives and survey data. In particular, by focusing on the estimation of a population total from measurements calculated at the unit level, and considering the main sources of uncertainty that may affect registers, the authors suggest a feasible calculation strategy for the GMSE. The performance of the proposed strategy is verified through a simulation study.

“*A Hybrid Technique for the Multiple Imputation of Survey Data*” by Razzak and Heumann discusses imputation techniques in surveys in which categorical variables suffer from a large number of missing values, and complex dependencies may exist. Since the implementation of multiple imputation (MI) techniques or chained equations (MICE) may be problematic in such situations, the authors propose a method to make dependent on categorical variables, previously imputed using latent class models, chained MI equations for continuous variables. A simulation study and a survey data example articulately describe the good performance of the proposed method with respect to alternative techniques.

2. Acknowledgments

On behalf of the Editorship, we would like to warmly thank all the authors who have submitted their manuscripts for our consideration. In addition, we would like to thank all the referees for taking the time to review all the manuscripts. Over 50 referees have contributed to the current special issue and their comments and suggestions greatly helped to improve the original versions of the manuscripts.

As a matter of fact, the participants in the NTTS conference and the readership of this special issue may belong to different fields: they may be, in fact, either academics or official statistics producers. This circumstance determines the fruitfulness and the added value of the NTTS conference, also mirrored in the variety of themes hosted by Journal of Official Statistics, thus allowing the diffusion of new proposals and the presentation of specific experiences.

Francesca Di Iorio
Guest Editor

Emanuele Baldacci
Dario Buono
Luca di Gennaro Splendore
Duncan Elliott
Rebecca Killick
Tiziana Laureti
Monica Pratesi
Natalie Shlomo
Guest Associate Editors