Italian Presidency of the Council of the European Union
italia2014.eu

FENStatS

eurostat

ESAC European Statistical Advisory Committee

Sapienza Università di Roma
Department of Statistical Sciences

SIS Società Italiana di Statistica

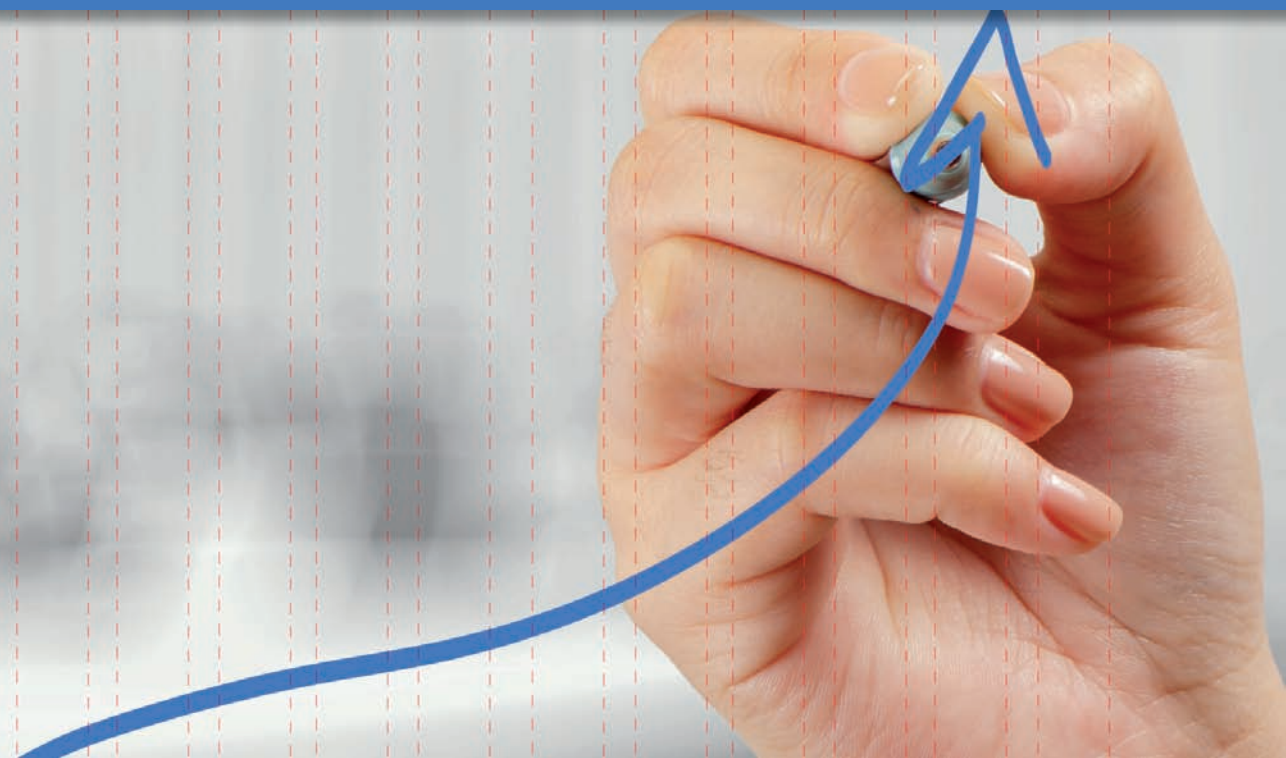
Istat Italian National Institute of Statistics

CONFERENCE OF EUROPEAN STATISTICS STAKEHOLDERS

Methodologists, Producers and Users of European Statistics

Rome 24 | 25 November 2014

ABSTRACTS



CONFERENCE OF EUROPEAN STATISTICS STAKEHOLDERS

Methodologists, Producers and Users of European Statistics

Rome 24 | 25 November 2014

Organized by
Federation of European National Statistical Societies (FENStatS)
EUROSTAT
European Statistical Advisory Committee (ESAC)
Italian National Institute of Statistics (Istat)
Società Italiana di Statistica SIS-VSP
Department of Statistical Sciences, Sapienza University of Rome

Scientific Steering Committee
Walter Radermacher, Director General of Eurostat, Co-organizer
Maurizio Vichi, President FENStatS, Co-organizer

Anne Gégout-Petit, President of the French Statistical Society,
Giorgio Alleva, President ISTAT,
Leandro Pardo LLorente, President of the Spanish Society of Statistics and Operations Research.
John Pullinger, UK National Statistician, Past President of The Royal Statistical Society,
Aurel Schubert, Director General Statistics European Central Bank,
Ineke Stoop, ESAC,
Wolfgang Schmid, President of the German Statistical Society,
Nicola Torelli, President of the Italian Statistical Society,

Scientific Program Committee
Emanuele Baldacci, Istat
Patrizia Cacioli, Istat
Roberto Barcellan, Eurostat
Martin Karlberg, Eurostat
Giuliana Coccia, SIS-Vsp
Fabio Crescenzi, SIS-Vsp
Donatella Vicari, Sapienza University of Rome

A selection of fully peer-reviewed papers will be published in the Springer Series
Selected Papers of the Statistical Societies (http://www.springer.com/series/10107).

CONTENT

CONTENT

CONTENT

CONTENT

CONTENT

CONTENT

CONTENT

CONTENT

CONTENT

Trends in the modernisation of official statistics

Mariana Kotzeva[1]

Abstract
The paper provides a short historical review of the efforts of official statisticians to modernize statistical production in response to the various methodological, technological and user driven challenges. Furthermore it considers the current modernization programmes run at national, European and global level in the historical perspective and on the basis of that highlights lessons learned, challenges ahead and the roles of users, producers and researchers in making official statistics relevant, efficient and agile.

[1] Mariana Kotzeva, Deputy Director General of Eurostat, Bulgaria; email: mariana.kotzeva@ec.europa.eu

Modernizing Modernization

Bert Kroese, Barteld Braaksma and Marton Vucsan [2]

Abstract

Statistics Netherlands has always invested a lot in innovating products and adapting processes. Major developments in the world indicate that modernisation is now more urgently needed than ever. Our current processes, methods and tools for collecting, processing and disseminating may rapidly become obsolete; or at least will need to be complemented by totally different and innovative alternatives. Are we doing enough to modernize, now that society and technology are accelerating at a rate never seen before? How should we proceed? Which challenges can we expect? In this paper we will examine these questions, starting from our current and past modernization initiatives and then looking at our plans to move forward. We argue that we must modernize our modernization.

[2] Bert Kroese, Statistics Netherlands; email: akse@cbs.nl

Modernisation of Statistical Production and Services – a Global Perspective

Steven Vale [3]

Abstract
This paper summarises the experiences of implementing the vision and strategy of the High Level Group for the Modernisation of Statistical Production and Services (HLG). This global group of chief statisticians has initiated a range of activities, driven by the needs of the official statistics community, to improve efficiency, and better meet the needs of users. The approach of "standards-based modernisation" has already delivered a number of successful results in terms of standards, models and architectures, and the focus is now on using these to deliver real benefits to statistical organisations, and their clients. The paper outlines what has been achieved so far, and what the likely next steps will be.

Key words:
standards-based modernisation, global, High Level Group.

[3] Steven Vale, UNECE; email: steven.vale@unece.org

A Business Architecture Model to foster standardisation in official statistics

Nadia Mignolli, Giulio Barcaroli, Piero Demetrio Falorsi, and Alessandra Fasano [4]

Abstract
Business Architecture (BA) is called to play a central role in a programme as complex as that of modernisation and standardisation of the official statistical information production. This study aims at illustrating a BA model for achieving a unity of views, so as to ensure the strategic alignment in each part of an Organisation and to carry out an innovation consistent with the standardisation objective that should be reached. This BA model individuates four different business lines (Strategy; Corporate support; Production; Capability) and is led by common infrastructures and principles that become instruments and guidelines for the implementation of each business line group of actions. Both principles and infrastructures facilitate and enhance the standardisation process.

Key words:
modernisation, standardisation, industrialisation, business environment, business lines, harmonisation, organisational and production processes.

[4] Nadia Mignolli, Italian National Institute of Statistics - Istat; email: mignolli@istat.it
Giulio Barcaroli, Italian National Institute of Statistics - Istat; email: barcarol@istat.it
Piero Demetrio Falorsi, Italian National Institute of Statistics - Istat; email: falorsi@istat.it
Alessandra Fasano, Italian National Institute of Statistics - Istat; email: fasano@istat.it

The EU Justice Scoreboard: Challenges in data gathering on the effectiveness of justice systems

Elsa García-Maltrás [5]

Abstract
The EU Justice Scoreboard is an information tool aiming to assist the EU and Member States to achieve more effective justice by providing objective, reliable and comparable data on the quality, independence and efficiency of justice systems in all Member States. The 2014 Justice Scoreboard confirms that the gathering of objective, comparable and reliable data on the effectiveness of justice systems covering all Member States remains a challenge.

Key words:
EU Justice Scoreboard.

References
1. The 2014 EU Justice Scoreboard, COM(2014)155 final

[5] Elsa García-Maltrás, Directorate-General Justice, European Commission, email: elsa.garcia-maltras@ec.europa.eu

Judicial Enforcement Procedures and Credit: Lessons from the Last Decade in Spain

Juan S. Mora-Sanguinetti and Miguel García-Posada [6]

Abstract
The efficacy with which the agents can enforce their contracts in an economy, and in this regard, the design and quality of the enforcement institutions, are crucial determinants for the development of the credit market. In our study we propose to analyze how local market conditions with respect to the quality of enforcement explain differences in the availability of credit in the years of the economic boom (2001-2009) in Spain. As in previous studies for other countries, this paper finds that higher inefficacy of the judicial system appears to reduce the availability of credit. This research, however, goes a step further and shows what kind of judicial procedures are those that specifically hinder credit market development. Specifically, only the quality of the "execution" stage matters. Moreover, this is the first study using real court data (at the local level) after the new Civil procedural Law in Spain of 2000.

Key words:
judicial efficacy, legal procedures, credit availability

References

1. Bae, K. H. and V. K. Goyal (2009), "Creditor Rights, Enforcement, and Bank Loans", Journal of Finance, 64(2), 823-860.
2. Chemin, M. (2009). The impact of the judiciary on entrepreneurship: Evaluation of Pakistan's Access to Justice Programme. Journal of Public Economics 93 (1-2), 114-125.
3. Círculo de Empresarios (2003). Justicia, economía y empresa, Madrid.

[6] Juan S. Mora-Sanguinetti is Senior Economist at the Banco de España – Eurosystem and lawyer. E-mail: juans.mora@bde.es. Tel. (+34) 913385197. Fax: (+34) 913385486. Banco de España. Alcala 48, 28014, Madrid.
Miguel García-Posada is Senior Economist at the Banco de España- Eurosystem and researcher at Universidad Carlos III de Madrid. Email: miguel.garcia-posada@bde.es. Tel. (+34) 913385479. Fax: (+34) 913385486. Banco de España. Alcala 48, 28014, Madrid.
We are grateful to Ildefonso Villan Criado (CGPJ) for his advice in the use of the judicial performance data. The views expressed are those of the authors and should not be attributed to the Banco de España or the Eurosystem.

Lights and shadows of the Italian judicial system: first steps towards accountability

Maria Giuseppina Muratore [7]

Abstract

When speaking about the Italian justice system, statistics have not been till now very helpful. In fact, in Italy there is not an exhaustive monitoring system in order to give a sound and panoramic overview of justice. Only recently, the measurement situation concerning civil justice is really improved, both from objective and subjective side, even if this is only the beginning. Datawarehouse implemented by Ministry of Justice allows having a better idea about what happen during a civil lawsuit, from its beginning to its end. Till now the only data regarded numbers of incoming cases, resolved cases and of suits pending. The module about civil justice, inserted in the Istat Annual Multipurpose survey on Everyday Life Aspects, give voice to citizens, that had a direct experience with the justice system, about justice evaluation, their awareness when they bring suit and costs awareness and about the obtained advantages from doing the suit. The framework reveals an inadequate justice system, where citizens are not considered as the main important subjects involved, where the system is not accountable, where norms, procedures and lobby's interest seem to be more important than actors involved in law cases. Italy also appears one of the worst European Countries as regards performance and efficiency of Courts, considering above all the processes length. Nevertheless, interviewing citizens about their satisfaction towards justice emerged a more complex and interesting panorama. The 52,1% is unsatisfied (22,8% are totally unsatisfied) versus the 44,1% of satisfied. Unsatisfied citizens underlined the need of important improvements as regards trials length, bureaucracy simplification, hearings punctuality, clarity of cases cost, judges availability, just to mention the most important ones. From legislative and organizational point of views, something changed and many other proposals are in a planning phase. There were for instance some interventions on Alternative Controversial Resolution, proposal of changes about attorneys' payments, others regarding the judicial machinery, as the geographical courts organization or some other inputs aimed at discourage the demand of justice. Furthermore there are many best practices examples widespread on the territory, best practices leading towards a more accountable system. The monitoring of these aspects and their efficacy is very important. Istat wants to do it, trying to design a more comprehensive evaluation system and its design will be illustrated in the presentation.

[7] Maria Giuseppina Muratore, Istat, Italy; email: muratore@istat.it

Measures of Risk: Advances and New Perspectives for Environmental Risk Assessment

José Miguel Angulo[8]

Abstract
Recent developments and advances in the theoretical foundations and construction of measures of risk, as well as related statistical methodology and computational aspects, are reviewed. Application, implementation and interpretation in the environmental risk assessment context, particularly in connection with structural dependence and behavior characteristics of random processes in space and/or time, are discussed. New directions for extensions and significant challenges in continuing research with reference to complexity modeling and uncertainty are highlighted.

Key words:
complexity, environmental risk assessment, measures of risk, space-time processes, uncertainty

[8] José Miguel Angulo, Department of Statistics and Operations Research, University of Granada, Spain;
email: jmangulo@ugr.es
Partially funded by Spanish Ministry of Economy and Competitiveness under grant MTM2012-32666 (co-financed by FEDER)

Use of atmospheric pollution data at INERIS, the French National Institute for Industrial Environment and Risks, to support the assessment and management of air quality

Laure Malherbe[9]

Abstract
The development of environmental monitoring in compliance with the European legislation, the European reporting obligations and increased exchange of information constantly add to the amount and availability of environmental data. Growing modelling capacities also make it possible to build up large databases of numerical simulation outputs which can be used in combination with measurement data. To give technical support to the Ministry in charge of the environment and as part of its service public mission, INERIS is involved in the maintenance, processing and statistical or geostatistical analysis of large environmental databases. This activity and the related stakes will be presented and illustrated with examples from the air quality field in which INERIS, as member of the Central Laboratory for Air Quality Monitoring (www.lcsqa.org) and major partner of the national PREV'AIR system (www.prevair.org), has been working for long time.

Key words:
environmental data, monitoring, modelling, data analysis, decision support

9 Laure Malherbe, Chronic Risks Division, National Institute for Industrial Environment and Risks (public research body under the supervision of the French Ministry in charge of the Ecology and Sustainable Development); email: laure.malherbe@ineris.fr

Environmental data production and exploitation at Arpa E-R, the Regional Environmental Protection Agency of Emilia-Romagna, Italy

Stefano Tibaldi[10]

Abstract
Arpa Emilia-Romagna is the regional Environmental protection Agency of the Regional Government of Emilia-Romagna. Its institutional duties will be reviewed, with particular attention to the management of the regional environmental information system and to the tasks of environmental data collection and processing, with the aim of producing indices and indicators of the state of the environment and its quality. The problem of making such data available as open data will be mentioned, also in the context of the development of the Italian Public Administration legislation.

Key words:
environment, environmental data, indices, indicators, open data

[10] Prof. Stefano Tibaldi, Director General Agenzia Regionale Prevenzione e Ambiente dell'Emilia- Romagna, Via Po 5, 40139 Bologna, Italy; email: stibaldi@arpa.emr.it www.arpa.emr.it

Multivariate hidden Markov models as tools for marine atlas development

Francesco Lagona[11]

Abstract
Marine atlases display the typical environmental regimes of marine or coastal areas, by clustering multivariate environmental data into a finite number of classes. Environmental clustering is however often complicated by multiple correlations, across time, space and between variables, and by the different supports on which the variables are observed, the real line and the circle. Multivariate hidden Markov models integrate these features of the data into a single framework and, simultaneously, allow for efficient data clustering. This paper describes the use of these models in the development of a new atlas of the Adriatic Sea, by discussing two case studies.

Key words:
Adriatic sea, circular data, Hidden Markov model, Gaussian Markov field, von Mises Markov field, sea surface, skewness

[11] Francesco Lagona, University of Roma Tre, Via G. Chiabrera 199, 00145 Rome, Italy; e-mail: francesco.lagona@uniroma3.it

Alternative model specifications for regression estimated on big spatial datasets

Giuseppe Arbia[12]

Abstract
The last decades have witnessed a formidable explosion of geo-referenced data collection and diffusion in all areas of human society. Geo-referenced data cannot be thought of as being a random sample, rather they tend to be spatially correlated with disturbing effects on inference. When referring to big datasets, most of the spatial statistical techniques are computationally prohibitive to implement even with powerful computer machines. To solve the problem many approximate solutions have been suggested in the literature (Smirnov and Anselin, 2001; Griffith, 2000, 2004) and alternative models have been developed. In this paper we will review some of these alternative models (LeSage and Pace, 2007; Arbia et al., 2013; Arbia, 2014).

Key words:
big spatial data; spatial correlation; spatial regression models

References

1. Arbia, G., Bee, M. and Espa, G. (2013) Testing isotropy in spatial econometric models, to appear on Spatial Economic Analysis.
2. Arbia, G. (2014) Pairwise likelihood inference for spatial regressions estimated on very large datasets,Spatial Statistics, 7, –39,
3. Griffith, D. A. (2000) properties and approximations of selected incidence matrices employed in spatial analysis, Linear algebra and its applications, , 95-112
4. Griffith, D. A. (2004) eigenfunctions of adjacency matrices for planar graphs employed in spatial analysis, Linear algebra and its applications, 388, 201-219.
5. Lesage J. and Pace, K. (2007) A matrix exponential spatial specification, Journal of Econometrics, 140, 1, 190-214.
6. Smirnov O. and Anselin, L. (2001) maximum mlikelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach, Computational Statistics and Data Analysis, 35, 301-319.

[12] Giuseppe Arbia, Università Cattolica del Sacro Cuore, Roma, Italy; email: giuseppe.arbia@rm.unicatt.it

Benchmarking methods for space and time unit level EBLUP estimators

Michele D'Alò, Lorenzo Di Biagio, Stefano Falorsi, Andrea Fasulo and Fabrizio Solari[13]

Abstract
In this paper benchmark strategies for SAE estimates are compared for the case of space and time unit level EBLUP estimation. In particular the procedure proposed in Di Fonzo and Marini (2011) has been developed to obtain coherent estimates for both temporal and contemporaneous constraints. Finally two empirical studies are presented. The first study is based on synthetic data generation finalized to show the computational performances of the method.
The second one considers the Italian Labour Force Survey (LFS) data.

Key words:
Time series, indirect estimator, small area estimation, serial autocorrelation

References

1. Di Fonzo, T., and Marini, M. (2011). Simultaneous and two-step reconciliation of systems of time series: methodological and practical issues, Journal of the Royal Statistical Society, Applied Statistics, 60(2), pp.143–164.
2. D'Alo', M., Di Consiglio, L., Falorsi, S., Ranalli, M.G., and Solari, F. (2012). Use of Spatial Information in Small Area Models for Unemployment Rate Estimation at Sub-Provincial Areas in Italy, Journal of the Indian Society of Agricultural Statistics 66(1), pp. 43–53.
3. D'Alo, M., Falorsi, S., Solari, F. (2014). Generalized and efficient formulation of space and time unit level EBLUP for repeated surveys, Journal of Survey Statistics and Methodology, under reference.

[13] Michele D'Alo' · Lorenzo Di Biagio · Stefano Falorsi · Andrea Fasulo · Fabrizio Solari, Istat, Via Cesare Balbo 16, e-mail: dalo@istat.it, lorenzo.dibiagio@istat.it, stfalors@istat.it, fasulo@istat.it, solari@istat.it

Use of geo-referenced information for statistical purposes

Elisabetta Carfagna, Simone Maffei and Andrea Carfagna[14]

Abstract

Developed and developing countries need accurate and timely agricultural, rural and agro-environmental statistics for policymaking; in several cases these are requested also at small administrative units level, that means small domains. At the same time, the statistical offices and the Ministries of Agriculture face scarce and often declining resources. Thus, there is a strong need to increase the efficiency of data collection methods and estimators. One way for reaching this result is using auxiliary information and developing more efficient statistical methods to be adopted at the design level as well at the estimator level. The cost of collection, processing and analysis of geo-referenced information is decreasing and its quality is improving.

In this paper, we discuss the implications of this use, in terms of efficiency, robustness, characteristics of sampling units and non-sampling errors generated by data aggregation, disaggregation, and integration.

Key words:

Agricultural statistics, cost-effectiveness, technological and methodological development.

[14] Elisabetta Carfagna, University of Bologna; email: elisabetta.carfagna@unibo.it
Simone Maffei, TERRA NOVA GIS; email: simone.maffei@gmail.com
Andrea Carfagna, Independent consultant; email: andreacarfagna@virgilio.it

Toward a data-and-policy spatially integrated system in the local context: From evidence-informed to policy-motivated statistical knowledge

Wlodzimierz Okrasa[15]

Abstract

Since systematic policy research that plays vital role in the evidence-based policy-making (EBPM) scheme – embracing a sequence from generating data and information to research knowledge and knowledge utilization - has essentially been missing from public information system during the pre-enlargement era in many of the CEE countries (including Poland), closing the gap (between evidence and policy) became one of the key objectives of the steadily modernized official statistics (in Poland). As the gap seemed in particular to be acutely felt among the policy practitioners and analysts working with sectoral (esp. social services, infrastructure, labor market, etc.) and overall development issues at the local level – including EU-supported interventions - much of the efforts has been made to establish a base for policy-relevant indicators at the NUTS5/LAU2 units (gmina in Poland).

Adopting the user's (policy analyst's) point of view, an exemplification of the EBPM chain in the local context is being given by showing how the multidimensional index of local deprivation (MILD) – possible to be constructed on the basis of the local level data sets – works for both objective type evaluation criteria (e.g. 'efficiency' of intervention) and for normative criteria of 'spatial justice'(including geographic targeting of social and development programs).

Thanks to the availability of geo-referenced data, the spatial distribution and inequalities of two types of indices – the MILD (for gminas) and the measures of subjective well-being (for their residents, in the selected regions of the country) – are presented to demonstrate benefits of the data-and-policy spatially integrated system (Okrasa, 2012), as an approach complementary to multilevel modeling (Subramanian, 2010).

This paper aims to illustrate a new type of approach to the EBPM process through adopting a perspective promoting policy-motivated statistical knowledge for so-called 'intelligent policy making' (e.g. Sanderson, 2011).

[15] Wlodzimierz Okrasa, Central Statistical Office of Poland, Warsaw, Poland; email: wlodek.okrasa@wanadoo.fr

Financial Statistics for Europe 2020

Aurel Schubert[16]

Abstract
A vision on the main features of financial statistics that could effectively contribute to the decision making processes of monetary policy and banking supervision by 2020. Key concepts are the reuse of existing information available for central banks, overcoming the silo approach by combining existing as well as new sources, with the help unique identifiers into an integrated enterprise-wide data management system. The role of a Statistical Data Dictionary, a European Reporting Framework, the Centralized Securities Database (CSDB), the planned Analytical Credit Dataset (AnaCredit), as well as the Legal Entity Identifier (LEI) are briefly outlined.

[16] Aurel Schubert, Director General Statistics, European Central Bank; email: aurel.schubert@ecb.europa.eu

How to measure the quality of Financial tweets?

Paola Cerchiello and Paolo Giudici[17]

Abstract
Systemic risk modelling concerns the estimation of the interrelationships between financial institutions, with the aim of establishing which of them are more central and, therefore, more contagious/subject to contagion. The aim of this paper is to develop a novel systemic risk model. A model that, differently from existing ones, employs not only the information contained in financial market prices, but also big data coming from financial tweets. From a methodological viewpoint, the novelty of our paper is the estimation of systemic risk models using two different data sources: financial markets and financial tweets, and a proposal to combine them, using a Bayesian approach. From an applied viewpoint, we present the first systemic risk model based on big data, and show that such a model can shed further light on the interrelationships between financial institutions.

Key words:
Twitter data analysis, Graphical Gaussian models, Graphical Model selection, Banking and Finance applications, Risk Management

References

1. Billio M., Getmansky M., Lo, A. and Pelizzon, L. : Econometric measures of connectedness and systemic risk in the finance and insurance sector. Journal of Financial Economics, 104(3), 535–559 (2012)
2. D'Alo', Cerchiello P., Giudici P. : On a statistical h index. Scientometrics, 99, 299–312 (2014).
3. Lauritzen, S. L. (1996). Graphical models. Oxford University Press. (1996).

[17] Paola Cerchiello and Paolo Giudici, University of Pavia, Italy; email: giudici@unipv.it, paola.cerchiello@unipv.it

Improving business statistics for small domains

Maria Rosaria Ferrante and Silvia Pacei[18]

Abstract

The improvement of small domains business statistics is becoming increasingly important in order to improve official statistics, to better monitor enterprise performance and to better plan policies promoting entrepreneurship at the local level. Using data on the Small and Medium Enterprises sample survey conducted by the Italian National Statistical Institute (ISTAT), we estimate two parameters, value added and labour cost, with reference to small domains defined by cross-classifying regions, industries and firm size. The survey considered is designed to provide reliable estimates for domains larger than those we are interested in, hence the number of firms sampled from many of our domains is too low to obtain reliable estimates using the "direct" estimation strategy currently employed by Istat. To improve estimates reliability we propose to use model-based small area estimators. One of the most relevant statistical issues which arise in business surveys is the skewness in the distribution of outcomes, due to the majority of small firms. To take this issue into account, we relax the normality assumption of the classic normal-normal small area model by using a model based on the skew-normal distribution. In addition we propose a bivariate extension of such skew-normal models, which enables us to take into account the high correlation between the target variables. Results highlight i) the importance of considering the asymmetry of data and the correlation between outcomes, ii) the relevance of obtained estimates for the regional economic studies.

[18] Maria Rosaria Ferrante, Department of Statistical Sciences, University of Bologna, Italy; email: maria.ferrante@unibo.it
Silvia Pacei, Department of Statistics, University of Bologna, Italy; email:silvia.pacei@unibo.it

International Trade and Investment: improving our knowledge about competitive advantages

Fabrizio Onida[19]

Abstract
The improving comparability of European statistics on industry and trade, as well as the growing availability of firms' microdata, allow us to foster our understanding of globalization patterns and (even more important from policy sides) impact on countries' and regions' growth of incomes and human capital. Here follows a short list of additional and/or improved statistical evidence that could greatly benefit us as occasional or systemic users. Technological content of products in international trade: integrating standard classifications into "high-medium high-medium low-low" categories based on R&D intensity (expenditure, patents) in developed countries with additional evidence on informal innovative content and product-process complexity, drawn from sample survey data (e.g. European CIS) in a number of more or less developed countries. Trade and labour skills: we should have access to labour statistics (employment, wages) with finer disaggregation of labour skills beyond the official traditional categories of white and blue collars (dirigenti-quadri-impiegati-operai) or of employment by educational degree. Data available to agencies of social security (INPS in Italy), that cover more disaggregated skill levels within general categories of white and blue collars (e.g. more or less specialized worker–technical staff, higher-lower class of white collars) are still largely unavailable to external users for computing the human capital content of exports and import substitutes. Knowing more about the human capital content of trade is crucial for a better evaluation of the "quality" of a country's specialization. Trade and firms' size: use of transition matrices of firms by size class over time (microdata) in order to assess the country's propensity to achieve economies of scale, given a highly fragmented industrial structure (analogy to transition matrix over various internationalization forms in 2007-2010, in Istat, Rapporto sulla competitività dei settori produttivi, 2013, par. 2.2). Firms' size and competitive strategies: survey data on heterogeneous reactions to economic crisis and to external price and non-price pressure (e.g. Istat Rapporto sulla competitività dei settori produttivi 2013 and 2014, par. 3.2). Repositioning along international value added chains, upgrading product quality-technological content, increasing product varieties, changing export markets mix, cutting prices and export profits, investing in foreign distribution network, insourcing-outsourcing-offshoring (see following point) etc. Export statistics (micro data): identifying the very large share of pure traders (small export-import companies) from domestic manufacturing producing exporters. Inward-Outward FATS: linking evidence on FDI to microdata on trade and employment, so as to achieve better assessment of FDI impact on trade balance and on demand of labour skills. Comparison of productivity and profitability of foreign affiliates vs. domestic plants (beware of fiscal optimization across host countries with different corporate tax rates). Surveys on FDI determinants (market seeking, cost-saving, resource and knowledge seeking) and patterns of internationalization strategies (outsourcing, offshoring, insourcing, equity vs. non equity investment, contractual agreements). Survey data on intra-firm/intra-group trade (WTO estimates one third of global trade?)

[19] Fabrizio Onida, University of Milan- Bocconi, Italy: email: fabrizio.onida@unibocconi.it

Big Data and Visual Analytics for Health Care Systems Performance Assessment

Orestis Tsigkas and Michail Skaliotis[21]

Abstract
This paper describes a novel approach towards assessing the (cost-) efficiency of health care systems across European countries based on big data and visual analytics. The proposed solution goes beyond the current approach of countries' benchmarking (often based on composite indicators) by allowing the policy analyst to look into the difference dimensions (pieces of evidence) and interactively investigate the often multi-disciplinary dynamics of the phenomenon under study. The ultimate goal is to enrich the policy analyst's arsenal with big data analytics tools and visual analysis techniques that will extend his understanding of the evolving and raising phenomena beyond his/her normal human perception. Such an approach can significantly contribute in the realization of evidence-based policy-making practices, avoiding the pitfall of relying on an incomplete or inappropriate set of indicators (pieces of evidence). Although the proposed solution is not meant to serve as an early-warning system, certain aspects of its implementation could facilitate its realization.

Key words:
big data, visual analytics, health care, evidence-based policy making

[21] Orestis Tsigkas, European Commission, Eurostat, 5 rue Alphonse Weicker, 2721 Luxembourg; email:orestis.tsigkas@ec.europe.eu
Michail Skaliotis, European Commission, Eurostat, 5 rue Alphonse Weicker, 2721 Luxembourg; email: michail.skaliotis@ec.europe.eu

Data-Visualisation: From Tool to Mother Tongue

Michael Neutze and Sibylle von Oppeln-Bronikowski[22]

Abstract
Mapping census data on housing and dwellings in Germany as an example, current efforts in communicating statistics and improving statistical literacy are discussed. The influence of tools is emphasized.

Key words:
data-visualisation, statistical literacy, census, mapping, Germany

[22] Michael Neutze, Federal Statistical Office of Germany; email: michael.neutze@destatis.de
Sibylle von Oppeln-Bronikowski, Head of Department: Strategy and Planning, International Relations, Research and Communication, Federal Statistical Office of Germany; email: sibylle.oppeln@destatis.de
Pina Grazia Ticca, ISTAT, Italy; email: ticca@istat.it

Web Services for spatial-temporal comparison of statistical indicators

Pina Grazia Ticca[23]

Abstract
The paper describes a new way for comparing geospatial statistical indicators using interactive tools based on spatial analysis approach; that has been implemented through geowebservices accessed by GIS web applications.
The prototype designed let the web users interact with statistical thematic maps so that spatial and temporal relations can be easily viewed and analyzed.
In the prototype implemented the statistical geospatial database used is composed of a subset of population census indicators linked to admin areas.

Key words:
GIS, geospatial analysis, statistical geowebservices.

[23] Pina Grazia Ticca, ISTAT, Italy; email: ticca@istat.it

Preparation of reproducible statistical reports

Andrej Blejec[24]

Abstract
Preparation of various reports for communication of statistical information is a very common task. Reports are the final product of the sequence of steps consisting problem definition, data gathering and analysis, table and figures organization and, finally typesetting of the report itself. In the usual framework, the steps are not connected and traces to origin of parts is lost. If some data change, become more complete or revised, it is necessary to prepare a new version of the report which tends to be tedious.
In some fields of research, tools and systems to support preparation of reproducible statistical reports are used. Reproducible report consists of documents, which are a combination of text, meant for human readers, and computer code that is executed by computers. This enables easy reproduction and traceability of analyses, tables, and figures. We will present reproducible reporting based on Sweave and knitr, which combines R and LaTeX (or markdown).

Key words:
Reproducible report, Sweave, knitr, markdown, R

[24] Andrej Blejec, National Institute of Biology, Vecna pot 111, 1000 Ljubljana, Slovenia; email: andrej.blejec@nib.si

A Methodological and Computational Framework for Statistical Disclosure Control

Matthias Templ, Alexander Kowarik and Bernhard Meind[25]

Abstract

Data privacy/confidentiality is one of the core businesses in official statistics. SDC becomes more and more important in the last years because of an increase of the awareness on data privacy and because of the fact that more and more data are provided to the public or to researchers.

Before providing microdata to the public or to researchers confidentiality/privacy has to be respected. Only data with low disclosure risk can be disseminated. Confidentiality can be achieved and disclosure risk can be estimated with statistical disclosure control (SDC) methods. This contribution will give an outline of problems and possibilities in the area of statistical disclosure control on microdata.

Existing anaonymisation and disclosure risk methods are briefly reviewed and summarized and modern efficient tools are presented. The discussed methods are then applied on popular large real-world data set. The application of few selected anonymisation methods leads to well-protected anonymised data with high data utility and low information loss. Moreover, the first time it is possible to also anonymise large data sets with millions of observations with efficient ready-to-use user-friendly software tools.

Key words:

statistical disclosure control, data utility, disclosure risk, R

[25] Matthias Templ, Austrian Statistical Society & Statistics Austria & Vienna University of Technology & Palacky University Olomouc & data-analysis OG, Austria; email: matthias.templ@gmail.com

The Economic e-Census in Tuscany: effects of the data collection strategy on the response rates

Luca Faustini, Sabina Giampaolo and Alessandro Valentini[26]

Abstract
During the 2010-2011 Census wave (Agricultural Census, Population and Housing Census and Economic Census), the Italian National Institute of Statistics (ISTAT) has adopted web survey techniques pushing the transition from the traditional census (wave 2000-2001) to the e-Census, with the general purpose to improve data quality and to reduce costs. This paper discusses in detail – using provisional data – the Italian experience of the Economic Census referred to the Enterprises Census (ENT) and Non-profit Institutions census (NP) in Tuscany, focusing the attention on the main characteristics of the data collection procedures, and on the impact of the introduction of web tools. In particular, results highlight the main dimensions affecting the web and paper response and advantages of the use of web.

Key words:
economic censuses, web data collection, frame list, fieldwork monitoring, response rates.

26 Luca Faustini, ISTAT; email: faustini@istat.itfaustini@istat.it
Sabina Giampaolo, ISTAT; email: giampaol@istat.itgiampaol@istat.it
Alessandro Valentini, ISTAT; email: alvalent@istat.italvalent@istat.it

Experiments on internet and mixed-mode data collection at Insee

Olivier Sautory, Gaël de Peretti and Tiaray Razafindranovona[27]

Abstract
To face a request of surveys always more demanding in quality terms in a general context of budgetary restriction, the use of internet as privileged or complementary mode of data collection is envisaged at Insee (French national statistical institute). Before generalizing the use of internet to collect data on households, probably in a mixed modes way, some issues need a bit of investigation: coverage, selection and measurement effects relevant to this mode... So Insee has decided to launch a series of experimental CAWI surveys. The standard protocol of the experiments is a sequential mixed modes data collection: first web, then paper questionnaires. The paper outlines the main principles of the experimentation, and gives some examples and results of past, on-going and future experimental surveys.

Key words:
household and individual surveys, internet data collection, mode effects

[27] Olivier Sautory, department of statistical methods, Insee, France; email: olivier.sautory@insee.fr
Gaël de Peretti, department of statistical methods, Insee, France; email: gael.de-peretti@insee.fr
Tiaray Razafindranovona, department of statistical methods, Insee, France; email: tiaray.razafindranovona@insee.fr

Istatel Project: computerization of the data collection process on population statistics

Roberto Bartoli, Carla Camilloni, Cristina Cesaroni and Carla Rubbo[28]

Abstract
The Italian National Institute of Statistics conducts statistical surveys concerning demography, the sources of which are to be found in the administrative archives of over 8,000 municipalities throughout the country. In 1997, the Italian National Institute of Statistics introduced the Istatel Project to better fulfil its institutional commitment, ensure quality assurance and disseminate timely and detailed statistical information. The results this project has achieved were an important requirement to enable the Regulation on European Demographic Statistics 1260/2013 to come into force, which also provides a transition to demographic accounting on an individual basis.

Key words:
istatel, population statistics, administrative archives

[28] Roberto Bartoli, Italian National Institute of Statistics (ISTAT), Integration, Quality, Research and Production Networks Development Department (DIQR) Development and Coordination of Territorial Network and National Statistical System Directorate (DCSR) Territorial Offices for Tuscany and Umbria (TEU); email: bartoli@istat.it
Carla Camilloni, ISTAT, Social and Environmental Statistics Department (DISA), Socio-demographic and Environmental Statistics Directorate (DCSA), Office of structure and dynamic migration (DEM) camillon@istat.it
Cristina Cesaroni, ISTAT/DIQR/DCSR/TEU; email: cesaroni@istat.it
Carla Rubbo, ISTAT/DIQR/DCSR/TEU; email: rubbo@istat.it

User-focused threat identification for anonymised microdata

Hans-Peter Hafner, Felix Ritchie and Rainer Lenz[29]

Abstract
When producing anonymised microdata for research, national statistics institutes (NSIs) identify a number of 'risk scenarios' of how intruders might seek to attack a confidential dataset. This paper argues that the strategy used to identify confidentiality protection measures can be seriously misguided, mainly since scenarios focus on data protection without sufficient reference to other aspects of data. This paper brings together a number of findings to see how the above problem can be addressed in a practical context. Using as an example the creation of a scientific use file, the paper demonstrates that an alternative perspective can have dramatically different outcomes.

Key words:
data protection measures, microdata anonymisation, big data

29 Hans-Peter Hafner, Saarland State University of Applied Sciences, Faculty of Engineering, 66117 Saarbruecken, Germany; email: hans-peter.hafner@htwsaar.de
Felix Ritchie, University of the West of England, Bristol, BS1; email: felix.ritchie@uwe.ac.uk
Rainer Lenz, Technical University of Dortmund, Statistics Faculty, 44221 Dortmund, Germany; email: lenz@statistik.tu-dortmund.de

Are the Central and East European Countries attracting the appropriate type of foreign direct investments?

Vasile Alecsandru Strat[30]

Abstract
When Measuring and studying the relationship between the inflows of FDIs received by a country and the openness of its market might shed some light on the typology of foreign investments attracted by that country. The FDIs can be clustered, as the literature states, in a dichotomous fashion, in vertical FDIs, searching for efficiency and horizontal FDIs mainly interested in acquiring new markets. Thus, it is of crucial importance for an economy to be able to design policies that will encourage those foreign investments which bring the most important benefits for the host country. In this research paper, using the T-Y procedure, we try to identify some aspects useful in identifying the general typology of the FDIs attracted by a sample of five central and east European countries (members of the EU).

Key words:
Foreign Direct Investment, Market openness index, Imports and Exports, Granger Causality analysis, Toda-Yamamoto procedure

[30] Vasile Alecsandru Strat, Department of Statistics and Econometrics, Bucharest University of Economic Studies Romania; email: strat_vasile@yahoo.com

Revised definitions for statistical units - methodology, application and user needs

Roland Sturm[31]

Abstract
Currently modified definitions for basic statistical units are being proposed. Starting from the main conceptual issues and bearing in mind the different functions of units the users are addressed. Statisticians need a clear picture of the user needs from which to delineate to appropriate choice of statistical units and their optimised application.

[31] Roland Sturm, Federal Statistical Office Germany Destatis; email: roland.sturm@destatis.de

Displaying empirical distributions with conditional quantile estimators: an application of symbolic data analysis to the cost allocation problem in agriculture

Dominique Desbois[32]

Abstract
This paper introduces some symbolic data analysis tools in order to display and analyze the conditional quantile estimates, with an application to the cost allocation problem in agriculture. After recalling the conceptual framework of the estimation of agricultural production costs, the first part presents the empirical data model, the quantile regression approach and the symbolic data analysis techniques used. The second part presents the comparative analysis of the econometric results for wheat between member states, using principal component analysis and hierarchic clustering of estimates and range of estimation intervals, discussing the relevance of the displays obtained for inter-country comparisons based on specific productivity.

Key words:
agricultural production, cost allocation, micro-economics, quantile regression, symbolic data analysis, confidence intervals.

[32] Dominique Desbois, Joint Research Unit Economie publique, INRA-AgroParistech; email: desbois@agroparistech.fr

The boundary of public sector in National Accounts versus IPSAS

Giovanna Dabbicco[33]

Abstract
The ongoing debate about the necessity of harmonized accruals-based public accounting standards and the possible implementation of an integrated reporting covering public accounts and government finance statistics (GFS) reporting, have widened the potential scope for comparative research on consolidation practices in Whole of Government Accounts (WGA) and National accounts, notably in the European Union. This development would probably add momentum to broaden the scope of reporting to WGA.
The paper analyses in depth the conceptual frameworks behind financial reporting and national accounts,to better understand the differences between the definition of public sector and its boundary in national accounts as compared with financial reporting.
This would form a useful input to the overall research agenda WGA.

Key words:
National Accounts/ GFS, IPSAS, General Government, Public Sector, WGA.

[33] Giovanna Dabbicco, ISTAT- Istituto Nazionale di statistica, University of Naples Parthenope; email: dabbicco@istat.it

Morbidity Statistics in the EU – key results from pilot studies in sixteen member states

Monica Pace, Hartmut Buchow[34]

Abstract
This article focuses on pilot studies on diagnosis-specific morbidity statistics conducted by sixteen EU Member States from 2005 to 2011 in view of establishing European Union (EU) morbidity statistics. The Eurostat 'Morbidity Task Force' analysed the results of the pilot studies and formulated a set of recommendations on the feasibility of a regular EU morbidity statistics data collection, based on a multi-source approach and best national estimates on incidence and prevalence for a selected list of diseases. This paper presents some examples of the main findings, problematic aspects and proposed solutions that the Task Force reported on potential sources and best estimates with regard to accessibility, usefulness, overall quality and comparability.

Key words:
Morbidity Statistics, Administrative data, Incidence of diseases, Prevalence of diseases, Data linkage

[34] Monica Pace, Hartmut Buchowemail: email: mopace@istat.it
Hartmut Buchow, European Commission – EUROSTAT, Luxemburg ; email: Hartmut.Buchow@ec.europa.eu

Developing comparability in Eurofound surveys: focus on input harmonisation and translation

Tadas Leoncikas[35]

Abstract

Eurofound carries out three European surveys: EWCS, EQLS, and ECS. By default, they cover all the EU Member States (as well as some enlargement countries and other countries) and are set to provide comparable information.

The design of each survey has specific features to cover their specific target population, yet all of them have a stringent procedure for translation of survey questions that involves both national and minority languages. Over the several survey rounds, the translation procedure has evolved into a model based on work of independent translators, use of both survey and linguistic expertise, and an expert-led adjudication process.

Improving the comparability of questions via translation process helps to narrow down a circle of other factors (e.g. related to sampling) affecting comparability and to subsequently address them.

Since some items in Eurofound surveys are same or similar to items existing in or introduced in surveys of the European Statistical System (ESS) or their ad hoc modules, the translation outcomes from Eurofound surveys (available in all the EU languages and beyond) could benefit future work on comparability improvements in ESS. Selected examples of comparing survey items from Eurofound and ESS surveys to reflect on comparability and country differences will be presented.

[35] Tadas, Leoncikas, Eurofound, Ireland; email:tle@eurofound.europa.eu

Choosing the Number of Categories in Agree-Disagree Scales

Melanie Revilla[36]

Abstract
Although a large body of research has documented the bias that results from acquiescence in responses to Agree-disagree (AD) scales and the higher measurement errors that they generate (Saris et al., 2010), still many researchers use them. Therefore, we investigate how many answer categories lead to a higher quality for these scales. Using multitrait-multimethod experiments implemented in the European Social Survey, we found that 5 answer categories lead to a higher quality than 7 or 11. Nevertheless, the quality estimates vary across countries and the differences between the quality estimates of scales with different numbers of response categories too. This means that, even if the scale with the highest overall quality is used, standardized relationships cannot be compared across countries without taking measurement errors into account.

Key words:
quality, MTMM, cross-cultural surveys, Agree-disagree scales, number of response categories

References

1. Saris, W.E., Revilla, M., Krosnick, J.A., and Shaeffer, E.M. (2010) "Comparing Questions with Agree/Disagree Response Options to Questions with Construct-Specific Response Options" Survey Research Methods, 4(1): 61-79.

[36] Melanie Revilla, RECSM, Universitat Pompeu Fabra; email: melanie.revilla@upf.edu

Seeing the Hidden Part of the Iceberg
Gauging the Real Dimension of International Migration

Adrian Otoiu and Emilia Titan[37]

Abstract
The reliability and comparability of international migration statistics are among the of the most important statistical issues, due to the importance of the correct dimensioning of the migration flows and stocks, and the consequences on the design of timely and effective policy measures to manage this phenomenon and contain its possible adverse outcomes. This paper makes an assessment of the migration statistics provided by Eurostat, reveals the most prominent discrepancies between stock and flow data, makes a summary of the main issues that affect the quality and completeness of the migration data, and identifies some solutions in order to improve the comparability, reliability and completeness. There is no one-size-fits-all solution, but an eclectic mix which extends the use of administrative and private data, matches the data coming from distinct sources in order to assess the full dimension of international migration, harmonizes the way in which data is compiled and reported by different countries, matches observed flows with (demographic) stock-based estimates, provides consistent estimates of the bilateral migration flows between countries, and improves the measurement of temporary and illegal/undeclared migration. Another recommendation is to preserve the existing migration data along with the improved estimates in order to gauge the specifics of international migration in the reporting countries and to properly inform national public policy measures.

Key words:
official statistics methodology, international migration, migration measures, data harmonisation, data dissemination

[37] Adrian Otoiu, Bucharest Academy of Economic Studies, Bulgaria; email: otoiu_adrian@yahoo.com
Emilia Titan, Bucharest Academy of Economic Studies, Bulgaria; email: emilia_titan@yahoo.com

The absolute measure of poverty: a key indicator to design policies

Cristina Freguja and Nicoletta Pannuzi[38]

Abstract
European Union has made extraordinary progress in the availability of indicators on poverty and social exclusion, improved together with the evolution of statistics, data collection and policy needs. Italy has implemented a measure of absolute poverty to identify and quantify the poorest among the poor people; to properly orient policies; to improve and complete the existing information. It is based on a minimum basket of goods and services (basic needs), it takes into account costs of living and different dynamics of prices during time. The unavailability of such an indicator means losing information on the population most at risk of economic distress (excluding extreme poverty), sometimes hidden by traditional indicators; the comparison between different measures confirms the heuristic value of this approach.

Key words:
absolute poverty approach, permanent poverty, deprivation measures.

References
1. Nolan, B. e Whelan, C.T., Using Non-Monetary Deprivation Indicators to Analyse Poverty and Social Exclusion in Rich Countries: Lessons from Europe?. In: Journal of Policy Analysis and Management, vol. 29, n. 2, pp. 305–325 (2009)
2. Istat, La misura della povertà assoluta, Metodi e Norme n. 39 (2009)
3. Istat, Reddito e condizioni di vita, anno 2012, Statistiche report (2013)
4. Istat, La povertà in Italia, anno 2013, Statistiche report (2014)

[38] Cristina Freguja, Italian National Institute of Statistics, Italy; email: freguja@istat.it
Nicoletta Pannuzi, Italian National Institute of Statistics, Italy; email: pannuzi@istat.it

The Italian Active inclusion System

Raffaele Tangorra[39]

Abstract

[39] Raffaele Tangorra, Ministry of Labour and Social Affairs, Italy

Behind and beyond the (headcount) employment rate

Andrea Brandolini and Eliana Viviano[40]

Abstract

This paper argues that we need more general statistical indices for the analysis of the European labour markets. First, the paper discusses some normative aspects implicit in the current definition of the employment rate, which is a fundamental policy target in the new strategy Europe 2020. Second, it proposes a class of generalised indices based on the intensive margin of labour supplied, as approximated by the total annual hours of work relative to a benchmark value, for both individuals and households. These indices provide a more nuanced picture of the European labour markets, which better reflects the diversity in the use of part-time and fixed-term jobs as well as other factors affecting the allocation of work across and within households.

[40] Andrea Brandolini , Bank of Italy, email: eliana.viviano@bancaditalia.it; Eliana Viviano, Bank of Italy, email: andrea.brandolini@bancaditalia.it

Measuring poverty and living conditions at small area level

Monica Pratesi and Alessandra Petrucci[41]

Abstract

Small area estimation (SAE) plays an important role in survey sampling due to growing demands for reliable small area statistics from both public and private sectors. This paper reviews some of the current techniques of small area estimation combined with spatial models available in the literature and focusing on the quality of the obtained results. Illustrative examples or applications are likewise presented in the context of Official Statistics where data sources, particularly surveys and censuses or surveys, administrative sources and also Big Data sources, has been combined using statistical models based on small area and domain estimation.

[41] Monica Pratesi, Department of Economics and Management, University of Pisa, Pisa, Italy; email: monica.pratesi@unipi.it
Alessandra Petrucci, Department of Statistics, Informatics, Applications, University of Florence, Florence Italy; email: alessandra.petrucci@unifi.it
Andy Teague, Office for National Statistics, UK, email:andy.teague@ons.gsi.gov.uk

Research into using administrative data and surveys to produce census and population statistics – ONS Beyond 2011 Programme, England and Wales

Andy Teague[42]

Abstract

The abstract In 2010, The UK Statistics Authority asked the Office for National Statistics (ONS) to review the options for the future provision of population statistics and the next census in England and Wales after the 2011 Census was concluded. ONS launched the Beyond 2011 Programme in 2011. It examined and compared various approaches to counting the population. A key focus of the research over the past three years has been on the use of anonymously-linked record-level administrative data and surveys to produce statistics about the size and characteristics of the population. In Summer 2013 and Spring 2014, a range of research papers were published that reported the progress that had been made in this area. In March 2014, the National Statistician recommended to the Board of the UK Statistics Authority that the future provision of population statistics and the next census should comprise:

1. An online census of all households and communal establishments in England and Wales in 2021 as a modern successor to the traditional, paper-based decennial census
2. Increased use of administrative data and surveys in order to enhance the statistics from the 2021 Census and improve annual statistics between censuses.

This recommendation was subsequently endorsed by the UK Statistics Authority and the Government. The presentation will summarise the research, the challenges of using administrative data and outline the ongoing programme of research to deliver the above recommendation.

[42] Andy Teague, Office for National Statistics, UK, email:andy.teague@ons.gsi.gov.uk

The Dutch Census 2011

Eric Schulte Nordholt[43]

Abstract
The Dutch 2011 Census tables were produced by combining existing register and sample survey data. Since the last census based on a complete enumeration was held in 1971, the willingness of the population to participate has fallen sharply. Statistics Netherlands no longer uses census questionnaires and has found an alternative in the register-based census, using only existing data. The register-based census is cheaper and more socially acceptable. The table results of the Netherlands are not only comparable with earlier Dutch censuses, but also with those of the other countries in the 2011 European Census Round.

[43] Eric Schulte Nordholt, CBS, NL; email: e.schultenordholt@cbs.nl

Statistical Matching in complex survey data: beyond the conditional independence assumption

Pier Luigi Conti and Daniela Marella[44]

Abstract

The goal of statistical matching is the estimation of the underlying joint distribution of variables separately available from independent sample surveys. The lack of joint observations of the variables of interest leads to uncertainty about the joint population distribution function. The aim of this paper is to analyze the statistical matching problem for complex survey data. More specifically, in the first place the class of all plausible joint distributions for the variables not jointly observed compatible with the available information is defined. Any distribution in such a class can be used as a "surrogate" of the actual joint distribution function and will be called a matching distribution. A measure to assess how reliable is the estimate of any matching distribution in the class is introduced and its asymptotic normality is proved for complex survey data. In the second place, a reasonable criterion to choose a plausible joint distribution function for the variables not jointly observed via the Iterative Proportional Fitting algorithm is proposed. Finally, an application to a real case is performed.

Key words:

Statistical matching, conditional independence assumption, uncertainty, Fréchet bounds, iterative proportional fitting algorithm.

[44] Pier Luigi Conti, Dipartimento di Scienze Statistiche, Sapienza Università di Roma; email: pierluigi.conti@uniroma1.it
Daniela Marella, Dipartimento di Scienze della Formazione, Università Roma tre; email: daniela.marella@uniroma3.it

Supporting the Statistical Utilization of Administrative Data Sources

Giovanna D'Angiolini, Pierina De Salvo and Andrea Passacantilli[45]

Abstract

The paper discusses the documentation and standardization requirements which are related to the statistical utilization of administrative data sources and illustrates the main features of the Istat's strategy for satisfying such requirements [1], which adopt an ontology-based approach to the information content specification and the data quality assessment. Generally speaking the Istat's strategy is aimed at collecting information about the available administrative data sources and producing standard documentation about their information content and quality, and modifying, when possible, the content of such data sources through adopting standard statistical definitions, classifications and data management conventions. In order to provide the users with proper knowledge about the content and the quality of the administrative data sources Istat is launching several systematic documentation activities, which concern different kinds of administrative data sources. The collected information is managed by means of a dedicated web-based metadata management system, called DARCAP (Documenting Public Administration Archives) in order to disseminate it to any potential statistical user of the documented administrative data sources. Moreover in order to support in-depth quality analyses on the most important administrative data sources we are studying a new Quality Assessment Framework for Administrative Data Sources.

Key words:

Administrative data source, Administrative data documentation, Administrative data quality, Data source ontology, Statistical data production

References

1. D'Angiolini G., De Salvo P., Passacantilli A., Istat's new strategy and tools for enhancing statistical utilization of the available administrative databases, European conference on quality in official statistics, Vienna (2014)

[45] Giovanna D'Angiolini, ISTAT, Italy; email: dangioli@istat.it
Pierina De Salvo, ISTAT, Italy; email: desalvo@istat.it
Andrea Passacantili, ISTAT, Italy; email: passacantilli@istat.it

Implementing the ESS Vision 2020 - which will be the new interfaces between the stakeholders of the European Statistical System

Roberto Barcellan[46]

Abstract
In a digital and information driven society, official statistics become a precious public good aimed at responding to the needs of many different users. To accomplish their mandate, statistical offices strive to adapt to major developments in society and the economy; in other words they try to stay relevant becoming modern.
The modernisation hype for statistical offices started in the aftermath of the crisis and foresees implementation strategies running for most organisations up to 2020. In May 2014, the European Statistical System (ESS) outlined its strategic goals at the horizon 2020 in the ESS Vision 2020. This paper illustrates how the collaborative functioning of the ESS is changing to address the challenges that the ESS Vision 2020 poses to European statistics.

Key words:
official statistics, modernisation, organisational changes

References
1. ESS Vision 2020
http://epp.eurostat.ec.europa.eu/portal/page/portal/pgp_ess/0_DOCS/estat/ESS_Vision 2020.pdf

[46] Roberto Barcellan, Eurostat, email: roberto.barcellan@ec.europa.eu

Modernization @ work: lessons learned

Emanuele Baldacci[47]

Abstract

Istat has undertaken in the last half decade a major effort to innovate the statistical production process and its products with a view to raising efficiency, enhancing data quality, and fostering the ability to respond to emerging user needs.

The strategic vision is consistent with Vision 2020, the medium term transformation program of the European Statistical System. It has its foundation in the role of technology as a driver for innovation and builds on the Enterprise Architecture model based on lean management and agile organisation to respond to change.

The key pillar of the strategy are process industrialisation based on service-oriented architecture and standardisation of statistical tools and IT services for data collection, process and analyse, and dissemination.

In these three areas of the production process, the modernisation strategy focuses on key priorities. First, digital data collection systems (including through web portals) to reduce respondent burden and provide tailor- based information to respondents. Second, extensive use of administrative data and new data sources, such as big data, for Registry-based integrated statistical production. Third, metadata systems as drivers of production industrialisation based on standards and reuse of data and statistical tools. Finally, dissemination systems based on integrated data warehouse for data browsing, accessible micro data for research, linked open data portals and related data analytics services, visual analytics tailored to user needs, and data analytics based on tools for policy decision and policy simulation.

As Istat is implementing the strategy of modernisation, key challenges and lessons have emerged. First, Identifying a Business Architecture model is key for designing process innovation based on agreed corporate principles and ensuring the elimination of domain-specific silos. Second, a sound governance to steer the strategy and change management process are instrumental for implementation based on sound project management frameworks. Third, promoting innovation through agile and horizontal corporate networks and public private partnerships drives corporate energy for change and build innovation champions that are critical facilitators for the strategy, Finally, communicating modernisation to stakeholders is important to demonstrate the benefits of change and build momentum for political support and corporate consensus for the change strategy.

[47] Emanuele Baldacci, Istat, email: baldacci@istat.it

Changing UK Official Statistics System

John Pullinger[48]

Abstract
In common with all most NSIs, the Office for National Statistics (ONS) – the executive arm of the UK Statistics Authority – has been through years of sustained change. As we prepare to implement the European Statistical System's 'Vision 2020' this paper looks at what we've learned from where things have gone wrong, but more importantly how to get it right. The insights provided here focus on our experiences from the UK. We believe the principle of delivering the right projects and programmes, in the right way, with the right people is one which is equally applicable across the European Statistical System and will help us to achieve the Vision together.

Key words:
change management, vision, people

[48] John Pullinger, National Statistician, UK Statistics Authority; email: national.statistician@statistics.gsi.gov.uk

The MIMIC model to predict economic public opinions through macro-economic indicators in Europe

Maurizio Carpita, Enrico Ciavolino and Mariangela Nitti[49]

Abstract
A Multiple Indicators Multiple Causes (MIMIC) model is proposed in order to catch the influence of national macro-economic indicators on the European public opinions about the economic situation, at both national and EU level.
This study is part of the SYRTO (Systemic Risk Tomography, syrtoproject.eu) European Project, which aims at creating an early warning system to identify potential threats to financial stability and to inform policy measures in order to prevent and manage systemic crises in the Eurozone.

Key words:
MIMIC models, measurement models, economic perceptions

[49] Maurizio Carpita, University of Brescia, Italy, email: maurizio.carpita@unibs.it
Enrico Ciavolino and Mariangela Nitti, University of Salento, Italy, email: ciavolino.enrico@gmail.com, m.nitti@hotmail.it

Summarizing Data using Partially Ordered Set Theory: An Application to Fiscal Frameworks in 97 Countries

J. Bachtrögler, H. Badinger, A. Fichet de Clairfontaine and W.H. Reuter[50]

Abstract
The widespread use of composite indices has often been motivated by their practicality to quantify qualitative data in an easy and intuitive way. At the same time, this approach has been challenged due to the subjective and partly ad hoc nature of computation, aggregation and weighting techniques as well as the handling of missing data. Partially ordered set (POSET) theory offers an alternative approach for summarizing qualitative data in terms of quantitative indices, which relies on a computation scheme that fully exploits the available information and does not require the subjective assignment of weights. The present paper makes the case for an increased use of POSET theory and compares POSET indices and composite indices (from previous studies) measuring the "stringency" of fiscal frameworks.

Key words:
Partially Ordered Set Theory, Composite Indices, Index Functions, Fiscal Frameworks, Fiscal Rules

[50] J. Bachtrögler, Vienna University of Economics and Business, Austria; email: julia.bachtroegler@wu.ac.at
H. Badinger, Vienna University of Economics and Business, Austria,;email: harald.badinger@wu.ac.at;
A. Fichet de Clairfontaine, Vienna University of Economics and Business, Austria;
email: aurelien.fichet.de.clairfontaine@wu.ac.at
W.H. Reuter, Vienna University of Economics and Business, Austria,;email: wolf.reuter@wu.ac.at

Interval Based Composite Indicators

Carlo Drago[51]

Abstract
Composite indicators are increasingly important in country comparisons and in policy making. At the same time, the robustness of the results obtained and in particular of the rankings and the conclusions obtained from the analysis it is usually accepted with doubts. In this sense our proposal is to use interval data in order to measure the uncertainty related to the different composite indicators based on the different assumptions used as input. In this sense where composite indicators can be considered as models, for this reason it could be necessary to assess the uncertainties related to the different choices in the construction. The uncertainty can be represented by the interval data. The intervals keep the information related to the initial value of the composite indicator, but at the same time give information on the range of the results.

Key words:
Composite Indicators, Interval Data, Robustness, Sensitivity Analysis,
Uncertainty Analysis

[51] Carlo Drago, Università degli Studi "Niccolo Cusano", Italy; e-mail: c.drago@mclink.it

Non-compensatory Composite Indices for Measuring Changes over Time: A Comparative Study

Matteo Mazziotta and Adriano Pareto[52]

Abstract
this paper, we analyze some non-compensatory composite indices for measuring multi-dimensional phenomena and monitoring their changes over time. In particular, we compare the Adjusted Mazziotta-Pareto Index (AMPI) and theMean-Min Function (MMF). The AMPI is a not full compensatory index based on the use of a penalty related to the unbalance among dimensions. The MMF allows compensability among dimensions with a cost that increases with unbalance and can be seen as an intermediate case between complete compensability and perfect complementarity among dimensions. An application to a set of individual indicators of development in the Italian regions is presented.

Key words:
index, ranking, non-compensability

[52] Matteo Mazziotta and Andrea Pareto, Italian National Institute of Statistics; email: mazziotti@istat.it, pareto@istat.it

Streamlining Data Editing at Statistics Spain

Silvia Rama and David Salgado[53]

Abstract
Among the efforts to streamline and modernise the production process at Statistics Spain, we have focused on the data editing phase. Upon the optimization approach to selective editing developed at this national statistical office we have satisfactorily conducted a pilot experience with real data to test these ideas and slightly generalized the generic EDIMBUS editing and imputation strategy including input editing (during data collection) and expressing the strategy in terms of statistical production functions. This proposal is currently in production for three short-term business statistics, and under implementation for other two. We detail the extended strategy in terms of the number of collected units at each time instant and the number of editing cycles undergone by each questionnaire.

Key words:
data editing, selective editing, optimization approach, statistical production function, editing and imputation strategy

References

1. Arbués, I., Revilla, P., and Salgado, D. An optimization approach to selective editing. J. Off. Stat. 29, 489–510 (2013)
2. de Waal, T., Pannekoek, J. and Scholtus, S. Handbook of statistical data editing and imputation. Wiley, Hoboken (2011)
3. EDIMBUS. Recommended practices for editing and imputation in cross-sectional business surveys. ISTAT, CBS, SFSO and EUROSTAT (2007)
4. López-Ureña, R., Mancebo, M., Rama, S. and Salgado, D. Application of the optimization approach to selective editing in the Spanish Industrial Turnover Index and Industrial New Orders Received Index survey. To appear as INE Spain Working Paper. INE, Madrid (2014)
5. Rama, S. and Salgado, D. Standardising the editing phase at Statistics Spain: a little step beyond EDIMBUS. To appear as INE Spain Working Paper. INE, Madrid (2014)

[53] Silvia Rama, Spanish National Institute; email: silvia.rama.garcia@ine.es
David Salgado, Spanish National Institute; email: david.salgado.fernandez@ine.es

Estimation of Non-Registered Usual Residents in the Netherlands, ultimo September 2010[54]

Bart F.M. Bakker, Susanna Gerritse, Peter G.M. van der Heijden, Jan van der Laan, Rik van der Vliet and Maarten Cruyff[55]

Abstract

From the 2011 Census round, the population "usual residents" is defined as those who have lived or intend to live for a period of more than 12 months in their place of usual residence. Most of the register-based statistics describe the total number of registered persons in e.g. a population register at a particular moment or period. To estimate the non-registered population of usual residents, capture – recapture methods are available. By making a three list estimation, making restrictions to one day for the period-based registers and a short period for the event-based register, applying a stringent linkage method and deletion of erroneous records, most of the assumptions of the capture – recapture method are met. However, to estimate the number of usual residents, we need to divide the estimated total number of persons into those who stay longer than a year in The Netherlands and those who did not. In this paper we present a method to estimate the number of usual residents even though the residence duration is observed in only one of the lists. We apply the method to Dutch empirical data.

Key words:

population estimates, capture-recapture methodology, triple system estimates, undercount, undercoverage

[54] The presented results are preliminary and temporary. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands. We would like to thank Peter-Paul de Wolf, Eric Schulte-Nordholt, Mila van Huis and Kees Prins for their valuable comments on earlier versions of this paper. Direct all correspondence to B.F.M. Bakker, bfm.bakker@cbs.nl

[55] Bart F.M Bakker,Statistics Netherlands and VU University Amsterdam; email: bfm.bakker@cbs.nl
Susanna Gerritse, Utrecht University
Peter G.M. Van der Heijden, Utrecht University and University of Southampton
Jan van der Laan, Statistics Netherlands
Rik van der Vliet, Statistics Netherlands
Maarten Vruyff, Utrecht University

Analysis of Survey Data under Informative Sampling Design and Nonignorable Nonresponse Mechanism

Abdulhakeem A.H. Eideh[56]

Abstract

In this paper, we study, within a modeling framework, the joint treatment of not missing at random response mechanism and informative sampling for survey data, by specifying the probability distribution of the observed measurements when the sampling design is informative. This is the most general situation in surveys and other combinations of sampling informativeness and response mechanisms can be considered as special cases. The sample distribution of the observed measurements model is extracted from the population distribution model, such as the normal distribution. The sample distribution is derived first by identifying and estimating the conditional expectations of first order sample inclusion probabilities, given the study variable, based on a variety of models, such as linear, exponential, logit and probit. Next, we consider a logistic model, probit and other models for the not missing at random response mechanism. The proposed method combines two methodologies used in the analysis of sample surveys for the treatment of informative sampling and not missing at random response mechanism. One incorporates the dependence of the first order inclusion probabilities on the study variable, while the other incorporates the dependence of the probability of nonresponse on unobserved or missing observations. The main purpose here is to consider how to account for the joint effects of informative sampling designs and of not missing at random response mechanism in statistical models for complex survey data.

Key words:

sample distribution, response distribution, nonignorable nonresponse, probability- weighted estimator, informative sampling design.

[56] Abdulhakeem A.H. Eideh, Department of Mathematics, Al-Quds University Abu-Dies Campus, Palestine; email: msabdul@science.alquds.edu

Pairwise likelihood estimation based on a sample of pairs

Ioulia Papageorgiou and Irini Moustaki[57]

Abstract
Pairwise likelihood estimation has been recently developed for estimating the parameters of latent variable and structural equation models. Pairwise likelihood is a special case of composite likelihood methods that use lower order conditional or marginal log likelihoods. The composite likelihood to be maximised is a weighted sum of marginal or conditional loglikelihoods. Weights can be chosen to be equal or unequal for increasing efficiency. In this paper, we approach the problem of weights from a sampling perspective. More specifically, we propose a sampling method for selecting pairs that is based on those pairs that contribute more to the total variance from all pairs. We demonstrate the performance of our methodology using simulated examples.

Key words:
principal component analysis, structural equation models, confirmatory factor analysis

[57] Ioulia Papageorgiou, Athens University of Economics and Business, Department of Statistics;
email: ioulia@aueb.gr
Irini Moustaki, London School of Economics and Political Science, Department of Statistics;
email: i.moustaki@lse.ac.uk

Processes and methods for Big Data: a traffic index based on huge amounts of road sensor data

Piet Daas, Marco Puts, Saskia Ossen and Martijn Tennekes[58]

Abstract

There are many challenges when using Big Data for official statistics. One of them is dealing with data produced in large amounts and at high velocity. This requires processing the data within a reasonable amount of time. A second challenge concerns accuracy and selectivity; essential for the production of official statistics. It is described how road sensor data is used for making reliable official statistics about traffic intensities on the Dutch main road network. Up until recent, only a selection of sensors -not specifically selected for statistical purposes- was used for the production of a traffic index. Because of the poor quality of the road sensor data and the fact that -in this Big Data era- all data can be processed within a reasonable amount of time, there are ways to improve such an index. The poor quality of the road sensor data is caused by three major issues: i) the data contains a lot of noise, ii) the data contains many missing values, and iii) the sensors are not uniformly distributed over (similar sized) road sections. By looking at the road sensor data from a Big Data perspective, we found ways to deal with these quality issues. The first issue is solved by using an adaptive filter tuned for Poisson distributed noise. The second issue is solved making use of the redundancy in the data. The third quality issue is resolved by including the uniqueness of the information provided by individual sensors in the aggregation step. Based on these methods a statistical process has been set up that starts with data collection, is followed by processing and cleaning of the data, and has a subsequent aggregation and dimension reduction step. The latter enables an efficient and nonetheless accurate estimation of the regional traffic intensity on which the monthly and quarterly traffic index are based.

[58] Piet Daas, Statistics Netherlands, pjh.daas@cbs.nl
Marco Puts, Statistics Netherlands, m.puts@cbs.nl
Saskia Ossen, Statistics Netherlands, sjl.ossen@cbs.nl
Martijn Tennekes, Statistics Netherlands, m.tennekes@cbs.nl

Handling missing data in large data sets

Agostino di Ciaccio[59]

Abstract
Often, analyzing administrative data we have a large number of units and variables and many missing observations. Sometimes it is necessary to merge large data sets, in which only some variables are in common.
The correct approach to handle these situations depends on the type of data and the purpose of the analysis. However we can not simply delete the incomplete records, because it amounts to a substantial loss of costly collected data.
In this paper we compare different approaches that can be implemented using specific software. Single imputation, multiple imputation and non-parametric methods will be considered with an application to a European statistical survey.

[59] Agostino di Ciaccio, University of Rome "La Sapienza", Italy; email: agostino.diciaccio@uniroma1.it

Big-data, self-selection, predictions in real-time, intangible variables, opinion and sentiment manipulation: the landing in the future of business statisticians

Furio Camillo[60]

Abstract
The paper derives from the oral report realized for Stakeholders Conference of European Statistics. The basic theme is to share and to discuss some recent experiences relating to the application of statistical analysis in companies and organizations, while maintaining the scenario of context that allows to clarify what are the frontiers of the future that are already apparent today. The report therefore proposes a sequence of problematic connected issues, showing some real cases where challenges to which the statistics will be submitted in the next few years are already evident, given "requests" and "needs" of business users.

Key words:
big data, business statistics, future landscape, real time modelling

References

1. ASQ Global State of Quality Research: Analysis, Trends, and Opportunities 2013, American Society for Quality Press, Milwaukee, 2013.
2. Liberati C., Camillo F.: Discovering hidden concepts in predictive models for texts' polarization, International Journal of Data Warehousing and Mining, (to be published)
http://www.rsc.org/dose/title of subordinate document. Cited 15 Jan 1999
3. Fan, L., Zhang, Y., Dang, Y., & Chen, H. (2013) Analyzing sentiments in Web 2.0 social media data in Chinese: experiments on business and marketing related Chinese Web forums. Information Technology and Management. 14, pp. 231-242.
4. Lee, T. Y., & Bradlow E.T. (2011). Automated Marketing Research Using Online Customer Reviews. Journal of Marketing Research. 68, pp.881-894.

[60] Furio Camillo, Department of Statistical Sciences, University of Bologna; email:furio.camillo@unibo.it

Data Linking: A Common Project for Official Statistics

Tiziana Tuoto, Phillip Gould, Allyson Seyb, Nicoletta Cibella, Monica Scannapieco and Mauro Scanu[61]

Abstract
In National Statistical Institutes (NSIs), current solutions to record linkage problems are commonly based on ad-hoc application-specific tools and procedures. Moreover, the linkage techniques are strictly dependent on the practitioner's skill and there is often little awareness of linkage errors in the analyses on integrated data. Against this background, the "Common and Unified Linkage Tool" project focuses on sharing, assessing and reusing methodological tools for record linkage, stressing the common nature of record linkage problems in NSIs with a view to align methodological solutions to standard procedures. This paper presents a review of six widespread tools proposed by or used in NSIs, on the basis of performance dimensions accounting for both operational and methodological aspects.

Key words:
data integration, data linkage, record linkage software

[61] Tiziana Tuoto, Italian National Statistical Institute; tuoto@istat.it
Phillip Gould, Australian Bureau of Statistics; phillip.gould@abs.gov.au
Allyson Seyb, Statistics New Zealand Tatauranga Aotearoa; allyson.seyb@stats.govt.nz
Nicoletta Cibella, Italian National Statistical Institute; cibella@istat.it
Monica Scannapieco, Italian National Statistical Institute; scannapi@istat.it
Mauro Scanu, Italian National Statistical Institute; scanu@istat.it

Some fallacies and remedies in secondary data analysis for survey data

Giancarlo Manzi, Sonia Stefanizzi, and Pier Alda Ferrari[62]

Abstract
In the last decades, international survey agencies such as Eurostat have been making a huge effort to increase the survey design reliability. Much of the focus has been on improving harmonization and standardization of data, collecting protocols, and eliminating discrepancies among surveys from different countries. However, unsolved problems still remain, especially at the level of secondary analysis from the side of the final researcher who wants to combine data from multiple surveys, different waves of the same survey or surveys referring to different domains. In this paper we first recall some possible sources of internal and differential bias commonly present in large scale surveys and then outline some research paths to overcome (or control) the effects of the lack of comparability and coherence among survey data sources by suitable statistical analysis.

Key words:
Secondary Data Analysis, Survey comparison, Survey estimates.

[62] Giancarlo Manzi, Department of Economics, Management and Quantitative Methods, University of Milan, Italy; email: giancarlo.manzi@unimi.it.
Sonia Stefanizzi, Department of Sociology and Social Research, University of Milan-Bicocca, Italy; email: sonia.stefanizzi@unimib.it.
Pier Alda Ferrari, Department of Economics, Management and Quantitative Methods, University of Milan, Italy; email: pieralda.ferrari@unimi.it.

The best model of a cat is the cat itself

Gejza Dohnal[63]

Abstract
A statistical model is a description of a part of the real world using mathematical concepts. As such in many cases we add a certain layer of abstraction to facilitate our inferential procedure (e.g. normality of measurement errors, compound symmetry in correlation structures etc.). It is almost impossible for a single model to perfectly describe a real-world phenomenon in its in infinite complexity. Nevertheless, successful statistical inference does happen as our world does have a certain degree of consistency which we exploit. Hence, our almost-always-wrong models do prove useful.

[63] Gejza Dohnal, Czech Technical University in Prague, Czech Republic; email: dohnal@nipax.cz

SiGeoS: a user-friendly tool to visualise and communicate statistics

Antonella Bianchino, Annamaria Grippo and Salvatore Cariello[64]

Abstract
In order to extend the manner of use, sharing, and interpretation of statistical data through effective dissemination and communication tools and encourage growth paths of statistical literacy, the Region of Basilicata and the Istat - Office of Basilicata started the project Geographic statistical System (SiGeoS).

Key words:
technology platform , communication data, statistical literacy

[64] Antonella Bianchino, Istat, Italy; email: bianchin@istat.it
Annamaria Grippo, Region Basilicata, Italy; email: annamaria.grippo@regione.basilicata.it
Salvatore Cariello, Istat, Italy; email: cariello@istat.it

Measuring the at-risk-of-poverty rate at local level

Ralf Münnich[65]

Abstract

In the past years, local poverty measurement raised increasingly the interest of politicians, journalists, and the public. This interest was furnished by European and countries' politics to reduce poverty. The European initiative introduced the indicators on poverty and social exclusion and the survey EU-SILC. For details, the reader may refer to the reports from the NET-SILC group or the FP7 projects AMELI and SAMPLE.

When considering local poverty measurement, however, the classical design-based poverty estimates suffer from possibly small or tiny subsamples in the different areas which yield estimates with low accuracy. This urges the needs of introducing small area methods to derive local poverty estimates. The presentation gives an overview of methods and discusses the problem of the impact of sampling designs on small area poverty estimates. An example will be given using at-risk-of-poverty estimates on district-level in Germany.

[65] Ralf Münnich, University of Trier; email: muennich@uni-trier.de

Extracting association rules from two-mode networks

Giuseppe Giordano, Raffaele Miele[66]

Abstract

In recent years the network paradigm has invested many scientific fields and encountered the favor of both hard and soft sciences since its capability to represent complex phenomena with rigorous and well-defined analytical tools and, in the same time, with high substantive interpretative rules. From one hand, Physics, Math, Engineering, Statistics and Informatics contributed with new theoretical results and new network-based measures are offered to applied researcher to explore and analyze its own domain. On the other side, Psychology, Sociology, Business, Marketing, Economy and other applied fields have exploited the substantive ability of a graph representation and, more in general, of network analysis to enhance studies on behaviors, knowledge flows, community formation, leadership, organization and so on. An important success-key for a network-based approach is the large availability of relational data arising from static and dynamic database, new data collection methods and on-line exchange of information among units of interest. This continuous phenomenon of data recording calls for the emergence of data storage, newly defined statistical treatment, and knowledge extraction from huge datasets. In this paper we cope with the problem of extraction of association rules derived by a very large network affiliation matrix and derive a general framework to deal with such kind of data. We define a funnel strategy where raw data are analyzed as in a bipartite graph of transactional data; association rules are derived and evaluated at macro level (the complete graph), meso level (sub-graphs) and micro level (egocentric networks).

[66] Giuseppe Giordano, Dept. of Economics and Statistics University of Salerno, Italy; email: ggiordan@unisa.it
Raffaele Miele, CRISMA

Identification and Estimation of Outcome Response with Heterogeneous Treatment Externalities

Tiziano Arduini, Eleonora Patacchini and Edoardo Rainone[67]

Abstract
This paper studies the identification and estimation of treatment response with heterogeneous spillovers in a network model. We generalize the standard linear-in-means model to allow for multiple groups with between and within-group interactions. We provide a set of identification conditions of peer effects and consider a 2SLS estimation approach. Large sample properties of the proposed estimators are derived. Simulation experiments show that the estimators perform well in finite samples. The model is used to study the effectiveness of policies where peer effects are seen as a mechanism through which the treatments could propagate through the network. When interactions among groups are at work, a shock on a treated group has effects on the non-treated. Our framework allows for quantifying how much of the indirect treatment effect is due to variations in the characteristics of treated peers (treatment contextual effects) and how much is because of variations in peer outcomes (peer effects).
Key words: Networks, Heterogeneous Peer Effects, Spatial Autoregressive Model, Two-Stage Least Squares, Efficiency, Policy Evaluation, Treatment Response, Indirect Treatment Effect.

[67] Tiziano Arduini, Sapienza University of Rome, Italy; email: tiziano.arduini@uniroma1.it.
Eleonora Patacchini Syracuse University, EIEF and CEPR, Italy: email: epatacch@maxwell.syr.edu
Edoardo Rainone, Banca d'Italia and Sapienza University of Rome, Italy; email: edoardo.rainone@uniroma1.it.
The views expressed here do not necessarily reflect those of Banca d'Italia.

Bayesian Network in effectiveness of education

Federica Cugnata, Giovanni Perucca and Silvia Salini[68]

Abstract
A broad literature focused on the effectiveness of tertiary education. In classical models a performance indicator is regressed on a set of individual characteristics of the individuals and fixed effects (FE) at the institution-level. The FE coefficients are interpreted as the pure value added of the universities. The innovative contribution of the present paper resides in the use of Bayesian Networks (BN) to assess the effectiveness of tertiary education. The results of an empirical study focused on Italian universities are discussed, to present the use of BN as a decision support tool for policy making purposes.

Key words:
value added models, do-calculus, casual models

[68] Federica Cugnata, Università degli Studi di Torino, Italy; email: federica.cugnata@unito.it
Giovanni Perucca, Politecnico di Milano, Italy: email: giovanni.perucca@polimi.it
Silvia Salini, Università degli Studi di Milano, Italy; email: silvia.salini@unimi.it

Detecting Influential Observations for a Graphical Model

Avner Bar-Hen and Jean-Michel Poggi[69]

Abstract
Graphical models allow to represent a set of random variables together with their probabilistic conditional dependencies. The focus of this paper is on individual observations diagnosis issues. The use of an influence measure is a classical diagnostic method to measure the perturbation induced by a single element, in other terms we consider stability issue through jackknife. For a given graphical model, we provide tools to perform diagnosis on observations. An application to a real gene expression dataset illustrates the proposals and an investigation of clustering with respect to influence is finally presented.

Key words:
Graphical model, Influence curve, Jackknife

References

1. Ben-Gal, I. (2007). Bayesian networks. In: Ruggeri F., Faltin F. & Kenett R. (Eds.), Encyclopedia of Statistics in Quality and Reliability, John Wiley & Sons.
2. Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. (2005), Robust Statistics: The Approach Based on Influence Functions, Wiley, New York,
3. Lauritzen, S. L. (1996). Graphical models. Oxford University Press.
4. Meinshausen N., Buhlmann P. (2006). High-dimensional graphs and variable selection with the Lasso. Annals of Statistics, 34:1436–1462.
5. Miller, R. G. (1974). The jackknife - a review. Biometrika, 61, 1–15.

[69] Avner Bar-Hen, MAP5, Université Paris Descartes, France;email: Avner.Bar-Hen@mi.parisdescartes.fr
Jean-Michel Poggi Laboratoire de Mathématiques LMO, Université Paris Sud, Orsay, France and Université Paris Descartes, France; email: Jean-Michel.Poggi@math.u-psud.fr

Three-step estimation of latent Markov models

Francesco Bartolucci, Giorgio E. Montanari and Silvia Pandolfi[70]

Abstract
Latent Markov models represent a powerful tool for the analysis of longitudinal categorical data. In the presence of many response variables for each time occasion and with individual covariates, full maximum likelihood estimation of these models may present some critical aspects. Moreover, the simultaneous estimation of the parameters of the measurement model and the latent process may not be easily interpreted by applied researchers. We propose an alternative three-step approach to estimate these models, which is based on a preliminary clustering of sample units on the basis of the time-specific responses only. This method is particularly useful to overcome the drawbacks of the full likelihood approach, with an advantage both in terms of interpretability of the estimation process and in terms of computing time.

Key words:
Expectation-Maximization algorithm, Latent class model, Pseudo likelihood methods

[70] Francesco Bartolucci, University of Perugia (IT), Department of Economics, e-mail: bart@sta.unipg.it
Giorgio E. Montanari and Silvia Pandolfi University of Perugia (IT), Department of Political Sciences, e-mail: giorgio.montanari@ unipg.it;pandolfi@stat.unipg.it

Modelling pollutants' evolution and association over time: a time-varying latent class approach

Antonello Maruotti[71]

Abstract
We introduce a parsimonious latent class model to deal with time depen- dence and dimensionality reduction in a multivariate time series setting. The model is motivated by the analysis of environmental data. Several environmental regimes are identified in the analysis and evidence of different associations structures be- tween pollutants are shown as well.

Key words:
Pollution, Hidden Markov Models, Dimensionality Reduction

References

1. BIERNACKI,C.,CELEUX,G.ANDGOVAERT,G.(2000). Assessing a mixture model for clustering with the integrated completed likelihood. IEEE T. Pattern Analysis, 22: 719–725.
2. FASSO , A., CAMELETTI, M. AND NICOLIS, O. (2007). Air quality monitoring using hetero- geneous networks. Environmentrics, 18: 245–264.
3. MARTINEZ-ZARZOSO, I. AND MARUOTTI, A. (2013). The Environmental Kuznets Curve: functional form, time-dependent heterogeneity and outliers in a panel setting. Environmetrics, 24: 461–475.
4. MENG, X.L. AND VAN DYK, D. (1997) The EM algorithm: an old folk song sung to the fast tune (with discussion). Journal of the Royal Statistical Society - Series B, 59: 511–567.
5. MORLINI, I. (2007). Searching for structure in measurements of air pollutant concentration. Environmetrics, 18: 823–840.
6. SAIDANE, M. AND LAVERGNE, C. (2007). A structured variational learning approach for switching latent factor models. Advances in Statistical Analysis, 91: 245–268.
7. ZUCCHINI, W. AND MACDONALD I.L. (2009). Hidden Markov models for time series: An introduction using R, London: Chapman & Hall.

[71] Antonello Maruotti, Università di Roma Tre, Via Gabriello Chiabrera 199, Roma - Italy; email: antonello.maruotti@uniroma3.it; University of Southampton, Highfield Campus, SO17 1BJ, Southampton-UK, email: a.maruotti@soton.ac.uk

Bayesian non-parametric model to longitudinally predict churn

Bruno Scarpa[72]

Abstract
Churn prediction is one of the most important purposes of customer base analysis for telecommunications companies. Since companies accumulate very large amount of data on customer behaviour, churn prediction is typically achieved by profiling and clustering traffic behaviour jointly with demographic data and contracts characteristics. Unfortunately most algorithms and models used for prediction do not take into account the longitudinal characteristics of data. In particular in telecommunications traffic analysis, it is well known the importance of decreasing pattern of traffic in customer life and it is relevant to aggregate all clients with such a pattern as likely churners. Our approach to tackle such a problem is based on specifying the distribution of functions as a mixture of a parametric hierarchical model describing the decreasing pattern segment and a nonparametric contamination that allows unanticipated curve shapes in subjects traffic. The parametric component is chosen based on prior knowledge, while the contamination is characterised as a functional Dirichlet process.

Key words:
Statistical methods for customer relationship management, Bayesian nonparametrics, functional data analysis, telephone traffic, churn prediction

[72] Bruno Scarpa, Dipartimento di Scienze Statistiche, Università di Padova, e-mail: scarpa@stat.unipd.it

WebGis and Geostatistical Techniques for Monitoring Potential Radon in Soil

Sabrina Maggio, Veronica Distefano and Alessandra Spennato[73]

Abstract

Updated territorial knowledge and its representation are fundamental elements of government directives and local environmental policies. In this paper, the usefulness of geostatistical techniques for monitoring soil radon concentration and the integration of these results in a WebGis, called RnWebGis, are described. This innovative tool offers dynamic scenarios for monitoring and analyzing different environmental variables. In particular, geostatistical analysis is used to identify potential areas of radon risk and allows the detection of clusters related to different levels of radon concentration in the study area. The WebGis system obtained by using an Open Source Software enables online spatial and environmental information to be easily got from the users.

Key words:

Geographic Information System,WebGis, Kriging, Variogram, Radon

[73] Sabrina Maggio , Department of Management, Economics, Mathematics and Statistics, University of Salento, Italy; email: Sabrina.maggio@unisalento.it
Veronica Distefano, Department of Management, Economics, Mathematics and Statistics, University of Salento, Italy; email: veronica.distefano@unisalento.it
Alessandra Spennato, Department of Management, Economics, Mathematics and Statistics, University of Salento, Italy; email: alessandra.spennato@unisalento.it

Generalised Regressions For Spatially and Spatio-Temporally Correlated Data

Lara Fontanella, Luigi Ippoliti and PasqualeValentini[74]

Abstract

Bayesian spatial factor models have already been used in multivariate spatial statistics and dynamic factor models are also common in multivariate time series analysis. This work discusses recent developments of regression models in the literature of spatial and spatio-temporal processes. In particular, we discuss the use of hierarchical Bayesian generalized structural equation models (GSEM) to study the spatial or the spatio-temporal association existing between several dependent and explanatory variables. The model, which encompasses a large class of spatial and spatial-temporal models, is developed for handling measurements belonging to the exponential family of distributions and allow the spatial and temporal components to be modelled conditionally independently via a latent factor analysis structure for the (canonical) transformation of the measurement mean function. The model formulation has an intuitive appeal and enjoy several advantages. For example, it avoids the curse of dimensionality commonly present in many spatial or spatio-temporal data and enables the modelling of the spatial or the temporal relationships between dependent and regressor variables in a latent space. It will be shown that this provides a natural solution to what is known as the misalignment problem (i.e the data may come from independent sources).

Key words:

Bayesian factor models, Structural equation models, Spatial and Spatio-temporal models.

[74] Lara Fontanella, DEc, University G. D'Annunzio Chieti-Pescara, email: lfontan@unich.it
Luigi Ippoliti, DEc, University G. D'Annunzio Chieti-Pescara, email: ippoliti@unich.it
Pasquale Valentini, DEc, University G. D'Annunzio Chieti-Pescara, email: pvalent@unich.it

Spatio-temporal modeling and some applications

Sandra De Iaco, Sabrina Maggio, Monica Palma and Donato Posa[75]

Abstract
Space-time correlation modeling is one of the crucial steps of traditional structural analysis, since space-time models are used for prediction purposes. This paper aims to illustrate the importance and convenience of variogram-based exploratory and prediction techniques to perform a complete spatio-temporal analysis. Moreover, the relevance of choosing a suitable model by taking into account the characteristic behaviour of the models is also provided. In particular an extensive case study regarding PM10 daily concentrations registered at some monitoring stations located in an area with high risk of particle pollution, is faced through the following steps: a) structural analysis, b) estimation of missing values, c) predictions, over the study area, of the PM10 concentrations at time points following the last available observation, d) estimation of the distribution function. Regarding the computational aspects, an ad-hoc program for spatio-temporal predictions is used.

Key words:
space-time random field, spatio-temporal variogram models, kriging, particle pollution

[75] Sandra De Iaco, Department of Management, Economics, Mathematics and Statistics, University of Salento, Italy; email: sandra.deiaco@unisalento.it
Sabrina Maggio, Department of Management, Economics, Mathematics and Statistics, University of Salento, Italy; email: sabrina.maggio@unisalento.it
Sandra De Iaco, Department of Management, Economics, Mathematics and Statistics, University of Salento, Italy; email: sandra.deiaco@unisalento.it
Donato Posa, Department of Management, Economics, Mathematics and Statistics, University of Salento, Italy; email: donato.posa@unisalento.it

Monitoring spatially dependent data coming from sensors: an approach based on functional data analysis

Elvira Romano and Antonio Balzanella[76]

Abstract

The Sensor data have prompted a wide range of research issues. Data delivered by sensors generate large and complex spatio-temporal series whose analysis presents several research challenges. One issue involves outliers detection and data summarization in spatio-temporal dependent data. This paper contributes to this issue by introducing a functional data analysis approach focused on managing sensor data. The method is based on the introduction of the concept of depth for functional data spatially dependent. It allows to get a center-outward ordering of functional data recorded by sensors placed on a geographic area. It is a measure in the framework of classical depths for functional data which accounts for the spatial covariance in the ordering process. It provides an informative graphical tool for visualizing the order of spatial dependence among functional data. Through an application on a real study, we show its effectiveness in monitoring spatio-temporal sensor data behavior.

Key words:

Sensor Data, Functional data Analysis, Depth functions

[76] Elvira Romano and Antonio Balzanella, Department of Political Science "Jean Monnet", Second University of Naples, Caserta, Italy; email: {elvira.romano, antonio.balzanella}@unina2.it

Small area estimation for classifying change: Bayesian modelling applied to crime data

Robert Haining[77]

Abstract
Governments and public and private agencies require data at the small area level to assist with administering programmes and allocating resources for such activities as poverty alleviation, crime reduction (etc) at the local level. The opportunity to meet these needs has perhaps never been better, but to do so in a timely and cost effective way it may be necessary to construct estimates from noisy small area spatial-temporal data. Such estimates provide the basis for mapping but also for highlighting places experiencing abrupt changes or displaying unusual temporal patterns. We describe a methodology to strengthen small area space-time estimates and to classify change in a statistically consistent way. The presentation will give an example using space-time crime data from the city of Peterborough, England.

Key words:
spatial and temporal autocorrelation; borrowing strength; Bayesian hierarchical modelling

References:
1. Li, G., Haining, R., Richardson, S. and Best N: Space-time variability in burglary risk: a Bayesian spatio-temporal modelling approach. Spatial Statistics, 9, 180-191 (2014).

[77] Robert Haining, Department of Geography, University of Cambridge, England; emailrph26@cam.ac.uk

A new paradigm for modelling ordinal responses in sample surveys

Maria Iannario and Domenico Piccolo[78]

Abstract
A growing interest in the current surveys is focussed on human and relational issues collected as ordinal variables. Standard approaches interpret them as manifest expressions of a continuous latent variable and the current methodology
is based on the relationship between the cumulative function of the ratings and the
subjects' variables. A different class of models, called CUB , is based on the statement that ordinal responses are a weighted combination of a personal feeling and an inherent uncertainty surrounding the decisional process. In this paper, the novel paradigm is presented and applied to real data sets to show the advantages of this method for analysing big data in the context of official statistics.

Key words:
Ordinal responses, Cumulative models, CUB models

[78] Maria Iannario, Department of Political Sciences, University of Naples Federico II, Naples, Italy;
email: maria.iannario@unina.it
Domenico Piccolo, Department of Political Sciences, University of Naples Federico II, Naples, Italy;
email: domenico.piccolo@unina.it

Childhood obesity and CUB models: a new perspective

C. Crocetta, E. Zavarrone, V. Russo, B. Cafarelli, L. Antonucci, L. Milani, S. De Francisci[79]

Abstract
The Italian studies on the childhood obesity depicts a painting multicolored where health, psychological, and sociological dimensions seem to be nested. However the polychromy of this landscape does not give a harmonious vision of the studies owing to the insufficient implementation of national guidelines for management of obesity .The application of new statistical techniques could help to better understand the behavior of the populations at risk of obesity. We propose the application of CUB models to the study of obesity. The data is obtained from a survey carried out in March 2008 on diet lifestyle in Milan with a questionnaire administrated in all primary schools (147). The potential population of respondents was 45.000 students families. The coverage rates for schools and respondents were, respectively, 74% and 51%. The survey included the following information: parents and children identification data; the level of parental education; the parents and children fruit and vegetable consumption; the children anthropometric measures, the physical activity frequency and the TV watching habits.

Key words:
Childhood obesity, CUB models, BMI

[79] Corrado Crocetta, B.Cafarelli, L.Antonucci, Univ. Foggia, Italy; email: corrado.crocetta@gmail.com
Emma Zavarrone , V. Russo, L. Milaniand S. de Francisci, Univ. Iulm, Italy; email: emma.zavarrone@iulm.it

Skewness-kurtosis adjusted confidence estimators and significance tests

Wolf-Dieter Richter[80]

Abstract
First and second kind modifications of usual confidence intervals for estimating the expectation and of usual local alternative parameter choices are introduced in a way such that the asymptotic behavior of the true non-covering probabilities and the covering probabilities under the modified local non-true parameter assumption can be asymptotically exactly controlled. The orders of convergence to zero of both types of probabilities are assumed to be suitably bounded below according to an Osipov-type condition and the sample distribution is assumed to satisfy a corresponding tail condition due to Linnik. Analogous considerations are presented for the power function when testing a hypothesis concerning the expectation both under the assumption of a true hypothesis as well as under a modified local alternative. Applications are given for exponential families.

Key words:
orders of confidence, orders of modified local alternatives, true non-covering probabilities, local non-true parameter choice, covering probabilities, Linnik condition, Osipov-type condition, skewness-kurtosis adjusted decisions, order of significance, error probabilities of first and second kind, exponential family

References

1. Linnik, Yu.V.: Limit theorems for sums of independent variables taking into account large deviations. I-III. Theor. Probab. Appl. 6 (1961), 131-148, 345-360; 7 (1962), 115-129; translation from Teor. Veroyatn. Primen. 6, 145-162, 377-391 (1961); 7, 121-134 (1962).
2. Osipov, L.V.: Multidimensional limit theorems for large deviations. Theory Probab. Appl. 20, 38-56 (1975); translation from Teor. Veroyatn. Primen. 20, 40-57 (1975).

[80] Wolf-Dieter Richter, Institute of Mathematics, University of Rostock, Germany;
email: wolf-dieter.richter@uni-rostock.de

A Procrustes-based approach for comparing relational datasets

Simona Balbi, Michelangelo Misuraca and Agnieszka Stawinoga[81]

Abstract
The availability of relational information is a crucial element for depicting processes and dynamics of a given social environment. From Social Network Analysis viewpoint this environment can be expressed in terms of regularities in the connections among social actors. Aim of the paper is proposing a strategy for analysing and comparing two kinds of relations observed on the same set of actors. Relational profiles of the units are firstly investigated by Correspondence Analysis. Subsequently, in order to compare the relational structures a Procrustes Analysis is performed, by superimposing one configuration onto the other one. The evidence of the proposal is showed by carrying out an empirical study.

Key words:
Social Network Analysis, Factorial Analysis, Procrustes Rotations

[81] Simona Balbi, Agnieszka Stawinoga Università Federico II di Napoli, Naples, Italy;
email: agnieszka.stawinoga@unina.it, sb@unina.it
Michelangelo Misuraca, Università della Calabria, Italy; email: michelangelo.misuraca@unical.it

Regulatory data used for macroprudential policymaking: the UK residential mortgage market

Mark H Robson[82]

Abstract

In April 2013 the regulatory system in the UK was radically transformed, when the Financial Services Authority was split into new Financial Conduct and Prudential Regulatory Authorities. The second was created as a subsidiary of the Bank of England, which also acquired a Financial Policy Committee with macroprudential responsibilities, alongside its Monetary Policy Committee established in 1997. In this presentation I explain how data collection, processing and analysis for policy are being conducted in the new legal framework, using the particular example of the new limits on lending at high loan to income (LTI) ratios introduced in October 2014.

[82] Mark H Robson, Head of Statistics and Regulatory Data, Bank of England; email: mark.robson@bankofengland.co.uk

Reaping the benefits of using integrated micro-data for statistical purposes

João Cadete de Matos[83]

Abstract

One of the most striking characteristics of our times is the speed and scope of the change that is under way in the financial landscape, which demands from the statistical function of central banks to be specially attentive and able to devise innovative solutions that are as flexible as possible, so as to allow it to swiftly respond to new and unforeseeable data needs. Against this background, the paper discusses how the Banco de Portugal has been exploiting the statistical potential of a number of available micro-databases, covering different areas of the economy and the financial system, with a view at enhancing the effectiveness and efficiency of its statistical system while keeping the respondents' burden at an acceptable level. The granular nature of such information, together with a good coverage of the relevant population, offers increased flexibility as regards the compilation of new statistics and a more rapid response to ad hoc data requirements from the users. The use of integrated micro-databases for statistical purposes constitutes the cornerstone of the Bank's long-term strategy as regards its statistics. We believe that this approach will pave the way to better address the challenges that lie ahead in this field, whichever they may be.

[83] João Cadete de Matos, Director, Statistics Department; email: jcmatos@bportugal.pt

Integration: a key-word for success in statistical production

Grazia Marchese[84]

Abstract

The paper presents the general features of the framework used by the Banca d'Italia to collect, compile and disseminate statistical data. The use of integrated reporting schemes to satisfy information needs from a variety of functions, the pivotal role of registries, the cross-domain integration of the general datawarehouse, the adoption of well-established standards for data exchange, are the main synergic components of the system. The benefits from such an approach are not for central banks only; on the contrary, they can be exploited by any institution involved in the production of official statistics. That paves the way for an even more intense and successful cooperation between the ESCB and the ESS.

[84] Grazia Marchese, Banca d'Italia; email: grazia.marchese@bancaditalia.it

Robust Clustering of Customs Data

Andrea Cerioli, Domenico Perrotta, Francesca Torti and Marco Riani[85]

Abstract
Robust methods are needed to fit regression lines when outliers are present. In a clustering framework, outliers can be extreme observations, high leverage points, but also data points which lie among the groups. Outliers are also of paramount importance in the analysis of data arising from customs declarations, which may provide information about anomalies like fraudulent transactions. In this work we show that robust techniques can fail when a large proportion of non-contaminated observations fall in a small region, which is a likely occurrence in many customs data sets. In such instances, the effect of a high-density region is so strong that it can override the benefits of trimming and other robust devices. Following the methodology developed by Cerioli and Perrotta (2014), we show how to solve the problem by sampling a much smaller subset of observations which preserves the cluster structure and retains the main outliers of the original data set. We then apply robust clustering methods to the sampled data set for the purposes of classification and outlier detection. We show the potential and the advantages of our method in several empirical applications to data collected through customs declarations in the EU Member States.

Key words:
Anti-fraud analysis, Customs declarations, Robust linear clustering, TCLUST, Thinning

References

1. Cerioli, A., Perrotta D.: Robust clustering around regression lines with high density regions. Advances in Data Analysis and Classification 8, 5-26 (2014)

Mode Validation for Nonparametric Clustering

Marco Perone Pacifico[86]

Abstrac
Nonparametric clustering based on modes is becoming more and more popular as it can be used with large and/or high-dimensional datasets without requiring distributional assumptions. To reduce the possibility of fake clusters, we propose to integrate nonparametric clustering procedures with a bootstrap test that permits to validate the detected modes. The mode test is based on nonparametric confidence intervals for the eigenvalues of the Hessian. In order to get valid bootstrap confidence sets even in presence of multiplicity of the eigenvalues, we use a bootstrap based on an elementary-symmetric-polynomial transformation.

Key words:
bootstrap, density estimation, elementary symmetric polynomials, mean shift algorithm, modes, nonparametric clustering

[86] Marco Perone Pacifico, Sapienza University of Rome, Italy; email: marco.peronepacifico@uniroma1.it

fclust: a toolbox for fuzzy clustering using the R programming language

Maria Brigida Ferraro and Paolo Giordani[86]

Abstract
Fuzzy clustering is used extensively in several domains of research. In the literature, starting from the well-known fuzzy k-means (fkm) clustering algorithm, an increasing number of papers devoted to fkm and its extensions can be found. Nevertheless, a lack of the related softwares for implementing these algorithms can be observed preventing their use in practice. Even the standard fkm is not necessarily available in the most common softwares. For this purpose, a new toolbox for fuzzy clustering using the R programming language has been introduced. The toolbox, called fclust, contains a suit of fuzzy clustering algorithms and fuzzy cluster validity indices. In this talk, after recalling the most used fuzzy clustering methods, the main features of fclust will be described by real data examples.

Key words:
clustering, fuzzy approach, R, fclust

[87] Maria Brigida Ferraro, Sapienza University of Rome, Italy; email: mariabrigida.ferraro@uniroma1.it
Paolo Giordani, Sapienza University of Rome, Italy; email: paolo.giordani@uniroma1.it

Clustering spatial networks through latent mixture models: an application to local labour market

Francesco Pauli, Nicola Torelli and Susanna Zaccarin[88]

Abstract

We consider a model based clustering technique that directly accounts for network relations between subjects and their position in geographical space. The work is aimed at analysing data on mobility patterns of workers, collected for instance in Italy in the Census, to partition territories into local labor markets, i.e., areas with a high proportion of workers both living and working. Social network analysis provides a natural framework to deal with this issue: municipalities are the nodes and the relationship between the nodes is measured by the commuting flows. In our model the flows are explained by the distances between the nodes in a three-dimensional space where two coordinates are the actual geographical coordinates and the third one is a latent variable. The model is completed by specifying a Gaussian mixture distribution for the coordinates: this model component allows us to obtain a clustering of the municipalities based on the flows (having discounted for the populations). The model is estimated on data from Friuli-Venezia Giulia region using the Bayesian approach.

Key words:

network clustering, geographical network, latent variable, Bayesian hierarchical model, local labour markets

[88] Francesco Pauli, Nicola Torelli and Susanna Zaccarin, DEAMS, Università di Trieste, Italy; email {francesco.pauli, nicola.torelli, susanna.zaccarin}@deams.units.it

Statistical problems in earth observing systems: the case of radiosonde collocation uncertainty

Alessandro Fassò[89]

Abstract
The EU Framework Programme for Research and Innovation known as Horizon 2020 has a specific call, namely H2020-EO-3-2014, aimed at mapping the observation capacity in the context of monitoring the climate change. One of the important objectives addressed there is the rigorous scientific assessment of geographical gaps in the system of surface-based and sub-orbital observing capabilities obtained by advanced geostatistical approaches. The quantification of measurement uncertainty of atmospheric parameters is a key factor in assessing the uncertainty of both earth observing system and forecasting numerical models. One of the critical contributions to the uncertainty budget is related to the collocation mismatch in space and time among observations made at different locations and given by vertical atmospheric profiles, obtained e.g. by radiosondes or lidar.
In this paper, we review recent results and future developments on statistical modelling for explaining the relationship between collocation uncertainty and a set of environmental factors, height and distance of imperfectly collocated trajectories.

Key words:
heteroskedastic functional regression, functional kriging, atmospheric vertical profiles.

References

1. Fassò, A, Ignaccolo, R, Madonna, F, Demoz, B. and Franco-Villoria M. (2014) Statistical modelling of collocation uncertainty in atmospheric thermodynamic profiles, Atmospheric Measurement Technology.7, 1803–1816, doi:10.5194/amt-7-1803-2014.
2. Ignaccolo R., Franco-Villora M., Fassò A. (2014) Modelling collocation uncertainty of 3D atmospheric profiles. Stochastic Environmental Research and Risk Assessment. On-line first. DOI 10.1007/s00477-014-0890-7.

[89] Alessandro Fassò, University of Bergamo, Italy; email: alessandro.fasso@unibg.it

Environmental monitoring through biodiversity functional measures

Tonio Di Battista, Francesca Fortuna and Fabrizio Maturo[90]

Abstract
Biodiversity is an essential topic for biologists and its assessment represents an important issue to evaluate the environmental impact. On the other hand, the statisticians deal with biodiversity problems in order to improve methodological aspects. This paper focuses on the use of diversity profiles to assess biodiversity using functional data analysis approach. The aim is to facilitate the stakeholder interpretation in this framework. The proposed approach allows to evaluate the profiles not only from a graphical point of view but also analyzing the profile functional characteristics. In particular, three functional tools has been proposed for biodiversity assessment: curvature, radius of curvature and length of the profile. In this way, biodiversity profile becomes an useful environmental monitoring model.

Key Words:
Biodiversity, biodiversity profiles, functional data analysis, radius of curvature, arc length

[90] Tonio Di Battista, University G. D' Annunzio, Chieti-Pescara, Italy; email: dibattis@unich.it
Francesca Fortuna, University G. D' Annunzio Chieti-Pescara, Italy; email: francesca.fortuna@unich.it
Fabrizio Maturo, University G. D' Annunzio Chieti-Pescara, Italy; email: fabmatu@gmail.com

EEA indicators supporting European policiesù

Roberta Pignatelli[91]

Abstract
This paper describes how the European Environment Agency places its activities of environmental information provider in the context of current European policies. Particular attention is given to the role played by the indicators developed and maintained by the Agency, as a means of connection between the information provided by the data collected and/or managed and the assessments that the agency provides to the policy makers to enable them to make fully informed decisions.

Key Words:
European policies, Indicators, Sustainable development

How to produce and communicate Agricultural Statistics

Eva Laczka[92]

Abstract
The purpose of the presentation is to present what kinds of tools Hungarian statisticians dealing with agricultural statistics have used or developed in the past one, one and a half decade to collect, process and publish agricultural data. Without being exhaustive I will present only the major developments and ideas. In the year 2000, it was a great challenge that we did not have adequate information on the vine and orchard plantations. The difficult task was to elaborate and prepare the "topographic maps" which consisted in using remote sensing data to create topographic maps showing the presumed place of the plantations. Topographic maps were very useful for planning and implementing the plantation surveys, but they were also efficient tools for the quality control of the collected data. The new data processing system developed was called HOMBÁR. HOMBÁR is an IT tool for data processing (not a software) governed and used by statisticians, which improves the quality of data processing, is quicker than traditional processing and decreases considerably the human resources needed. After 5 years of successful functioning of the HOMBÁR, the Hungarian Central Statistical Office has decided to extend the new data processing system to other fields. Currently the data processing system covering nearly two third of the data processing carried out in our Office is called the Standardized Data Processing System (EAR). Besides the development of statistical tools we have also focused on the development of the communication tools. This gave birth to the publication prepared especially for data providers working in agriculture, the "Account calendar", the statistical maps and interactive graphs on agriculture under preparation aimed at presenting plantation data with the use of GIS tools (ÜST) and the series published every year on "the role of agriculture in the national economy".

Key words:
topographic maps, data processing system, respondent's calendar, GIS tools

[92] Eva Laczka; email: eva.laczka@ksh.hu

Tracking users for a targeted dissemination

Ph. Bautier, Ch. Laevaert and B. Le Goff[93]

Abstract
How to build a dissemination and communication strategy in a world where users have easy access to a deluge of data and information from various origins and where IT tools and design standards change so quickly that users behaviour and their expectations are continuously modified? Taking into account the ESS Vision 2020, the first challenge of Eurostat is clearly to know what users want: we know our different types of users but we have to identify how they get our data, what they do with our data, how they react to our outputs and which sort of new service they would like us to propose. This information requires continuous and "real time" feedback from our users. The objective of this paper is to share Eurostat's experience in identifying user needs and to show how concretely this information has been taken into account and integrated.

Key words:
Newsroom, Social media, digital platform

[93] Ph. Bautier, Ch. Laevaert, B. Le Goff, Eurostat; email: philippe.bautier@ec.europa.eu, bernard.Le-Goff@ec.europa.eu

Innovation goal: recent developments of Istat Communication

Giulia Mottura, Anna Maria Tononi[94]

Abstract

Strengthening the position of NSIs is enabled by the promotion of official statistics as a strategic asset for governance and democracy of a country. This is why it is very important to define a strong and recognizable NSI corporate image, consolidate its reputation and the confidence of citizens in official statistics. The development of client-oriented communication to enhance value creation by official statistics is a key factor in Istat innovation and modernization strategy. Innovative communication tools can help respond to increasing and changing user needs. A clear and simple language, the correct use of social media, user-friendly web tools to browse and visualize data, the focus on re-use, open data and effective use of data by media are key drivers to spread innovation in Istat statistical business chain. Communication services are geared towards different targets - policymakers, stakeholders, media, civil society, data users and internal staff. The latter should have access to other tools such as intranet portals, sharing platforms, web seminars, and e-learning systems to enhance work efficiency in a cooperative environment.

[94] Giulia Mottura, Istat, Italy; email: mottura@istat.it
Ariella Martino, Istat, Italy; email: amartin@istat.it

Digital Content: Official Data in a New Light

Alan Smith[95]

Abstract
The increasing maturity and sophistication of open web standards offers huge opportunities for National Statistics Institutues (NSIs) to revitalise their outputs. But what are the characteristics of effective digital content and what implications do they have for traditional product portfolios? This session will feature some new work from ONS' inter-disciplinary Digital Content team along with a discussion of its mission to redefine the ONS product portfolio.

[95] Alan Smith, ONS; email: alan.smith@ons.gsi.gov.uk

Beyond Data Visualization: (Communication) Design Tactics to Embrace Complexity

Paolo Ciuccarelli[96]

Abstract
The complexity of social and natural phenomena becomes more and more evident. As the boundaries of intelligence and decision-making processes are increasingly wider and open, the need of making complexity accessible, understandable and usable - especially for non-experts - is growing. A set of design strategies is presented, with the aim of exploring the potential of Data and Information Visualization and Communication Design in the representation of complex issues. Every design strategy is introduced and discussed through cases and examples.

Key words:
Data Visualization, Information Visualization, Communication Design, Complexity, Uncertainty

References

1. Masud, L.; Valsecchi, F.; Ciuccarelli, P.; Ricci, D.; Caviglia, G., "From Data to Knowledge - Visualizations as Transformation Processes within the Data-Information-Knowledge Continuum," Information Visualisation (IV), 2010 14th International Conference , vol., no., pp.445,449, 26-29 July 2010. doi: 10.1109/IV.2010.68
2. Ciuccarelli, P. "Mind the Graph: From Visualization to Collaborative Network Constructions." Leonardo 47.3 (2014): 268-269 MIT Press Journals.

[96] Paolo Ciuccarelli, Polytechnic University of Milan ; email: paolo.ciuccarelli@polimi.it

Equitable and Sustainable Wellbeing: from measurement to policy

Enrico Giovannini[97]

Abstract
After ten years of initiatives launched around the world, the "Beyond GDP measurement agenda" has produced important results. Now we have a much deeper statistical knowledge of key social and environmental phenomena, especially in Europe. At the same time, the Beyond GDP policy agenda" is still facing huge difficulties, including the strong resistance from orthodox economists who are often providing advices to policy makers. The current crisis should be taken as a historical opportunity to radically change the European political agenda, also to achieve the Europe 2020's and the Sustainable Development's objectives. The presentation will discuss all these issues, showing existing opportunities and challenging the way the revision of the Europe2020 strategy is currently conducted.

[97] Enrico Giovannini, University of Rome "Tor Vergata"; email: enrico.giovannini@uniroma2.it

Europe 2020 at the half-way point: working toward better indicators for social policy making in the EU

Radek Maly[98]

Abstract

Nearing mid-way through the Europe 2020 Strategy, it is clear that its key social targets cannot be met without a significant policy change. The new European Commission stresses the rapidly increasing importance of high quality social indicators in the policy process. A tangible progress toward better social statistics has been achieved over the recent years by the European Statistical System. The experience of the crisis and the renewed policy focus on social indicators represent a unique political opportunity to build on this progress and bring social statistics to a qualitatively new level.

[98] Radek Maly, Head of Unit Social Analysis at DG Employment; email: Radek.Maly@ec.europa.eu

Partnerships to measure progress

Denise Lievesley[99]

Abstract

The vision expressed in Europe 2020 poses exciting challenges for statisticians. It lays out a strategy which "is about more than just overcoming the crisis from which our economies are now gradually recovering. It is also about addressing the shortcomings of our growth model and creating the conditions for a smart, sustainable and inclusive growth. This requires us to think in new ways about the data and analyses which will inform national and European policies. The strategy also sets new targets for measuring progress which we will need to translate into appropriate, sensitive and measurable indicators.

Denise will discuss the importance of strong partnerships to meet these challenges – across the statistical communities in different European countries, linking official statisticians and academic researchers, and building communities of practice drawing statistical skills together with IT and substantive topic expertise.

[99] Denise Lievesley, King's College London, UK; email: denise.lievesley@kcl.ac.uk

Text Mining Infrastructures for Search

Sophia Ananiadou[100]

Abstract

Advanced search systems are primarily enabled by the availability of structured semantic metadata over textual collections. Text mining (TM) plays a key role in semantic metadata generation, driving the extraction of structured information from unstructured documents, which results in annotations in the form of named entities or fine-grained relations between them. Furthermore, text mining allows for accurate clustering and ranking of documents, leading to more meaningful search results. However, many text mining resources differ from each other in terms of various dimensions, e.g., subject domain, input/output format, language and semantic representation. The ability to develop complex solutions by combining different TM resources is thus hindered by issues of incompatibility. This talk will explain how such issues can be alleviated by integrative text mining infrastructures. Specifically, we will demonstrate how the Argo text mining workbench fosters interoperability of resources by giving its users access to an environment that supports the integration of diverse corpora and tools into unified solutions, i.e., workflows composed of elementary processing components. Various use cases supported by Argo workflows, e.g., semantic metadata extraction, will also be illustrated.

[100] Sophia Ananiadour, National Centre For Text Mining School of Computer Science University of Manchester; email: sophia.ananiadou@manchester.ac.uk

iSA: a novel, fast, scalable and accurate algorithm designed to perform sentiment analysis over social network data

Andrea Ceron, Luigi Curini, Stefano Maria Iacu[101]

Abstract
In this paper we present a novel, fast, scalable and accurate version of an algorithm for aggregated sentiment analysis called iSA (integrated Sentiment Analysis) especially designed for the analysis of texts coming from the social network sphere (Twitter, blogs, etc.) Instead of working on individual classification and then aggregating the estimates, iSA estimates directly the aggregated distribution of sentiment over an entire corpus of texts. The analyses presented here show that iSA outperforms other commonly used techniques of individual classification as well as the only other alternative for aggregated sentiment analysis called ReadMe. The reason of this performance is that, while most statistical models or text mining techniques are designed to work on corpus of texts from a given and well defined population, i.e. without misspecification, in reality texts coming from Twitter or other social networks are usually dominated by noise, no matter how accurate is the data crawling. iSA is especially designed to work in this situation, moreover being a supervised technique, is completely independent of the language of the texts, up to the ability of hand coding of the source data.

Key words:
Big Data, Sentiment Analysis, Social Media.

[101] Andrea Ceron and Luigi Curini, Department of Social and Political Sciences, University of Milan; email: andrea.ceron@unimi.it, luigi.curini@unimi.it
Stefano Maria Iacus, Department of Economics, Management and Quantitative Methods, University of Milan, email: stefano.iacus@unimi.it

The Symbolic Approach for Survey Analysis: From Micro-Data to Macro-Data

Paula Brito[102]

Abstract

Symbolic Data Analysis (SDA) provides a framework allowing taking variability intrinsic to the data explicitly into account. This is accomplished by introducing new variable types, whose values may be intervals, finite sets or distributions. SDA is of particular of interest in (large) survey analysis, when the observed sample is partitioned into specific groups which are the focus of interest. This is often the case in official statistics when groups defined e.g., on a regional basis, should be compared and analysed together. The obtained groups, whose descriptions convey their inherent variability – the macro-data - could then be compared and studied together by multivariate methods. By aggregating the micro-data in such way, confidentiality issues are properly dealt with. Moreover, SDA makes it possible to merge independent surveys made on a same population, at a macro-level. Aggregating the concerned surveys using the same criteria (i.e., forming the same "groups"), provides data which may be analysed together, by juxtaposing the columns referring to the (symbolic) variables formed in each case. Data from a same survey applied in different populations may also be merged together providing aggregation is performed on the same variables; this allows for a comparative analysis of corresponding groups in the considered populations.

Key words:

macro-data; symbolic data; survey analysis ; variability in data.

References

1. Brito, P. (2014). Symbolic Data Analysis: another look at the interaction of Data Mining and Statistics. WIREs Data Mining Knowl Discov, 4, 281–295. doi: 10.1002/widm.1133.
2. Noirhomme-Fraiture, M., Brito, P. (2011): Far Beyond the Classical Data Models: Symbolic Data Analysis. Statistical Analysis and Data Mining, 4 (2), 157-170.

[102] Paula Brito, University of Porto, Portugal; email: mpbrito@fep.up.pt

A Multilevel Analysis of Population Registers' Coverage Error in the 15th Italian Census

Luca Mancini and Simona Toti[103]

Abstract
The 15th Italian Population and Household Census was assisted for the first time by Municipal Population Registers (LAC). Millions of household questionnaires were mailed out to residents on the basis of their registered dwelling addresses. A post-censual assessment of LAC's coverage error revealed significant heterogeneity across municipalities in spite of an overall acceptable adherence of registers to census. Pressure is mounting in the public debate to make the transition from a register-assisted to fully-fledged register-based census starting from the next round. The idea is to take LACs at face value whenever they have a good record of being reliable and administer a coverage survey in those municipalities whose registers are known to be less accurate. This paper aims to contribute to the debate by looking more closely at the determinants of coverage error both at the individual and municipal level. The results are expected to be informative as to establishing not only "which" but also the "why" some LACs are more prone than others to misrepresent the true count and composition of the municipal resident population.

Key words:
coverage error, population registers, census, multilevel regression analysis.

[103] Luca Mancini, Istat, Italy; email: lmancini@istat.it
Simona Toti, Istat, Italy; email: toti@istat.it

Distributed Access to Linked Microdata: the Example of ICT and Exports

Eva Hagsten[104]

Abstract
This study describes the novelty of how the relationship between ICT usage in firms and exports behaviour can be analysed across countries by use of the Distributed Microdata Approach (DMD). The approach allows access to official firm-level data over national borders in dimensions not earlier available. The process starts with extensive metadata analysis and harmonisation of microdata (in this case from business registers, production, trade, education and ICT statistics) to ensure comparability over time and across countries. An identical code is then run, linking and aggregating national results to a level where disclosure issues disappear. Aggregated data and analytical results are then pooled into a cross-country dataset called the Micro Moments Database (MMD). Initial results show that there is a relationship between the decision to export and the ICT usage of firms, although the kind of ICT most beneficial vary across countries.

Key words:
Microdata linking, distributed microdata research, ICT, exports
JEL codes: D220, F610 and O33

References

4. Bartelsman, E. J. (2004) The Analysis of Microdata from an International Perspective, STD/CSTAT (2004)12: OECD.
5. Bennett, R. (1997) Export marketing and the internet: Experiences of web site use and perceptions of export barriers among UK businesses. International Marketing Review, 14(5), 324-344.
6. Eurostat (2013) The Multifaceted Nature of ICT, Final report of the ESSNet on Linking of Microdata to Analyse ICT Impact.
7. Hollenstein, H. (2005) Determinants of international activities: are SMEs different? Small Business Economics, 24(5), 431-450.
8. Mathews SW and Bianchi C (2010) The role of the Internet on international market growth of Australian firms: an exploratory study. In: Proceedings of the Australia and New Zealand International Business Academy (ANZIBA) conference 2010, 15-17 April 2010, Sydney, Australia.
9. Melitz, M. J. (2003) The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. Econometrica, 71(6), 1695-1725.

[104] Eva Hagsten, Stockholm University, email: eva.hagsten@scb.se

Hard to reach populations as a challenge for European statistics – experiences based on the survey research of the EU Agency for Fundamental Rights

David Reichel and Sami Nevala, Ursula Till-Tentschert[105]

Abstract
This paper addresses the challenge of obtaining representative and comparable samples of respondents belonging to hard to reach sub-populations in the multi-country survey context of the European Union. FRA – with its mandate to collect reliable and comparable data related to fundamental rights in the EU – has carried out several surveys among specific groups deemed vulnerable to discrimination, victimisation and unequal treatment. Such groups are, amongst others, ethnic minorities, immigrants as well as lesbian, gay, bisexual and trans (LGBT) persons. This paper presents successes and challenges in efforts to draw samples from these groups.

Key words:
sampling, hard to reach populations, fundamental rights

[105] David Reichel, European Union Agency for Fundamental Rights, david.reichel@fra.europa.eu
Sami Nevala, European Union Agency for Fundamental Rights, sami.nevala@fra.europa.eu
Ursula Till-Tentschert,,European Union Agency for Fundamental Rights

Applying LDA topic model to a corpus of Italian Supreme Court decisions

Paolo Fantini and Pierpaolo Brutti[106]

Abstract
We investigate the use of topic models such as latent Dirichlet allocation (LDA) on two real-world problems: classifying documents and retrieving similar documents to a given document, with reference to a corpus of Italian Supreme Court decisions. A topic model is a generative model that specifies a simple probabilistic procedure by which documents in a corpus can be generated. In the LDA approach, a topic is a probability distribution over a fixed vocabulary of terms, each document is modeled as a mixture of K topics, and the mixing coefficients can be used for representing documents as points on K-1 dimensional simplex spanned by the topics. Approximate posterior inference is performed in order to learn the hidden topical structure from the observed data, i.e., the words in the documents.

Key words:
topic model, LDA, hidden variable models, approximate posterior inference, document classification, document similarity.

[106] Paolo Fantini, Statistical Service of the Italian Ministry of Justice, Italy; email: paolo.fantini@giustizia.it
Pierpaolo Brutti, Department of Statistical Sciences, Sapienza University of Rome, Italy;
email: pierpaolo.brutti@uniroma1.it

The merits of micro-aggregated linked firm-level data: An application of labour demand by SMEs for 14 European Countries

Martin Falk and Eva Hagsten[107]

Abstract
In the absence of access to linked firm-level data sets, this paper demonstrates that micro-aggregated data based on existing official data sources can be useful for analysing dynamic models of firm behaviour. The dimensions emphasized are industry affiliation and firm size for 14 EU countries for the period 2002-2010. The results based on dynamic panel data models show that employment depends significantly on wage costs, output, capital stock, skill intensity and technological innovations. In particular, the employment effects of new market products are positive with larger effects for small and medium-sized than for large enterprises. In large firms there is a negative employment effect of process innovations whereas in SMEs employment benefits from process innovations. Overall, micro-aggregated firm level data offer new prospects for analysing dynamic models of firm behaviour by secondary usage of existing official data sources.

Key words:
cross-country panel data, firm size, labour demand, technological innovations, dynamic panel data models.

References

1. Bartelsman, E, Hagsten, E and Polder, M. (2013), Cross-country analysis of ICT impact using firm-level data: Micro Moments Database and Research Infrastructure, ESSLAIT project.
2. Hamermesh, D. S. (1996). Labor Demand, Princeton, NJ: Princeton University Press.
3. Harrison, R., Jaumandreu, J., Mairesse, J., and Peters, B. (2008). Does innovation stimulate employment? A firm-level analysis using comparable micro data on four European countries. NBER working paper 14216.

[107] Martin Falk, Austrian Institute of Economic Research (WIFO),email: martin.falk@wifo.ac.at
Eva Hagsten, Stockholm University; email: eva.hagsten@scb.de

Assessment of statistical literacy: a pilot survey on college students. The case of University of Pisa

Alessandro Valentini, Monica Pratesi, Bianca Maria Martelli, Silvia Da Valle, Luca Faustini, Linda Porciani, Graziella Sanna and Claudia Tinelli[108]

Abstract
The promotion of Statistical Literacy (SL) is one of the strategic goals of Istat, even if little attention has been devoted to the assessment of it. For the first time in Italy, a pilot field study was realized, thanks to the cooperation between the Tuscany and Umbria Territorial Office of Istat (TEU) and the Department of Economics and Management (DEM) of the University of Pisa. The questionnaire (named QValStat) was administered to more than 600 students of the courses in statistics held in DEM. Results revealed the not so high level of SL and the need of specific coordinated actions adopted both by Istat and DEM. Assessment procedures like QValStat could contribute to drive policies for a "science with and for society" in line with Horizon 2020 work program.

Key words:
statistical literacy dissemination, assessment, pilot survey

[108] Alessandro Valentini, Bianca Maria Martelli, Silvia Da Valle, Luca Faustini, Linda Porciani, Graziella Sanna and Claudia Tinelli , Istat, Direzione Centrale per lo sviluppo e il coordinamento della rete territoriale e del Sistan. Email: alvalent@istat.it, bmartelli@istat.it, davalle@istat.it, faustini@istat.it, grsanna@istat.it, tinelli@istat.it
Monica Pratesi, Università di Pisa, Dipartimento di Economia e Management; email: monica.pratesi@unipi.it

JRF DATA – NEW interactive data visualisation tool for official statistics about Poverty

Aleksandra Collingwood[108]

Abstract
The Joseph Rowntree Foundation (JRF) is the place to get the facts on poverty in the UK. We have built a fantastic data resource of official statistics on our website with interactive graphs, maps and data to download, to appeal to both statisticians, journalists, academics and the general public www.jrf.org.uk/data.
This resource is linked to our annual State of the Nation Poverty and Social Exclusion report of which the 2014 edition will be published in November. We have a lot of interest from the Royal Statistical Society UK in our new resource and I have presented aspects of it at their annual national conference. As well as being informative and educational to the delegates I hope it will also open discussion as to how we can potentially expand our resource and also present European official data.

[108] Aleksandra Collingwood, Statistics Specialist and Programme Manager of JRF Data, Joseph Rowntree oundation, aleks.collingwood@jrf.org.uk

Minos: a journey through sport, theatre and statistic culture

Carbonari Cristina, Ciaffarafà Michela, Paradisi Francesca[110]

Abstract
To bring statistics to centre stage using theatrical performances, in which the concepts of statistics blend with aspects of daily life, is an experimented tool used to diffuse the complex concepts of this subject and make it easy to understand and within everyone's grasp. Thus, it puts into practice new ways of learning other than the traditional methods, in response to the principles of lifelong learning. It creates an opportunity to confront a text, a message and a series of data, to implement an innovative way of visualising and communicating statistics. A cultural activity devised in this way constitutes an interdisciplinary model of cultural transmission, which can be reproduced in more than one social environment and can be diffused via multimedia tools.

Key words:
statistical culture, life-long learning, interdisciplinary, theatrical performance.

[110] Cristina Carbonari, Istat, Italy; email: carbonar@istat.it
Michela Ciaffarà, Istat, Italy; email: mciaffa@istat.it
Francesca Paradisi, Istat, Italy; email: paradisi@istat.it

How to increase statistical literacy in primary school: a project of the Istat territorial network

Susi Osti, Silvia Da Valle, Paola Francesca Cortese and Alessandro Valentini[111]

Abstract
As well known, statistical literacy assumes a key role in the growing process of our knowledge society. As a consequence Istat, the Italian NSI, devotes a great deal of attention to this question. Aim of the present paper is to illustrate methods and results of the action plan ideated by the territorial network of the Institute to support teachers in their work with primary school students. After a brief outline of the role of territorial experts, core of this paper is the description of the educational tools, available for free download on Istat institutional website. The tools were specifically designed and field-tested (by around 800 students and 50 teachers) in 45 primary school classes. Finally, some perspectives for the future are proposed.

Key words:
statistical literacy, educational tools, lifelong learning, citizenship

[111] Susi Osti, Istat, Italy; email: osti@istat.itosti@istat.it
Silvia Da Valle, Istat, Italy; email: davalle@istat.itdavalle@istat.it
Paola Francesca Cortese, Istat, Italy; email: pacortes@istat.itpacortes@istat.it
Alessandro Valentini, Istat, Italy; email: alvalent@istat.italvalent@istat.it

Communicating through Social Networking Tools in Istat

Ariella C. Martino and Michela Troia[112]

Abstract
Social media are considered as broadcasting channels with a high potential to reach the target audience, yet inexpensive, versatile and often interoperable with each other. Social networks can also enlarge the audience and encourage interactivity with citizens, enabling listening, involvement and engagement. For these reasons, Istat decided to be present on more than one social media, even reachable now from the corporate website and the RSS feeds. Through the social web Istat changes its point of view and it establishes a privileged channel of communication with traditional and new generation stakeholders

Key words:
web, social network, information, communication

[112] Ariella C. Martino, Istat, Italy; email: armartin @istat.it
Michela Troia, Istat, Italy; email: mitroia @istat.it

Quality and aggregation issues in GDP and beyond: insights from the OECD Better Life Initiative

Carlotta Balestra[113]

Abstract
The OECD Better Life Initiative measures well-being in OECD countries and in a few emerging economies. It relies on a dashboard of headline and secondary indicators that are built on various datasets mostly coming from official statistics. The OECD dashboard includes a few data from non-official statistics, which are considered as place-holders until harmonised official surveys will become available. The presentation first discusses quality issues of official and non-official statistics in the context of the OECD work on well-being; it stresses that, given the limitations of non-official statistics, NSOs should be encouraged to put in place an integrated survey that cover all relevant well-being domains and that this survey ought to be harmonised at international level. The presentation then discusses aggregation issues when interpreting and analysing the well-being picture emerging from a dashboard of indicators in the light of the Better Life Initiative experience. It is argued that dashboards and composite indices are not necessary in contradiction to each other but they can be used at different stages of well-being analysis, providing meaningful insights on a range of different factors. The presentation concludes by highlighting the benefits for NSOs to make use of an integrated comprehensive well-being survey, notably to understand the joint correlation of well-being outcomes at individual level.

[113] Carotta Balestra, OECD; email: carlotta.balestra@oecd.org

Lost in Aggregation

Michaela Saisana, Andrea Saltelli and Paolo Paruolo[114]

Abstract
This presentation uses global sensitivity analysis to discuss potential flaws of weighted arithmetic averages - popular tools in the public discourse when it comes to assessing countries' performance. It highlights how the variables' weights cannot be considered as measures of importance of the components due to their covariance structure. Two approaches are compared: the Pearson correlation coefficient and a fully nonparametric estimation of the Pearson correlation ratio. We use three case studies on environment, governance and innovation to suggest: (a) how the importance of variables often diminishes through the different aggregation levels; (b) some of the statistical properties of a well-built composite indicator.

Key words:
composite indicator, weights, linear aggregation, correlation coefficient

References

1. Paruolo P., Saisana M., Saltelli A.: Ratings and Rankings: voodoo or science?. J Royal Statistical Society A, 176(3), 609—634 (2013)
2. Saisana M., D'Hombres B., Saltelli A.: Rickety Numbers: Volatility of university rankings and policy implications. Research Policy, 40, 165—177 (2011)
3. Saisana M., Saltelli A., Tarantola S.: Uncertainty and sensitivity analysis techniques as tools for the analysis and validation of composite indicators. J Royal Statistical Society A, 168(2), 307—323 (2005)

[114] Michaela Saisana, European Commission, Joint Research Centre, Ispra, Italy;
email: michaela.saisana@jrc.ec.europa.eu
Andrea Saltelli, European Commission, Joint Research Centre, Ispra, Italy;
email: andrea.saltelli@jrc.ec.europa.eu
Paolo Paruolo, European Commission, Joint Research Centre, Ispra, Italy;
email: paolo.paruolo@jrc.ec.europa.eu

Component-based Multi-block Path Modeling for building Composite Indicators

Laura Trinchera, Giorgio Russolillo and Vincenzo E. Vinzi[115]

Abstract
According to Saisana et al. (2002), a composite indicator is a mathematical combination of single indicators that represent different dimensions of a concept whose description is the objective of the analysis. In this optic, Multidimensional Data Analysis (MDA) approach seems to be the most natural tool to compute composite indicators In the talk we will discuss the use of PLS-Path Modeling (PLS-PM) (Tenenhaus et al. 2005; Esposito Vinzi et al., 2012) as well as of other component-based approaches whose features naturally fit the aim of computing composite indicators (Trinchera et al., 2008). Until now composite indicators have been obtained only as a mathematical (mostly linear) combination of (quantitative) indicators. Nevertheless, considering also categorical indicators in building composite indicators is very relevant. Therefore, in the talk we will also discuss the Non-metric PLS-PM approach (Russolillo, 2012). Finally, in the same framework, we will discuss how to deal with the problem of either observed or unobserved heterogeneity across a group structure in the data

Key words:
PLS-PM, categorical indicators.

References

1. Esposito Vinzi, V., Russolillo, G.: Partial least squares algorithms and methods, WIREs Comp Stat, 5, 1–19 (2012).
2. Russolillo, G.: Non-metric Partial Least Squares. EJS, 6, 1641-1669 (2012).
3. Saisana, M., Tarantola, S.: State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development, EUR 20408 EN, European Commission-JRC: Italy (2002).
4. Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.M., Lauro, C.: PLS Path Modeling, CSDA, 48, 159-205 (2005).
5. Trinchera, L., Russolillo, G., Lauro, C.: Using categorical variables in PLS Path Modeling to build system of composite indicators, Statistica Applicata, 20 (3/4), 309-330, RCE Multimedia, ISSN: 1125-1964 (2008).

[115] Laura Trinchera, NEOMA Business School; email: laura.trinchera@neoma-bs.fr
Giorgio Russolillo, CNAM; email: giorgio.russolillo@cnam.fr
Vincenzo Esposito Vinzi, ESSEC Business School; email: vinzi@essec.edu

Official Statistics and Composite Indicators: Developing the Innovation Output Indicator

Carsten Olsson[116]

Abstract
The trend towards evidence based policy making has created a corresponding need for more and better data to monitor performance, progress, and the impact of the implementation of new policies. Such data should ideally provide details on specific performance elements related to these policies as well as global information on the progress towards the goals of the policy initiatives and their impact.
This paper examines the perimeter within which official statistics contributes to monitoring the impact of medium scale policy initiatives and how composite indicators can function as a supplementary tool for policy makers. This is illustrated by a case study on the Innovation Output Indicator. Some of the methods for constructing Composite Indicators are highlighted.

Key words:
Composite indicators, Principles for Official Statistics

[116] Carsten Olsson, Head of Unit G4 "Innovation and information society", Eurostat - European Commission; email: Carsten.olsson@ec.europa.eu

How to Improve Statistical Literacy?

Anuška Ferligoj[117]

Abstract
While there are several definitions of statistical literacy, most are based on the definition given by Katherine K. Wallman (1993) in the speech she delivered when she became President of the American Statistical Association: "Statistical literacy is the ability to understand and critically evaluate statistical results that permeate our daily lives – coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions". In the first part of the paper current initiatives and latest publications with several ideas and good practices for improving statistical literacy are highlighted. In the second part some recommendations for the main actors dealing with statistics are offered. These actors are: educational institutions, statistical offices and other statistical institutions, statistical societies and media. It is pointed out that the coopertion of these actors is essential for improving statistical literacy.

Key words:
statistical literacy, statistical thinking, statistical reasoning, statistical data, educational institution, statistical office, statistical institution, statistical society, media

References

1. Wallman, K.K.: Enhancing Statistical Literacy: Enriching Our Society. JASA, 88, 1-8 (1993)

[117] Anuška Ferligoj, Faculty of Social Sciences, University of Ljubljana, Slovenia;
email: anuska.ferligoj@fdv.uni-lj.si

EMOS - European Master in Official Statistics

Markus Zwick[118]

Abstract

The European Master in Official Statistics (EMOS) is a project aimed at developing a programme for training and education in Official statistics within existing Master programmes at European universities.

In 2012, Eurostat launched the feasibility study 'Towards a European Master in Official Statistics (EMOS)'. During the study, more than 700 European universities were asked about their Master programmes' structure, most of them via an online survey. In 14 European countries face-to-face interviews with universities and NSIs took place. Based on the results of the study, presented in December 2013, a Group of Experts (GExp), from universities and NSIs together with Eurostat, developed a concept of how to implement EMOS at European university level. At the 21st meeting of the European Statistical System Committee (ESSC) in May 2014, the Directors General of the NSIs agreed to the concept. Further to this, on 26-27 June 2014, an international EMOS workshop presented the concept to NSIs and universities.

The concept for the EMOS label foresees a flexible curriculum structure, suitable to different types of Master degrees such as Masters in Statistics, Survey Methodology, Economics or Social Sciences. It also foresees a concrete syllabus, implementation plan, and an EMOS governance model.

A first 'Call for Interest' runs until 15 September 2014. All European universities that are interested in having the EMOS label for their Master programmes may participate. The Group of Experts will preselect the universities that are best prepared to reach the quality requirements of EMOS. These universities will subsequently be invited to further develop their EMOS programme and apply for the EMOS label by end of February 2015. Master programmes which fulfil the EMOS criteria will be awarded the EMOS label by the ESSC at their meeting in May 2015.

One of the objectives of the project is also to explore opportunities how EMOS could be used for further training of staff of the European Statistical System. One of the solutions is the modern interactive online technologies used to share the EMOS university courses on a wider scale. The eSummerschool organised by the Universities of Trier and Pisa together with Eurostat in 2013 is a good examples illustrating these opportunities.

[118] Markus Zwick, Eurostat, Luxembourg; email: markus.zwick@ec.europa.eu

Statistical Professionalism: Recognising Impact for the Public Good

Roeland Beerten[119]

Abstract
The Royal Statistical Society (RSS) has been at the forefront of setting professional standards in the statistics profession for over twenty years, and is a world-wide leader in awarding recognition of the professional skills of statisticians. Through the use of a clear framework for both educational and professional standards the Chartered Statistician award is a clear external recognition of the skill set of the professional statistician, and is a guaranteed kite mark to ensure the wider world can trust the quality and professionalism of the statistician's work. This presentation will give an overview of the frameworks and processes we use for professional accreditation and continuing professional development. It will conclude by explaining how the RSS ensures statistical skills continue to be relevant to today's debates by illustrating how statisticians can make an impact in the wider world beyond statistics.

Key words:
statistics, professional accreditation, continuing professional development.

[119] Roeland Beerten, Director of Professional and Public Affairs, Royal Statistical Society, UK;
email: r.beerten@rss.org.uk

Statistics at MOX: an excursion in high dimensional and complex data analysis

Piercesare Secchi[120]

Abstract
The analysis of high dimensional and complex data requires the development of new statistical tools urged by new challenges which require a strong
interplay of statistics with other scientific disciplines, including maths, engineering, scientific communication design, computer graphics, computer science, and information technology. This mix of expertises profiles a new type of statistician whose skills go far beyond the classical statistical academic education in the direction of a more eclectic background in which statistics is not the unique piece of the puzzle but the one which all other pieces are connected to. By means of a few case studies, I will disseminate the experience developed by the statistics group at the MOX laboratory of the Department of Mathematics of Politecnico di Milano.

Key words:
mathematical engineer, data scientist.

[120] Piercesare Secchi, MOX-Department of Mathematics, Politecnico di Milano, Italy; email: piercesare.secchi@polimi.it

Collecting survey data to support fundamental rights policies in the EU

Sami Nevala and David Reichel, Ursula Till-Tentschert[121]

Abstract
In 2007, the European Union Agency for Fundamental Rights (FRA) was established with the specific task of providing independent, evidence-based advice on fundamental rights. In order to fulfil its mandate, the Agency is engaged in collecting and analysing comparable data for EU-28 – this includes designing and carrying out specialised EU-wide surveys, targeting special topics and groups. The work of FRA is therefore not only challenging in terms of collecting statistical data on sensitive topics and hard-to-reach groups in an multi-country setting, but also with respect to communicating the results to policy makers and other stakeholders. This paper provides an overview of the experiences FRA has accumulated through collecting comparable data by means of survey research among minority and immigrant groups, LGBT persons and Jewish people as well as a general population survey concerning violence against women in the EU.

Key words:
Fundamental Rights, European Union, evidence-based advice, violence against women

[121] Sami Nevala, European Union Agency for Fundamental Rights, sami.nevala@fra.europa.eu
David Reichel, European Union Agency for Fundamental Rights, david.reichel@fra.europa.eu
Ursula Till-Tentschert, European Union Agency for Fundamental Rights

Estimating equivalised income inequality in small areas

Enrico Fabrizi, Maria Rosaria Ferrante and Carlo Trivisano[122]

Abstract

In this article, we focus on the estimation of the Gini concentration coefficient for small areas. Specifically, we consider the estimation of the coefficient for the concentration of equivalised disposable household income for small geographic regions (health districts) in Italy. The methodology for estimating the coefficient from large sample surveys is well established but the available samples are very small, the bias of ordinary design based estimators of the coefficient are non negligible and associated variances too large to allow for reliable inferences. A first objective of this paper is to propose a modified design-based estimator of the Gini coefficient characterized by a smaller bias in small samples and equivalent to the usual survey weighted estimator when the sample size grows large. To improve precision of direct estimators we resort to small area methods relying on area level models. Specifically, we propose a Beta regression model that we analyze adopting a Bayesian approach. The proposed methodology proved to be very effective and corrects the main flaws of direct estimation.

[122] Enrico Fabrizi, DISES, Università Cattolica del S. Cuore, Piacenza, Italy; email: enrico.fabrizi@unicatt.it
Maria Rosaria Ferrante and Carlo Trevisano, Dipartimento di Scienze Statistiche 'P. Fortunati', Università di Bologna

Poverty in Europe: the obscured message of official statistics

Jean-François Royer[123]

Abstract
One of the five target-indicators of the EU-2020 strategy is devoted to poverty. In Eurostat's publications of social statistics, this indicator is highlighted and analyzed. This indicator is far from being straightforward: it is a complicated construction produced by political negotiations. Hence the description of poverty in Europe given to the general public by the statistical institute is obscured. This raises the problem of the choice of concepts and definitions for statistical investigation: shouldn't statisticians have some degree of independence in that choice?

Key words:
poverty, independence

[123] Jean-François Royer, Société Française de Statistique, France; email: royerjf@free.fr

The causes of Polish emigration: a latent class analysis approach

Ewa Genge[124]

Abstract
Latent class analysis can be viewed as a special case of model-based clustering for multivariate discrete data. It is assumed that each observation comes from one of a number of classes, groups or subpopulations, with its own probability distribution. We used the model based clustering approach to find groups of Poles with similar attitudes to emigration (having the similar causes for emigration). We analyzed data collected as part of the Polish Social Diagnosis using the poLCA package of R.

Key words:
latent class analysis, model-based clustering, emigration

[124] Ewa Genge, Department of Economic and Financial Analysis, University of Economics in Katowice, Poland; email: ewa.genge@ue.katowice.pl

Data reduction and visualization for categorical data

Alfonso Iodice D'Enza, Michel van de Velden and Francesco Palumbo[125]

Abstract
Much of the data that are daily collected are categorical: survey data represent the most typical example of categorical data. However, many others exist: categorical data are automatically and continuously collected on a large (even huge) number of units. Cluster analysis represents one of the most powerful and widely used approaches to find homogeneous groups in data. Clustering methods have been developed and used in most of scientific fields with the aim of identifying general typologies and regularities in data. The majority of clustering methods proposed in the literature are designed to deal with numerical data rather than categorical data. The categorical data clustering problem has some peculiarities that make it not straightforward to adapt clustering methods originally designed for numerical data.
We propose a comparison among four methods for clustering categorical data that are based on the simultaneous data reduction and clustering strategy, and we highlight the advantages of such integrated approach in some real contexts.

[125] Alfonso Iodice D'Enza, Università di Cassino, Italy; email: iodicede@unicas.it
Michel Van de Velden, Erasmus University of Rotterdam, Netherlands; email: vandevelden@ese.eur.nl
Francesco Palumbo, Università degli Studi di Napoli Federico II, Italy; email: fpalumbo@unina.it

Multidimensional Scaling for Latent Dissimilarity Data

José Fernando Vera[126]

Abstract
In this paper, applications are shown of a general MDS model for latent dissimilarity data, for which the actual observed dissimilarities are considered indicators of an unknown dissimilarity matrix. In the context of least squares MDS allowing transformations, an alternating estimation procedure is employed where the unknown dissimilarity matrix is estimated in a covariance structure framework, while the objects are represented in a low dimensional space. Applications of the model to the situation in which several observed dissimilarity matrices are indicators of an unknown overall dissimilarity matrix (i.e., three-way data) and to the problem of MDS for asymmetric data, are showed in the context of this proposed Structural Equation Multidimensional Scaling (SEMDS) framework.

Key words:
Multidimensional Scaling, Structural Equations Models, SEMDS, Least squares.

[126] J. Fernando Vera, Department of Statistics and O.R. University of Granada, Spain; email: jfvera@ugr.es

A multi item scale to assess entrepreneurship levels in the Italian agriculture

Benedetto Rocchi, Gianluca Stefani, Chiara Landi, Massimo Greco, Daniela Fusco, Paola Giordano and Valerio Moretti[127]

Abstract
The entrepreneurship attitude of the farm manager is one of the main factors affecting the productivity in agriculture and hence the survival of farms. This paper presents a multi-item index aimed at evaluating the entrepreneurship capacity of Italian Farmers. The index is based on several variables such as the presence of bookkeeping, the use of ICT services, the market orientation of activities etc. The data stems from the 2000 and 2010 Census. Entrepreneurship scores are assigned at the single farm level and are aggregated at the municipality level. Results show evident spatial patterns in the distribution of agricultural entrepreneurship in Italy. The index has been validated through a test of internal coherence, a principal component analysis, and testing the association of entrepreneurship levels with farms survival and multifunctionality.

Key words:
entrepreneurship capacity, multi-item index, spatial analysis, Principal component analysis.

[127] Benedetto Rocchi and Gianluca Stefani, University of Florence, Italy; email: benedetto.rocchi@unifi.it and gianluca.stefani@unifi.it
Chiara Landi, University of Pisa, Italy; email: chiara.landi@for.unipi.it
Massimo Greco, Daniela Fusco, Paola Giordano and Valerio Moretti, Istat, Italy; email: msmagrec@istat.it, dafusco@istat.it, pgiordano@istat.it and vmoretti@istat.it

Modelling the Passenger Satisfaction of the Public Transportation Services: The Metro Brescia Survey

Luigi D'Ambra, Maurizio Carpita and Enrico Ciavolino[128]

Abstract
The present contribution shows an integrated approach based on Simple Component Analysis (SCA) and Second-order Partial Least Squares - Path Model (PLS-PM) to evaluate the passenger satisfaction and to define a decision support system for the local administration. An application to a Metro Brescia Survey is presented.

Key words:
Passenger Satisfaction, Simple

[128] Luigi D'Ambra, Department of Economics, Management, Institutions - University of Napoli Federico II, Italy;
email: dambra@unina.it
Maurizio Carpita, Department of Economics and Management - University of Brescia, Italy;
email: maurizio.carpita@unibs.it
Enrico Ciavolino, Department of History, Society and Human Studies - University of Salento, Italy;
email: enrico.ciavolino@unisalento.it

Big data may be more personal than it appears!

Achim Klabunde[129]

Abstract
The concept of big data is often understood as an invitation to be less selective and cautious when collecting and retaining data from different sources. Economic incentives for deleting data disappear when storage prices plummet and potential benefits from big data analytics might be enormous. Legal and ethical problems are supposed to be overcome when data is chosen that does not contain identifiers. However, with increasing density and precisions of data sources, new data sources created by sensors and internet of things devices, the data web may become so dense that individuals can still be found and identified in data that did not appear to be related to any identity.

[129] Achim Klabunde, edps europa; email: achim.klabunde@edps.europa.eu

Policies for a fair re-use of data: Big Data and the application of anonymisation techniques

Giuseppe D'Acquisto[130]

Abstract
The legal framework on data protection sets a high standard in Europe with regard to the possible re-use of personal data. Principles like purpose limitation and data minimization challenge the emerging Big Data paradigm, where the "value" of data is linked to its somehow still unpredictable potential future uses. Nevertheless, the re-use of data is not impossible. The EU's Article 29 Working Party published, in April 2014, an Opinion on the application of anonymisation techniques, which can be implemented to enable potential re-use of previously collected data within a framework of safeguards for individuals. The paper proposes a fair anonymisation policy in order to make fair reuse of data possible, giving an overview of the legal and technical aspects related to anonymisation, and pointing out the many misconceptions on the issue.

Key words:
Data Protection, Policy, Anonymisation, Accountability

[130] Giuseppe D'Acquisto - Garante per la protezione dei dati personali - Piazza di Monte Citorio n. 121 – 00186, Roma; email: g.dacquisto@gpdp.it

Privacy principles under pressure in the age of Big Data analytics

Catharina Nes[131]

Abstract
Big Data entails a new way of looking at data, revealing information that may have been previously difficult to extract or otherwise obscured. To a large extent, Big Data involves the reuse of data. The value of the data may be linked to its ability to make predictions about future actions or events. Such a view of data may challenge key privacy principles, in particular the principles of purpose limitation and data minimisation. The "Big data-mindset" is based on the assumption that the more data you collect and have access to, the better, more reasoned and accurate decisions you will be able to make. But collection of more data may not necessarily entail more knowledge. More data may also result in more confusion and more false positives. Extensive use of automated decisions and prediction analyses may have adverse consequences for individuals. There is a risk that Big Data may consolidate existing prejudices and stereotyping, as well as reinforce social exclusion and stratification. In spite of the fact that Big Data entails several privacy challenges, it is possible to make use of this type of analysis without infringing on key privacy principles. The International Working Group on Data Protection in Telecommunications (aka the Berlin-group) has recently issued a working paper that makes recommendations regarding how Big Data may be used in ways that will respect the privacy of each individual. Recommendations from this paper will be outlined in the presentation.

Key words:
Big Data, privacy challenges, discrimination

References

1. International Working Group on Data Protection in Telecommunications (2014), "Working Paper on Big Data and Privacy - Privacy principles under pressure in the age of Big Data analytics", file:///C:/Users/cane/Downloads/WP_Big_Data_final_clean_675.48.12%20(2).pdf

[131] Catharina Nes, datatilsynet; email: catharina.nes@datatilsynet.no

Indirect Questioning and Protection of Privacy in Sample Surveys

Tasos C. Christofides[132]

Abstract
In sample surveys, when the issue involved is sensitive or stigmatizing, many people refuse to participate. Even in cases they agree to participate, it is expected that many of them provide untruthful responses. As a consequence, the results of the survey are very much in doubt. In this review paper, for the benefit of all statistics stakeholders, in particular users, we present in an elementary way indirect questioning techniques used to overcome nonresponse and other nonsampling errors resulting in such situations. Furthermore, the issue of the protection of privacy is also presented.

Key words:
Indirect questioning, randomized response, item count technique, negative surveys, nonrandomized response techniques, protection of privacy.

[132] Tasos C. Christofides, Department of Mathematics and Statistics, University of Cyprus; email: tasos@ucy.ac.cy

Big Data Analysis: Experiences and Best Practices in Official Statistics

Monica Scannapieco, Giulio Barcaroli and Stefano De Francisci[133]

Abstract
Big data are becoming more and more important as additional data sources for Official Statistics (OS). In this paper, we focus on the analysis phase as it has been conducted on some Big Data sources in Istat projects. In particular, we will discuss: (i) the specific methodology applied to estimate mobility profiles starting from CDRs (Call Detail Records) [1]; (ii) the model based estimation methods used to exploit the time series query share extracted from Google Trends [2]: (iii) the analysis phase carried out on Web extracted texts, aimed at retrieving information about E-Commerce activities by enterprises directly from their Web sites [3].The contribution of the paper is twofold: on one side we review some concrete experiences on Big Data analysis in OS; on the other side, we provide best practices for statistics stakeholders to design the analysis phase of Big Data projects.

Key words:
Big Data, Web Scraping, Mobile Data, Data Mining, Predictive Models, Text Analysis

References

1. Furletti, B., Gabrielli, L., Garofalo,G., Giannotti,F., Milli, L., Nanni, M., Pedreschi,D., Vivio, R.: Use of mobile phone data to estimate mobility flows. urban population and inter-city mobility using big data in an integrated approach. SIS 2014.
2. Bacchini, F. , D'Alò, M., Falorsi,S., Fasulo, A., Pappalardo,A.: Does Google index improve the forecast of Italian labour market?. SIS 2014.
3. Barcaroli, G., Scannapieco, M., Nurra, A, Scarnò,M., Salamone, S., Summa, D.: Internet as Data Source in Istat Survey on ICT in Enterprises. Submitted to Austrian Journal of Statistics, 2014

[133] Monica Scannapieco, Istituto Nazionale di Statistica – Istat; email: scannapi@istat.it
Giulio Barcaroli, Istituto Nazionale di Statistica – Istat; email: barcarol@istat.it
Stefano De Francisci, Istituto Nazionale di Statistica – Istat; email: defranci@istat.it

Data Mining, Machine Learning and Official Statistics

Gilbert Saporta and Hossein Hassan[134]

Abstract
We examine the issues of applying Data mining and Machine Learning techniques to official data, and try to explain their underuse in National Statististical Institutes.

Key words:
data mining, machine learning, official statistics, NSI

[134] Gilbert Saporta, CEDRIC-CNAM; email: gilbert.saporta@cnam.fr
Hossein Hassani, Bournemouth University,; email: hhassani@bournemouth.ac.uk

Changing landscape in environmental data

Massimo Craglia and Jacopo Grazzini[135]

Abstract

This presentation reflects on the changing landscape in the production, management, and consumption of environmental data and what lessons maybe relevant in the context of official statistics. It draws on the experience as coordinators of the INSPIRE Directive, which sets up an infrastructure to share spatial environmental information in Europe, and the early results of pliots in the areas of remote and social sensing.

[135] Massimo Craglia, European Commission Joint Research Centre; email: massimo.craglia@jrc.ec.europa.eu
Jacopo Grazzini, European Commission Joint Research Centre;

The value proposition of the Official Statistics Research Community in the field of Big Data– and what Big Data research could bring to Official Statistics

Martin Karlber[136]

Abstract

As has been stated in many fora, official statistics is omnipresent in evidence-based policymaking, but for precisely that reason (not being a "policy" itself, but rather supporting various policies) not easily compatible with the thematic structure of the Horizon 2020 Research Framework Programme [1]. This paper outlines two possible strands of official statistics related research activity in the field of Big Data, showing how the official statistics research community could be a contributor to the Horizon 2020 research agenda. It has to be acknowledged that it is the role of the official statistics research community to actively bring forward the advantages of research involving official statistics experts – and research aimed at improving official statistics. No other stakeholders are likely to do it on our behalf.

Key words:

Big data, research, value proposition, official statistics, Horizon 2020, evidence-based policymaking.

References

1. http://ec.europa.eu/programmes/horizon2020/

[136] Martin Karlberg, Eurostat; email: Martin.Karlberg@ec.europa.eu (Any errors and omissions are the sole responsibility of the author. The opinions expressed in this paper are personal and do not necessarily reflect the official position of the European Commission)

The Cambridge Compromise - A Proposal for the Composition of the European Parliament

Friedrich Pukelsheim[137]

Abstract

The Cambridge Compromise is a proposal how to apportion the seats in the European Parliament among its Member States. It proposes to provide each Member State with a base of five seats, and to divide the remaining seats between the Member States in proportion to their populations observing a ninety-six seat capping. The recommended "base+prop" formula rounds fractional allocations of seats upwards, thereby guaranteeing a minimum of six seats to every Member State. Variants of the Cambridge Compromise that avoid the capping step are also discussed. They involve a non-linear downweighting of the population figures until the largest State is allocated exactly ninety-six seats. The pertinent calculations of the variants are described, and their relative constitutional merits are discussed.

References

1. .R. Grimmett, J.-F. Laslier, F. Pukelsheim, V. Ramirez-Gonzalez, R. Rose, W. Slomczynski, M. Zachariasen, K. Zyczkowski (2011): "The allocation between the EU Member States of the seats in the European Parliament." European Parliament, Directorate-General for Internal Policies, Policy Department C: Citizen's Rights and Constitutional Affairs, Note 23.03.2011 (PE 432.760).
2. G.R. Grimmett, K.-F. Oelbermann, F. Pukelsheim (2012): "A power-weighted variant of the EU27 Cambridge Compromise." Mathematical Social Sciences 63, 136-140.
3. J. Martinez-Aroza, V. Ramirez-Gonzalez (2008): "Several methods for degressively proportional allocations. A case study." Mathematical and Computer Modelling 48, 1439-1445.
4. F. Pukelsheim (2014): "Proportional Representation - Apportionment Methods and Their Applications. With a Foreword by Andrew Duff MEP." Springer International Publishing, Cham (CH) 2014.

[137] Friedrich Pukelsheim, Institute for Mathematics University of Augsburg, Germany; email: Pukelsheim@Math.Uni-Augsburg.De

Tailoring products to inform decision making: Sharing good practice from across the UK

Stephanie Howarth[138]

Abstract

This paper describes a range of successful products and initiatives used by the Government Statistical Service in the UK to maximise the impact and accessibility of official statistics and increase their influence on decision making. These include the creation of concise, engaging summary statistics for Ministers; the use of online data explorers for people that want to delve deeper in to the statistics; and infographics and interactive tools to communicate statistics to citizen decision-makers. The paper highlights the importance of understanding who uses statistics and the decisions that they influence, and how this knowledge can be used to tailor statistical products. The paper also discusses the role of social media in disseminating statistical products, to ensure that they reach the people that need them.

Key words:

official statistics, decision making, influence

[138] Stephanie Howarth, Good Practice Team, UK Statistics Authority; email: stephanie.howarth@statistics.gsi.gov.uk

Which wellbeing indicators can really influence the political arena?

Donato Speroni[139]

Abstract

"Let's scrap Gdp" , for many years, has been an ensign of political movements, with different visions, but with the common idea that the production of goods and services is not the real indicator of the well being of a community. Since 2004 (Oecd Palermo Conference) the process in order to find alternative measures of progress took speed all over the world. Specially after the Stglitz Report in 2009, it became clear that the goal was not going "beyond Gdp", but "Gdp and beyond". Statistics about the Product are useful (specially if presented in a different way, as suggested by the Stiglitz Commission) but in order to describe the effective well-being of a community, subjective and objective, we need a battery of other indicators. Squeezing these statistical series in one composite indicator like Gdp probably is not the right solution. We need a "dashboard". The Italian Bes by Istat and Cnel represents one of the most advanced experiences in this field.

For media communication and political use, however, dashboards are hard to use because they are more complex and confusing than Gdp. The media don't like too many numbers and the political decision-makers tend to present only the data that support their vision. On the other hand, the new indicators are just an academic exercise if they are not useful for improving political action. Recent projects at the European level (Brainpool, Eframe) have emphasized the need to propose not only a statistical dashboard, but a framework of indicators of well-being, on which to build a broad consensus. The main problems to solve are the coherence and comprehensibility of information to put in this framework. It should be designed to describe not only the current well-being, but also the economic, environmental, human and social effects of our individual and collective actions on the future, taking into account inequalities between social groups, intra-generational and intergenerational sustainability. In addition, to have a political impact, the new indicators should be coherent (like an econometric system in macroeconomics), globally comparable but with local breakdowns, updated and readily available in order to be useful for social policies as Gdp is fuseful or monetary and fiscal policy. All this looks like a very difficult challenge, an almost impossible "statistical revolution". However, this process might receive a big boost from the building within next year of the new Sustainable Development Goals (Sdg), valid from 2016 to 2030, by the United Nations. Without scrapping Gdp, the new statistical indicators which will substantiate the Sdgs can be an important opportunity for building a global frame of data, as announced by the Un Secretary General Ban Ki-moon.

[139] Donato Speroni, Istituto per la Formazione al Giornalismo, Urbino University, Italy; email: donatosperoni@gmail.com

New challenges in detecting learning by exporting effects on firms' productivity distribution

Maria Rosaria Ferrante, Marzia Freo[140]

Abstract

This paper investigates whether firms that move from the status of non-exporter to the status of exporter experience an increase in productivity during the period following their entry into the export market. Although an extensive stream of empirical literature on international trade shows that the most productive firms undergo a self-selection process to enter foreign markets, few and far from conclusive contributions support the assumption that the performance of firms increases after they begin to export. The learning-by-exporting (LBE) hypothesis is usually tested in the impact evaluation context by using the propensity score matching approach, that aims at estimating the average treatment effect.

This paper is a further contribution that aims at verifying the presence of the LBE effect on the total factor productivity growth. Our proposal, as usual in this context, starts by dealing with the pre-entry selection bias but it differs from the literature in several aspects.

First, we estimate the net LBE effect, where "net" refers to the premium actually imputable to the entry into international markets, by working in the perspective of the evaluation literature but simultaneously by enlarging the focus on the whole distribution on the net premium. We consider the effect on different quantiles in order to not limiting the comparison only to central tendency indicators. Second, since TFP is sensitive to cycle movements and demand pulls we estimate the LBE effect explicitly taking into account this feature. In our study the presence of economic cycle is explicitly recognised. Third, since, as usual in developed countries, the low number of firms entering in the international market each year is quite low, in order to obtain reliable statistical results we design an experiment by aligning and pooling cohorts of firms. Fourth, the evaluation of LBE needs of individual information observed for long periods so we base our analysis on a panel survey covering ten years. Since, as usual, in a so wide time window a not negligible number of firms leave the market, we also deal with the bias due to drop out. The analysis is based on a newly available panel data set of firms recently developed jointly by the Italian National Institute of Statistics. It contains firms' micro-data for the period 1998 through 2007 and it has been realised by matching different census and sample surveys of the firms. This very rich longitudinal firm-level data allows for the proper evaluation of the causal effect of exporting on the performance of firms by also dealing with the self-selection bias. Results show that the post-entry productivity growth is not present by only adjusting for the pre-entry selection bias, whereas the LBE effect arises when estimates are adjusted even for the drop out selection bias and obtained by considering business cycle effect. With regard to the heterogeneity of the LBE effect, the post-entry growth of the TFP premium is larger for firms at the top section of the distribution of the TFP premium growth. The more growing starters firms begin to accelerate as regard as domestic firms just after the entrance and as regard as incumbents after one year. Thus, for the less performing firms a positive post-entry effect may be detected at least two years, while for the more performing firms the post-entry effect may be detected just after the entry and suddenly or after one year for. As a consequence, the most performing firms show a remarkable LBE effect two years after entrance.

[140] Maria Rosaria Ferrante, Department of Statistics, University of Bologna, Italy; email: maria.ferrante@unibo.it
Marzia Freo, Department of Statistics, University of Bologna, Italy; email: marzia.freo@unibo.it

How to cope with modelling and privacy concerns? A regression model and a visualization tool for aggregated data

Rosanna Verde, Antonio Irpino and Dominique Desbois[141]

Abstract

Regulation of official statistical institutes does not allow the diffusion of microdata for privacy-related purposes. In general, it is easier to obtain aggregated data of a set of individuals. Most of the modelling tools in statistics (e.g., regression) works on microdata and cannot be easily extended to macrodata. In this paper, we show the use of a regression method developed for aggregated data, where both the explanatory and the response variables present quantile distributions as observations. The analysis has been performed on a dataset of economic indicators related to the specific cost of agriculture products in the France regions. We show how the proposed method can predict distribution values. In conclusion a PCA method on quantile data is used in order to visualize relationships between the predicted distributions.

Key words:

Macrodata, distribution-valued data, Wasserstein distance, Regression, Principal Component Analysis

[141] Rosanna Verde, Antonio Irpino, Second University of Naples, Dept. of Political Sciences "J. Monnet", Caserta, Italy;
email: {rosanna.verde} {antonio.irpino}@unina2.it.
Dominique Besbois, INRA-Économie publique - AgroParistech, Paris, France;
email: desbois@agroparistech.fr

New statistical approaches for the evaluation of healthcare structures

Salvatore Ingrassia, Giorgio Vittadini, Paolo Berta and Antonio Punzo[142]

Abstract

In recent years, increasing attention has been directed toward problems inherent to Quality Control in the Healthcare Services. Several factors have brought about this increased interest in the quality of the operational aspects of the healthcare field. The healthcare system has been going through a transition from a system where the resources were assigned to structures with a reimbursement system that was independent of the way in which they were used, to one where such allocation is distributed according to prefixed figures for each pathology. The medium amount of resources used to cure patients having a given pathology is found by means of the Diagnostic Related Group system (DRG). Therefore public health structures depend on national and regional political decisions but become autonomous from an operational point of view and characterized by competition between various healthcare structures. Hospitals are encouraged to search for an organizational model that allows the reduction of costs by optimizing the use of available resources (efficient use of resources). However the system, based on the DRG, has the disadvantage of inducing hospitals to place a higher priority on the economic aspects instead of the effective necessity of care. This phenomenon could lead to a selection of patients that are affected by a specific pathology associated with optimal clinical conditions: the evaluation is essential to find financing criteria complementary to DRG. Therefore it is also necessary to measure effectiveness of health structure, the ability to improving healthcare outcomes (Donabedian 1988) definable as the "technical result of a diagnostic procedure or specific treatment episode" (Opit 1993), a "result, often long term, on the state of patient well-being, generated by the delivery of a health service" (Aitkin & Longford 1986, Goldstein & Spiegelhalter 1996). Of particular importance is the concept of relative effectiveness, i.e. the effect of hospital care on patients used to compare different healthcare institutions in terms of "healthcare outcomes". Healthcare outcomes are influenced by covariates concerning the case mix of the patients, definable as the variability of their clinical and social demographic aspects and are related to the organization, resources, facilities and other characteristics of hospitals (Zaslavsky 2001). Therefore, to allow comparisons between healthcare institutions, relative effectiveness needs to be adjusted for patient-specific and hospital-specific variables by means of risk adjustment statistical methods. Evaluation of the healthcare structures is usually carried out through statistical approaches like multilevel models and risk adjustment methodologies. However, these approaches are not suitable when we cope with data with large heterogeneity. In particular, a potentially large source of unobserved heterogeneity resides in variation of the regression coefficients between groups of individuals sharing similar but unobserved background characteristics. In this work, we consider new statistical approaches based on Mixtures of Multilevel regression models, introduced in Muthén & Asparouhov (2009), and Cluster Weighted Models developed in Ingrassia et al. (2012), Ingrassia et al. (2014). These approaches prove to reduce such drawbacks of both multilevel models and risk adjustment methodologies.

[142] Salvatore Ingrassia, Antonio Punzo, Department of Economics and Business, University of Catania, Italy; email: {s.ingrassia} {antonio.punzo}@unict.it.
Giorgio Vittadini, Paolo Berta, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy; email: giorgio.vittadini@unimib.it, paolo.berta@crisp-org.it

Operationalising 'safe statistics': the case of linear regression

Felix Ritchie[143]

Abstract
The recent growth in research access to confidential government microdata has prompted the development of more general 'output-based statistical disclosure control' (OSDC) which go beyond tabular protection. Central to OSDC is the concept of 'safe/unsafe statistics', allowing researchers and data owners to make informed judgments about the types of research output that pose a disclosure risk. While increasingly accepted in specialist environments, in the wider community this novel approach causes some concern: how can 'safe' be unconditional? This paper therefore demonstrates the new approach using linear regression, a key research output, as an example. In doing so, the paper reconsiders the objectivity of SDC decision-making, arguing that 'safety' be explicitly acknowledged as a relative concept.

Key words:
statistical disclosure control, confidentiality, safe statistics.

[143] Felix Ritchie, University of the West of England, Bristol., UK; email: felix.ritchie@uwe.ac.uk

Exploiting TIMSS and PIRLS combined data: multivariate multilevel modelling of student achievement

Leonardo Grilli, Fulvia Pennoni, Carla Rampichini and Isabella Romeo[144]

Abstract

We propose a multivariate multilevel model for the analysis of the Italian sample of the TIMSS\&PIRLS 2011 Combined International Database on 4th grade students. The multivariate model jointly considers educational achievement on Reading, Mathematics and Science, thus allowing us to test for differential effects of the covariates on the three scores. This approach represents an advance with respect to official reports, where the three scores are analysed separately. Moreover, the multilevel approach allows us to disentangle student and contextual factors affecting achievement, also considering territorial differences in wealth. The proportion of variability at class level is relevant even after controlling for the observed factors. The residual analysis allows us to locate classes with extremely high or low effectiveness.

[144] Leonardo Grilli, University of Florence, Firenze, Italy; email: grilli@disia.unifi.it
Fulvia Pennoni, University of Milan Bicocca, Milan, Italy
Carla Rampichini, University of Florence, Firenze, Italy
Isabella Romeo, University of Milan Bicocca, Milan, Italy

The MIP scoreboard: a set of macroeconomic indicators for European Union governance

Rosa Ruggeri Cannata, Dario Buono[145]

Abstract

The Macroeconomic Imbalance Procedure (MIP) is a system for monitoring economic policies and detecting potential harms to the proper functioning of the European Union (EU). It is part of a surveillance system for budgetary and economic policies, implemented via the European Semester, the EU's policy-making calendar. An essential tool in the procedure is the MIP scoreboard - a set of eleven headline indicators intended to screen internal and external macroeconomic imbalances, covering a time span of ten years for the twenty-eighth EU Member States. It acts as a first filter in a broader process seeking to disentangle the existence and severity of macroeconomic imbalances in the EU Member States, starting every year in autumn with the Alert Mechanism Report identifying countries and issues for which a closer analysis is deemed necessary. In this process, the scoreboard is used by policy makers, together with a set of auxiliary indicators for their economic reading. Policy makers need an as complete as possible picture of the economy; it might be then necessary to apply statistical methodologies to ensure optimal data coverage, in particular when statistics are subject to events which could disrupt time series length, such as the adoption of new classifications or changes in the production process.

Considering the definitions of some of the MIP headline and auxiliary indicators (e.g. averages or percentage changes over several years), needed data coverage can reach twenty years. In this context, statisticians have to ensure the availability of the required length for time series needed for MIP indicators and, where necessary, to apply statistical techniques such as backcalculation for this purpose.

Section 1 of this paper introduces the main characteristics of the MIP, and its role within the context of the economic governance; section 2 introduces some relevant quality issues related to the MIP headline indicators; section 3 illustrates examples of statistical techniques applied to improve data coverage of MIP headline indicators in order to fulfil policymakers' needs; finally, section 4 contains some concluding remarks and directions for future work.

Key words:

economic governance, policy decision, time series

[145] Rosa Ruggeri Cannata, Eurostat, email: rosa.ruggeri-cannata@ec.europa.eu
Dario Buono, Eurostat, email: dario.buono@ec.europa.eu

The Macroeconomic Imbalance Procedure: statistics from a user perspective

Jonas Fischer[146]

Abstract
The Macroeconomic Imbalance Procedure (MIP) aims to strengthen surveillance and prevention of the building up of external and internal imbalances. While it's broad design is similar to the one of the EU fiscal surveillance framework (with a preventive and corrective arm), the role of statistics is different in that it is not used as an automatic trigger for policy action in a similar way. Nevertheless, the statistical use and needs are large. In its first stage, through an economic reading of a scoreboard of indicators, with indicative alert thresholds, countries are identified for which a more in-depth economic analysis on imbalances and risks is warranted. This economic reading follows a number of economically based broad principles and there is no automaticity. In the more analytical stage of the ensuing in-depth reviews, the basis of which policy action is decided, the analysis draws on all available and relevant statistics and information. To this end the IDR analyses is supported by an ambitious analytical agenda where the availability of the needed statistics are key for further advancement.

Key words:
economic governance, policy decision, indicators

[146] Jonas Fischer, DG ECFIN, email: Jonas.fischer@ec.europa.eu

Journalistic information deals with statistical data, economic government and the observance of the fiscal compact

Kevin Costelloe[147]

Abstract

[147] Western Europe Economy Editor Bloomberg News, Rome

National Reform Program: Italian evidence

Paolo Reboani[148]

Abstract

[148] Paolo Reboani, President of Italia lavoro

A new method for synthesizing multidimensional measures – an application on comparing wellbeing patterns in Italian society

Marco Fattore[149]

Abstract
In this study, we introduce an innovative methodology for synthesizing multidimensional multidimensional measures, based on ordinal data. We apply the methodology to data about subjective wellbeing in Italy, for years 2007 and 2010. The analysis reveals the existence of different levels across the Country and between males and females. Interestingly, the temporal dynamics of subjective suffering suggests that people living in economically more developed regions worsen their self-perception, across the beginning of the economic crisis.

Keywords
Subjective well-being, Quality of life, Italy, Partially ordered set.

References

1. Alkire. S., & Foster J. (2011) "Counting and multidimensional poverty measures", Journal of Public Economics. 95, 476 – 487.
2. Davey, B. A., & Priestley, B. H. (2002) Introduction to lattices and order, Cambridge: Cambridge University Press.
3. Diener, E. (1984) "Subjective Wellbeing" in Psychological Bullettin, 95.
4. Diener, E. (2000) Subjective Well-Being. The Science of Happiness and a Proposal for a National Index. American Psychologist, 55, 34-43.
5. Fattore, M. (2014) "Partially Ordered Set" (entry title), in: Michalos AC (Ed.) Encyclopedia of Quality of Life and Well-Being Research. Springer Dordrecht, Netherlands: Springer, pp 4627-4631.
6. Fattore, M., & Arcagni, A. (2014) "PARSEC: an R package for poset-based evaluation of multidimensional poverty", in: Bruggemann R. Carlsen L., Wittmann J. (Eds.) Multi-Indicator Systems and Modelling in Partial Order, Springer, Berlin.
7. Fattore, M., & Maggino, F. (2014) "Partial orders in socio-economics: a practical challenge for poset theorists or a cultural challenge for social scientists?", in: Bruggemann R. Carlsen L., Wittmann J. (Eds.) Multi-Indicator Systems and Modelling in Partial Order, Springer, Berlin.

[149] Marco Fattore, University Milan-Bicocca, Italy; email: marco.fattore@unimib.it

Complex Phenomena Measured on Individuals: The Challenge of Mixed Data for Composite Indicators

Giovanna Boccuzzo and Giulio Caperna[150]

Abstract

The concept that features Composite Indicators (CI) is traditionally connected to the use they had at the beginning of their diffusion: some characteristics of a country or a big institution (i.e. university,...) summarized together to obtain a unique score useful for comparison, evaluation and policy making. Nowadays the reputation of CI has increased and several attempts to measure complex phenomena at individual level have been proposed. Fundamental guidelines for CI construction are still serving well, anyway this evolution requires a deep change in the scientific approach to CI. Indeed, since the unit of interest is the individual, elementary indicators could be often ordinal or dichotomous variables, especially in case of subjective indicators. This is not the case of CI referred to countries, given that elementary indicators are percentages or rates. Therefore, the aim of this work is to discuss the recent advances in the use of dichotomous and ordinal variables for the construction of composite indicators. We will introduce some of the most interesting approaches developed in the last twenty years on the use of qualitative or mixed variables, pointing to the new criteria adopted for overcoming the classic linear and compensative approach. Starting from the proposals more similar to the classic approach to CI, we will discuss several methods, some of them that completely changed the traditional theory of CI, avoiding linear aggregation. First of all the Benefit of the Doubt method, which bases its roots in the linear programming tool called Data Envelopment Analysis (DEA), and enables to overcome the debate on standardization and weighting rules. It has been extended to deal with mixed data. The Multicriteria approach (MCA) includes a set of wide and flexible methods, that permit a large number of different multi-criterion aggregation procedures (MCAP) There is a large number of schemes for MCAP, useful for quantitative or mixed variables. Partial Least Squares (PLS) path modeling and LISREL methods allow to compute CI as latent variables of a Structural Equation Model. Some recent formulations allows to find an estimation in case of qualitative or mixed data. Finally, the Partial Order Set (POSET) method will be discussed as a completely new tool to describe a complex concept refusing the step of aggregation, that is replaced by an evaluation procedure. We will discuss and compare these methods, and draw attention to their strengths and weaknesses.

Key words:
composite indicator, mixed data, non-compensative approaches

[150] Giovanna Boccuzzo, Department of Statistical Science, University of Padova;
email: Giovanna.boccuzzo@unipd.it
Giulio Caperna, Department of Statistical Science, University of Padova; caperna@stat.unipd.it:

Evolution of the Bes project

Saverio Gazzelloni, Adolfo Morrone, Fabiola Riccardini[151]

Abstract
"Equitable and Sustainable Well-being (BES)" initiative aims at analysing the levels, time trends, distribution of the various components of well-being and sustainability over time and space of well-being. Two points are going to be developed of the BES project. The first line of development at present is the initiative facing the challenge to build a composite indicator for each of the 12 Bes domains with the aim to provide a quick outlook of the main trends. The composite indicators will be of paramount importance to enhance the communicability of the Bes and to stir the political debate on Italian well-being. he second line of development is related to sustainability. In Italy we have approached the measuring of sustainability in the context of BES initiative, in line with the United Nations 2013 declaration in which is affirmed that there is sustainable development when well-being of people is pursued over time. The main idea for sustainability measurement is to define a system of measures that give evidences of the interrelations among BES domains during a certain period and for a defined space, interrelated with other spaces, and verified if the level of well-being achieved at least can be maintained over time. The methods selection was focused mainly on observations and forward-looking models.

Key words:
Bes, composite indicators, sustainability of wellbeing, sustainability over time and space, complexity, vulnerability, resilience, forward-looking models

[151] Saverio Gazzelloni, Istat, Italy; email: saverio.gazzelloni@istat.it

Estimating the size of the Romanian shadow economy using the mimic approach

Adriana AnaMaria Davidescu(Alexandru)[152]

Abstract
This paper aims to estimate the size of the Romanian shadow economy (SE) using a Structural Equation Approach using quarterly data for the period 2000-2013. The shadow economy is modelled like a latent variable using the MIMIC model, revealing that the best model is the model with four causal variables (unemployment rate, self-employment, government employment and 12 months real interest rate) and two indicators (index of real GDP and currency ratio M1/M2). The shadow economy measured as percentage of official GDP records the value of 40% in the beggining of 2000 and follows an descendent trend until 2008. For the last years, the size of the unreported economy oscillates around the value of 28%-29% of official GDP.

Key words:
shadow economy, MIMIC models, Romania, Lisrel

References

1. Dell'Anno, R., Solomon, O.H. Shadow economy and unemployment rate in USA: is there a structural relationship? An empirical analysis, Applied Economics, 1-19,(2007).
2. Schneider, F., Buehn, A., Montenegro, C. Shadow Economies All Over The World: New Estimates For 162 Countries From 1999 To 2007, Working Papers wp322, University of Chile, Department of Economic, (2010).

[152] Adriana AnaMaria Davidescu(Alexandru) Department of Statistics and Econometrics, Bucharest University of Economic Studies, National Scientific Research Institute for Labour and Social Protection, Bucharest, Romania, adrianaalexandru@yahoo.com

National stakeholders for the European Master in Official Statistics – the Italian case

Maria Pia Sorvillo and Nereo Zamaro[153]

Abstract
Quality in the production of Official Statistics relies in the National statistical system, and the whole network of organizations being part of it can benefit from key modernization initiatives such as the Eurostat VIP project EMOS (European Master in Official Statistics). To be fully effective, EMOS has to fit in the training and recruitment policy of the NSI, integrating in the framework of present opportunities and constraints. But it could be highly beneficial also for the NSS at large. A discussion about the NSS' present staff characteristics and its training needs - in the light of two extensive fact-finding surveys carried out by Istat – shows that in several cases statistical competencies should be reinforced to ensure the production of high quality statistical information.

Key words:
European Master, Official Statistics, National Statistical System,
training needs, competencies.

[153] Maria Pia Sorvillo, Istat; email: sorvillo@istat.it
Nereo Zamaro, Istat; email: zamaro@istat.it

Statistics as a new Skill for new Jobs

Barbara Ascari, Francesco Michele Mortati and Marina Peci[154]

Abstract
Workers in the 21st century must have a stock of information-processing skills (OCSE, 2013). Statistics is the main competence for this purpose and must be considered as a key skill for employment and social inclusion. For this reason Istat (Italian National Institute of Statistics) has been actively engaging in strongly disseminating statistical literacy towards civil society. According to this mission, the targets are represented by public administration, media, the world of education, students and teachers and by the citizens in generally speaking.

Key words:
statistical literacy, ICT, education, e-learnig, data journalism, competences.

[154] Barbara Ascari, (expert), Italian National Institute of Statistics (ISTAT), Italy; email: ascari@istat.it
Francesco Michele Mortati, (expert), Italian National Institute of Statistics (ISTAT), Italy;
email: mortati@istat.it
Marina Peci, (senior researcher), Italian National Institute of Statistics (ISTAT), Italy; email: peci@istat.it

Experiences and Suggestions for Lifelong Learning in Permanent Census

Antonella Bianchino, Giulia De Candia, and Stefania Taralli[155]

Abstract
In censuses round 2010-2011 Italian National Institute of Statistics introduced many innovations also in training field. Traditional classroom training was supported by new products and services in e-learning. This paper describes the lessons learned from this experience and proposes same ideas for the next demographic census, the characteristics of which require the design of a plan for lifelong learning.

Key words:
lifelong learning, e-learning, u-learning, censuses, survey network

[155] Antonella Bianchino, Istat – office for Basilicata; email: bianchin@istat.it
Giulia De Candia, Istat – office for Liguria; email: decandia@istat.it
Stefania Taralli, Istat – office for Marche; email: taralli@istat.it

Measuring university attractiveness in Italy. An analysis of the determinants of students' mobility

Francesca Giambona, Mariano Porcu, and Isabella Sulis[156]

Abstract
A central question for education authorities has become "which factors make a university attractive?". This contribution makes an attempt to identify the most important factors that could affect student mobility in Italy by completing a macro-perspective analysis of competing university territorial areas. The Bradley-Terry modelling approach based on pair comparisons has been adopted to define the attractiveness of competing universities and assess how much the detected divergences can be attributed to university policies and how much is outside the control of the universities' ruling bodies

Keywords:
student mobility, Bradley-Terry Model, university attractiveness, League Tab

[156] Francesca Giambona, Università di Cagliari, email: fgiambona78@gmail.com
Mariano Porcu, Università di Cagliari, email: mrporcu@gmail.com
Isabella Sulis, Università di Cagliari, email: isulis@gmail.com

AUTHORS INDEX