Accepted Posters

Poster Session and Mixer, June 11, 2015, at 5-7 p.m. and
Friday 12, 9.45-10.15 (during Coffee Break)

Scandic Hotel Hasselbacken, Stockholm, Sweden.

Poster title, abstract and submitting author and/or presenter

---

Missing Data and Multiple Imputation: New Analyses With Old Methods
Murray Aitkin, University of Melbourne, Australia

This poster describes a new analysis method for accommodating randomly missing data on covariates in generalised regression models, by combining the Data Augmentation algorithm (Tanner and Wong 1987) with a nonparametric (multinomial) distribution for the incomplete covariates. The method uses the MCMC component of multiple imputation, but does not require a rich imputation model or combined analyses on multiply imputed data sets. Instead it obtains the posterior distribution of the regression model parameters directly through MCMC with the nonparametric distribution for the covariates. The approach can be applied to GLMs and to GLMMs, accounting for randomly missing covariate data at all levels of the design. It can be further extended with a nonparametric distribution for the response variable, providing a fully efficient but not model-dependent analysis for a wide range of non-informative designs. Simulation studies in small normal regression models show that the method is superior to both complete case analysis and multiple imputation using the mvn option in Stata.

Quality, Analytic Potential and Accessibility of Linked Administrative, Survey and Publicly Available Data
Manfred Antoni, Institute for Employment Research (IAB), Germany

Longitudinal data continuously gain in importance especially for life course studies. However, surveys increasingly face the problem of unit-nonresponse due to increasing data protection concerns and panel attrition or declining reachability and cooperation of respondents. Quality issues arise with item-nonresponse or misreporting, especially when recall error in retrospective interviews occurs. Particularly longitudinal interviews lead to high costs and response burden.

One potential remedy for quality and costs issues is the linkage with administrative or open data. Their purpose of their collection may initially have been other than creating research data but they usually offer precise and reliable information covering long periods of time. Data linkage thus potentially results in higher cost efficiency and data quality. Linked data also provide higher analytic potential for substantive analyses than their separate parts, either by combining their sets of variables, by adding observational levels (e.g. employees within establishments within companies) or following respondents after their panel drop out.

Moreover, research on the quality of either data source gets possible by applying validation, unit- or item-nonresponse analyses or by examining the selectivity of consent to and success of record linkage.

Our presentation will focus on the potential, quality and accessibility of linked data of the Research Data Centre of the German Federal Employment Agency. On the one hand they comprise administrative data on a daily basis with exact information on the income and receipt of benefit since 1975. On the other hand these data can be linked to several survey data sets on individuals, households or establishments which focus on fields of research like poverty, education, intergenerational mobility or lifelong learning.

Interviewers' Influence on Bias in Reported Income

Manfred Antoni, Institute for Employment Research (IAB), Germany

In this paper we investigate characteristics of respondents and interviewers influencing the accurateness of reported income by comparing survey data with administrative data. Questions on sensitive topics like respondents' income often produce relatively high rates of item nonresponse or measurement error. In this context several analyses have been done on item nonresponse (e.g. Essig/Winter 2009), but little is known about misreporting. Existing evidence shows that it is unpleasant for respondents to report very low or very high income. The generally observed high rates of item nonresponse at both tails of the income distribution support this hypothesis (Riphahn/Serfling 2005).

One possible explanation of such misreporting is social desirability bias, which may lead to overreporting of desirable attributes or underreporting of undesirable ones in order to present oneself in a positive light (Stocké/Hunkler 2007). Because an experienced and competent interviewer may be able to inhibit such behavior, interviewer characteristics as well as their interaction with respondent characteristics should be of particular importance. Moreover, the bias should decrease with perceived social closeness between respondent and interviewer (Diekmann 2008).

Using linked survey and administrative data we are able to measure the extent of deviation between reported and recorded incomes and explore the influence of respondent and interviewer characteristics on it. The starting point for the linkage is data from the German National Educational Panel Study (NEPS), Starting Cohort 6 (see Allmendinger et al. 2011), which surveys adults from birth cohorts 1944 to 1986. More than 90% of the respondents consented to a linkage of their survey information with administrative data from the German Federal Employment Agency. These longitudinal earnings data are highly reliable as they are based on mandatory notifications of employers to the social security system (Jacobebbinghaus/Seth 2007).

We include interviewer and respondent characteristics as well as their interactions into our model to estimate their respective impact on the incidence and size of any bias in reported incomes. This allows us to control for latent interviewer traits that might have influenced the respondent's answering behavior during each interview of a given interviewer.

Sampling Design and Cross-sectional Weighting for the Survey of Household Finances and Living Conditions in Korea

Jeeseon Baek, Statistics Korea, South Korea

Since 2012, Korean Survey of Household Finances and Living Condition (KSHFLC) has been conducted by Statistics Korea to provide annual family income, assets, and liabilities in Korea. The KSHFLC is a panel survey of sample size 20,000 dwellings. To reduce the response burden and also to produce more

precise estimates, Statistics Korea decided to change the sample design into a rotation scheme consisting of five panels from 2015. A rotation group remains in the survey for 5 years, after which it is replaced by a new one. In this paper, we discuss the sampling design and weighting method for the new KSHFLC. In the sampling design, the original sample is partitioned to maximize the homogeneity between the rotation groups. In the weighting stage, the design weights are first computed to obtain unbiased estimation and then further adjusted to improve the efficiency of the resulting estimator. We use a novel application of the EM algorithm to obtain an efficient weighting. Some results from a numerical study are presented.

Administrative Data for the Redesign of Surveys and Reorganization of the Data Collection Process

Silvia Biffignandi, University of Bergamo, Italy

The use of administrative data for statistical purposes has brought the NSIs to the creation of a set of statistical registers. For this reason, there is now a growing interest in planning register based surveys and in revisiting and reorganizing the existing ones (Wallgren and Wallgren, 2007; Vale, 2013). This poster focuses on how statistical registers, based upon available administrative data, can be used in the framework of Business Statistics to restructure and redesign surveys. With reference to statistics on businesses in Italy, it is investigated how surveys can benefit from administrative data and from their integration with the survey data. The study relies on changes with reference to variables to be collected and to the improvement of the data collection process, such as use of the available information for designing optimal survey strategies (Falorsi, 2012) or in an adaptive design framework (Groves and Heeringa, 2006; Luiten and Schouten, 2013; Lundquist and Särndal, 2013). Examples, experiments and related methodological aspects are presented in the poster with reference to business statistics data collected in the context of Official Statistics in Italy (Istat). In particular, R&D survey and Business statistics surveys are considered to investigate potentials of statistical registers for the redesign of surveys and the implementation of adaptive design schemes. Web as a data collection tool is taken into account as a further aspect of innovation in the review process.

Longitudinal Business Data: Two Construction Approaches for Economic Analysis

Silvia Biffignandi, University of Bergamo, Italy

Reducing the response burden and widening available statistical information require new approaches in the NSI production process. Our study focuses on longitudinal data needs for economic analysis and on how to obtain them without increasing response burden. Two construction approaches in building experimental longitudinal database are described and their quality evaluated. Various methods for the longitudinal data generation can be adopted: 1) repeated cross section by ignoring tracking individuals over time. This approach may be useful for understanding aggregate changes in a variable; 2) individuals may be tracked over time. If the interest is in studying the time-varying economic characteristic (like individuals employment) this approach is much more informative. So the two databases were built in Istat using different approaches: the first one (called IDB)comes out from the use of a repeated cross-section information (approach 1). The second (called PANEL) is based on tracking individuals over time (approach mentioned above as 2). The databases are obtained through sources integration (different surveys, register and administrative databases). Similarity between two databases and actual population is evaluated with reference to the distribution of the main economic variables. Rank correlation, and the Fligner-Policello test are applied for similarity comparison. No relevant differences are found. In addition, representativeness R indicators are computed for the PANEL db. Rindicators measure deviations from representativeness, are useful in supporting a comparison of different surveys or registers. The value of R indicator is quite high in each year. So it is possible to be fairly confident that the level of representativeness is appropriate in a global context i.e no differences are found between distributions.

Creating a Model for a Central Metadata Repository

Klas Blomqvist, Statistics Sweden, Sweden

Statistics Sweden uses a process based organisation according to the GSBPM structure. We focus on harmonising the basis for statistical production by providing common tools and methods.

Statistics Sweden has adopted a general strategy towards the use of registers and data warehousing technology. A common metadata repository plays a central role in providing an overview of data in the envisioned environment.

Today micro data are stored and managed locally adjacent to each statistical product, whilst in future they will be stored in data warehouses. Metadata will enable users to use metadata to connect micro and macro data to the statistics being produced, i.e. through all the stages of the GSBPM.

Externally, in a changing society NSI:s face an increasing demand on current pre-compiled information sets as well as on new combinations of information, which may make current statistical products less relevant. Internally, the development of a well-structured data warehouse involves a number of production related issues: efficiency improvements in terms of access and fewer data transformations, flexible data extraction and less duplicate data. Quality improvements can be achieved in terms of reduced risk of errors, improved consistency, standardization of populations and variables, traceability and documentation.

Currently the project is in its initial implementation phase, focusing on activities directed towards well-structured, good quality metadata in order to identify relevant variables, populations and objects in the data store. The development of a GSIM based conceptual model for a central metadata repository covering GSBPM has been a key step.

Agreement Between Observation and Theoretical Model: Anderson Darling Statistics

Sorana D. Bolboaca, Technical University of Cluj-Napoca, Romania

Application of any statistical test is done under certain assumptions and violation of these assumptions led to misleading interpretations and unreliable results. Anderson-Darling test is one statistical test used to assess the distribution of data (H0: Data follow the specified distribution vs. HA: Data do not follow the specified distribution). The interpretation of the Anderson-Darling test is done by comparison of statistic with the critical value for a certain significance level (e.g. 20%, 10%, 5% and 1%). Our study aimed to identify, assess and implement an explicit function for p-value associated to Anderson-Darling statistic able to take into consideration the value of the statistics and the sample size. A Monte Carlo simulation study was conducted to determine the function able to estimate the probability associated to Anderson-Darling statistics (AD). The following algorithm was used: Step 1. Generate data sets of samples sizes from 2 to 46 using uniform continuous distribution [0,1]; Step 2. Repeat Step 1 for k×u times (where k=number of thresholds, e.g. 1,000; u=number of repetitions); Step 3. Compute AD statistics for each replication in Step 2 and for each sample size; Step 4. Order the AD value obtained in Step 3, and select a list of k values corresponding to success rates; Step 5. Identify that function which best fit according to both sample size and the AD value. Two equations has been obtained, one to be applied for the value of AD lower than a threshold line and the other one to be applied if the AD is higher than or equal with the value from the threshold line. The calculation was implemented online and is available at http://l.academicdirect.org/Statistics/tests/AD/. The input data are the sample size ($2 \leq n \leq 1000$) and the value of Anderson-Darling statistics ($0.1 \leq AD \leq 10$) while the output is represented by the probability to be observed a better agreement between the observed sample and the hypothetical distribution being tested (producing reliable output when $1.0\times10\text{-}11 \leq \min(p,1\text{-}p)$).

A Modification to the Sampling Frame of the Survey of Consumer Finances

Jesse Bricker, Federal Reserve Board, U.S.A

Family wealth in the United States is concentrated at the top of the distribution. The Survey of Consumer Finances (SCF) has overcome this data collection challenge by selecting an oversample of wealthy families to supplement the nationally-representative area-probability sample of families. The sampling frame is a set of administrative records derived from tax returns; models that relate income to wealth are used to predict household-level wealth from these data. This prediction is then the basis for sorting wealthy families into several high-end strata from which the over-sample is drawn. The most timely frame data are used to select the oversample. However, these frame data are usually not available until shortly before the SCF is set to enter the field and this has led to delays in closing the SCF field period.

In this paper, we propose shifting the sampling frame one year back and explore the costs and benefits of such a shift. The main expected benefit is that the SCF will be completed on-time. The most recent frame data are released to us as we begin locating and contacting for the area-probability sample. The oversampled families are much harder to locate and contact. Currently, contacting and locating the oversampled families often begins many months into the field period; the amount of time needed to locate and convince these families delays the completion of the SCF.

The main cost of this shift is using frame data that may be outdated. This cost is mitigated if the distribution of family wealth at the top of the distribution, where the oversample occurs, does not change year-to-year. Using a panel of frame data, we will explore the year-to-year variability of wealth rankings in the frame data. Overall, these rankings are stable at the top of the wealth distribution, with a spearman rank correlation near 0.95 for these families. We also propose a potential adjustment to the survey weights that incorporates the most recent frame data.

Questionnaire Re-Design: An Experiment in Business Innovation and R&D Surveys

Anneleen Bruylant, Belspo, Belgium

This paper examines how re-designing business surveys can improve the overall response rate, as well as the quality of the answers. While there is abundant literature to be found on designing household surveys, there is need for more investigative work concerning business survey design. As most business surveys are conducted by government agencies - or for the purpose of making policy decisions - and since business surveys most often look like unwieldy forms (Dillman, 2009), it is important to assess the efficiency of re-design as a tool to improve the quality of measurement and the representativeness of the collected data, especially for non-compulsory surveys. Based on Belgian data from the biennial Community Innovation Survey (CIS) and Research and Development survey, both official but non-compulsory surveys, this paper analyses the effect of a change in certain survey design aspects, such as question wording, the order of key questions (using a split ballot design), pre-filling for certain basic economic data, a change in cover letter, and a complete overhaul of the survey layout. The objectives of our experiments were twofold: a significant increase in response rates and better quality of the data. The former would be measured by comparing both unit response rates and item response rates with those obtained in the previous version of the questionnaire, the latter by measuring the amount of data-editing necessary. This paper presents different sets of results showing a significant impact of re-design, which is consequently investigated according to the size, age and technology level of the firms. In particular, larger companies are more sensitive to re-design than very small and young firms. The latter, in general, present lower response rates and their answers need more correcting. Finally, based on these results, this paper provides some suggestions for re-designing establishment surveys.

An Estimator for the Growth Rates in Short-Term Business Statistics Using Calibration

Diego Chianella, ISTAT, Italy

Two different methods exist to estimate the growth rate for economic variables in the short-term business statistics domain when sample data not completely overlapping are used (*Knottnerus, 2012*). The first method consists in the use of an *estimator for variation* ($\widehat{G_{OLP}}$), that is based on enterprises observed on both occasions. The second one uses the estimated totals on both occasions (*estimator for expansion* $\widehat{G_{STN}}$). In 2011 when Istat launched new surveys to increase the indices coverage for the quarterly turnover in the service sector (*Bacchini et al, 2015*), an analysis of the methodology for the calculation of the indices was carried out. It was introduced an estimator for expansion $\widehat{G_{STN}}$, instead of the *variation estimator* $\widehat{G_{OLP}}$ used for other economic sectors. In particular, the estimates of the total turnover were calculated using a calibration estimator, that corrects the initial sample weights using an auxiliary variable to account for non-response. Then the index number was built comparing the estimated total on current quarterly with the mean in the base year. Because a part of the sample is annually updated, these estimates are performed on two different samples (one at t-time and one at base year) with only a part of the companies overlapping. It was realized that this method is unstable in presence of *business demography* and *delays in reference population register*. For this reason, an *intermediate estimator* between the two described ($\widehat{G_{INT}}$) was implemented. Each quarter an estimate for the total of the current quarter and one for the same quarter of previous year are built. These estimates are calculated for both periods on the same sample at time *t*. We conducted analysis to evaluate the relative efficiency of the new estimator and to disseminate confidence intervals of the trend variation. Moreover an application to real data has been carried out for a comparison with the old estimator $\widehat{G_{OLP}}$ obtaining very similar results.

Statistics Without Borders: Can Pro Bono Statistical Consulting Meet Real Needs of Official Statistics and Para-official Statistics?

Asaph Young Chun, US Census Bureau, ASA Statistics Without Borders, U.S.A

Statistics Without Borders (SWB), an official outreach group of the American Statistical Association, provides pro bono survey statistics consulting to organizations across the world that serve disadvantaged people, particularly those non-profit organizations and governmental agencies in developing countries which do not have the resources for such services. SWB has grown rapidly from a handful of volunteers in 2008 to over 1,500 volunteer members today joining from diverse groups across the globe. The purpose of this poster is to present how it started, what impacts SWB has made to provide timely and nimble statistical support to governmental and non-governmental organizations, and what challenges remain to provide statistical information useful and relevant to stakeholders. Our poster will feature select survey projects of SWB relevant to official statistics and para-official statistics, demonstrating agile survey practices that SWB volunteers have applied in the past several years to help save many people's lives or help improve quality of life of disadvantaged people particularly in developing countries.

Accounting for Implicit Geographic Stratification in Census Microdata: An Analysis of Census Microdata Samples from IPUMS International

Lara Cleveland, University of Minnesota, U.S.A

The IPUMS International comprises the world's largest collection of integrated, harmonized, public use census microdata samples in the world. In partnership with national statistics offices throughout the world, the IPUMS International project makes these data samples available through a user-friendly web dissemination system. Approved researchers can create and download customized pooled data extracts

containing only the variables they require for research from single samples or as pooled multi-years and multi-country data files. Like most census microdata, IPUMS samples contain individual level data, clustered by household, and often stratified and differentially weighted. However, researchers commonly apply methods designed for simple random samples.

In some cases, the IPUMS samples incorporate explicit stratification by such factors as household size or geographic location. More often, however, the data are implicitly geographically stratified as a result of the sample design. Many IPUMS-International samples are systematic random samples, typically drawn by selecting every tenth household in the source file after designating a random starting point. Since the data are often sorted according to low-level geographical information, the sample retains implicit geographic stratification.

Using full count census data from four countries, we evaluate the impact of sample design on standard error estimates of microdata samples from the IPUMS International. We compare replicate standard error estimates from the full count data files to estimates from the 10% public use samples using 3 methods: subsample replicate, Taylor series linearization incorporating a new geographic pseudostrata variable, and estimates using simple random sample assumptions. We suggest that, for some types of analyses, especially those involving highly clustered or stratified variables or for analysis done on subsets of the population, researchers should consider adjusting for such influences in their estimation procedures. We further describe our work creating a pseudo-strata variable developed to help researchers incorporate information about the implicit geographic stratification resulting from the sample design in their analytic strategies.

Reliability and Stability of Survey Attitudes

Edith de Leeuw, Utrecht University, The Netherlands

A high response of the sample units approached is one of the cornerstones of survey research (Groves, 1989) and the growing nonresponse has been a constant worry of survey statisticians all over the world (De Leeuw & De Heer, 2002). Several theories on why nonresponse occurs have been developed over the years (Stoop, 2005) and survey climate and attitudes towards surveys are key concepts in these theories (Loosveldt and Storms, 2008).

To measure survey attitude a brief nine-question scale was developed for use in official statistics and (methodological) survey research. This survey attitude scale is based on earlier work by multiple authors (Cialdini (1991), Goyder (1986), Singer (1998), Stocke (2006), and Rogelberg et al (2001). Key dimensions in the development of this scale were survey value (value ascribed to surveys), survey enjoyment (reflecting the assumption that respondents do like participating in surveys), and survey burden (reflecting that too many surveys are being undertaken, increasing respondent burden).

Preliminary research with this scale in a Dutch online survey (CenTERpanel) and a German telephone (PPSM) survey indicate the survey attitude scale is reliable and has predictive validity. Survey attitude is weakly related to survey nonresponse in the expected direction (more enjoyment, high value, and less burden is associated with less nonresponse). Although these results are promising there is one potential threat to the usefulness of the survey attitude scale in nonresponse studies. Loosveldt & Storms (2008) question the permanency of someone's survey attitude and the intention to participate. They argue that these kind of opinions are an element of the (heuristic) decision making process, which is situational.

In this study we investigated whether survey attitude and its three subdimensions is stable over time, using structural equation modeling (SEM) with state-trait models (Kenny & Zautra, 2001). Data of six consecutive waves of the LISS-panel are used. The LISS-panel is a probability-based online panel of the Dutch population. The results of the analyses show support for the theoretical assumption that this scale measures a trait-like concept.

SC.A.F.FOLDING STATISTICAL PARADATA: the "Call Attempts Form" Revolution

Alessandra Federici, ISTAT, Italy

In the pluriannual Program "Stat2015", designed by the Italian National Institute of Statistics (ISTAT) following both the European Recommendations and the Wiesbaden Memorandum on the renewal of social statistics, the innovation in data collection's techniques constitutes the core and crucial point.

This ambitious program of renewal involves the reengineering, standardization and industrialization of processes, as well as the integration between different sources (sample surveys and administrative sources) and products' innovation.

Fundamental socio-economical ISTAT surveys have therefore been recently involved in deep methodological changes as far as data capture techniques are concerned, abandoning the PAPI technique ("*Paper and Pencil Interview*") towards the CAPI technique ("*Computer Assisted Personal Interview"*), or the more economic CAWI (*Computer-assisted web interview*) and CATI (*Computer Assisted Telephone Interview*) techniques.

Starting from this global methodological (r)evolution, this poster concentrates on the crucial role of the "*Call Attempts Form*" (CAF) in guaranteeing data quality. In fact, thanks to computer-assisted techniques of interview, the registration of all the provisional and definitive outcomes of contacts/non contacts with sampled families, allows a detailed analysis of the delicate and problematic key elements propedeutic to interview and influencing its realization:

a) the reception of the ISTAT President's advance letter, which informs families about the aims of the survey and the guarantee of anonymity;
b) the number of call attempts with the family;
c) the number of contacts and non contacts;
d) the completed interview;
e) the appointment code;
f) busy telephone number;
g) the final coding (response and nonresponse coding)
h) the reasons which determine "nonresponses" or "partial responses", both before the starting of the interview and during the interview;
i) close/terminate sample unit;
j) the response rates;
k) the substitution rates.

Previously, with the PAPI technique, this information extremely precious for data quality, would be both partially collected that analyzed at the end of the data collection, not allowing a field/survey continuous monitoring. Thus, the highlighted surveys' weaknesses would be not corrected during the survey, but faced only during the planning of the following waves.

Thanks to this poster, it would be possible to easily understand how the renewal of the "*Call Attempts Form*" in computer-assisted surveys is expected to allow a direct and timely control on statistical data production, through the collection of complex and rich survey's paradata.

Automatic Coding of Occupation Using Spell Checking and Machine Learning

Michael Franzén, Statistics Sweden, Sweden

In survey data, it is often of interest to categorise objects into different groups defined by a standardised classification. One such classification is the socio-economic classification (SEI) coded by occupation. The data is coded using a dictionary (automated coding using dictionary). If an object cannot be coded automatically it is coded manually by personnel using a reference file (index) and other information that will help to classify the object. Statistic Sweden's currently used algorithm encodes about 50% of the survey objects and is based on the occupation specified in the questionnaire being listed in the dictionary; if the occupation is not listed, the object is not coded (exact matching).

The process to manually code objects provides data of higher quality, but is expensive. It is therefore of interest to maximise the number of automatically coded objects and in the same time keep a certain level of quality. In order to do so, we study two other methods for automatic coding, spelling correction and machine learning. The spell checking algorithm corrects the spelling of the occupation, using Levenshtein distance, and then compare against the dictionary. The machine learning algorithm represents the survey objects using a bag-of-words model. This model is used to train a Support Vector Machine. Since the problem involves multiple classes, a One-vs-all strategy is used.

In this study, these methods are compared and their performance is compared to the currently used method. We conclude that the best performance is obtained by a combination of these three algorithms. When the algorithms are combined, they complement each other in a satisfying way. We note that both the spell checking algorithm and the machine learning algorithm improve coding performance, and compared to the original algorithm, the number of coded objects increases by up to 10 percent with an acceptable decrease in coding quality.

The Official Personalised Pension Information and the Projection of Future Pension Incomes

Dina Frommert, DRV Bund, Germany

Currently, the projection of future pension incomes enjoys a lot of interest from social policy makers. Recent reforms of old age security systems created a need for information on future entitlements which is not easily met by surveys or official statistics. A cheap way of gathering up to date information on current and future entitlements might be to ask for the data provided by the yearly personalised pension information. Most people know about the official information letters and asking for this concrete data might reduce the risk of respondents making a more or less informed guess during a survey. But there is a drawback in using the data as some kind of projection of future entitlements: There is no way of determining how close the current information might be to the actual future entitlements. The proposed poster will present calculations on completed work histories of pensioners to shed some light on this problem. Virtual pension forecasts are simulated at different times of the life course and are then compared to the actual pension entitlements accrued over the lifetime. In this way it is possible to determine – at least for the past – how adequate the pension information would have been. The analysis uses data provided by the Research Data Centre of the German Federal Pension Insurance (Scientific Use File Vollendete Versichertenleben SUF VVL 2010). The longitudinal dataset provides monthly data on the work histories of pensioners and can be used to simulate the forecasts presented in the personalised pension information. By using information on today's pensioners and simulating virtual pension information data, it is possible to show in how many cases and from what age a good fit of pension forecast and pension entitlement can be expected. Moreover, it is possible to examine which personal and life course characteristics make a good fit more likely.

Does the Switch to a Mixed-Mode Design Increase Panel Attrition? Evidence from the UKHLS Innovation Panel

Alessandra Gaia, University of Milan – Bicocca, Italy

In the last decade, in developed countries, panel surveys suffered high attrition rates. Mixed-modes designs has been adopted in an attempt to reduce survey costs and increase response rates for some subgroups of respondents. In this paper I evaluate the effect of a switch to a mixed-modes design in a longitudinal survey on panel attrition at the subsequent wave. I use experimental data from the Innovation Panel (IP) of *Understanding Society (UKHLS)*. At the fifth wave the sample was randomly allocated to two experimental groups: one assigned to a unimode face-to-face design and another to a mixed-mode design (web with a face-to-face follow up).

I use logistic regression to model the effect of the experimental allocation on attrition at the sixth wave. Furthermore, the treatment (mixed-modes allocation) is interacted with sample members' characteristics, household structure, mode preference and *paradata* from previous waves. Finally, the samples obtained with the two different designs are compared both across them and with the sample at the 4th wave.

I do not find evidence that a mixed-modes design increases attrition. On the contrary, for the original sample non respondents at wave 4 a mixed-modes design reduces attrition. No effect is found for the entire sample, or for the original sample respondents or for the refreshment sample (added at the fourth wave). The finding that a mixed-mode design decreases attrition for previous wave's non respondents is particularly promising since this group is at higher risk of attrition.

This study constitutes a novel contribution as the evidence on the effect of mixed-mode on panel attrition is scarce. This and other similar studies may contribute to the decisions on the mode of data collection for Understanding Society and other panel surveys. Further research is needed to assess whether this effect persists with future waves, or if this is eroded over time. Moreover, future research may compare costs and data quality in the two protocols.

A European Panorama of Social Survey Sampling Frames – and the Resulting Implications for a Possible Integrated System of Social Surveys

Britta Gauckler, Eurostat, Luxembourg

1. INTRODUCTION

In 2011, the European Statistical System Committee adopted the Wiesbaden Memorandum, which calls for the development of a modern architecture for social statistics. Eurostat subsequently launched a project on the streamlining and integration of the European social surveys; this project incorporates a major methodological study, concluded in 2014, which took the first steps towards the development of a methodological toolbox to deal with estimation and sampling issues in the context of an integrated system of social surveys.

A necessary step to achieving integration across surveys is the alignment of the sampling frames used; here, we will present the outcome of the analysis conducted within the study in this regard.

2. METHODS

Based on existing documentation (mainly quality reports), an overview was made of the sampling frames applied across the European Statistical System (ESS) of the following core social surveys: the Labour Force Survey (LFS), the European Statistics on Income and Living Conditions (SILC), the Adult Education Survey (AES), the Information and Communication Technology household survey

[ICT(HH)], the Household Budget Survey (HBS) and the Harmonised European Time Use Survey (HETUS).

3. RESULTS

As could be seen from Figure 1, countries are typically rather homogenous internally when it comes to the type of frame used (register, census or area-based).

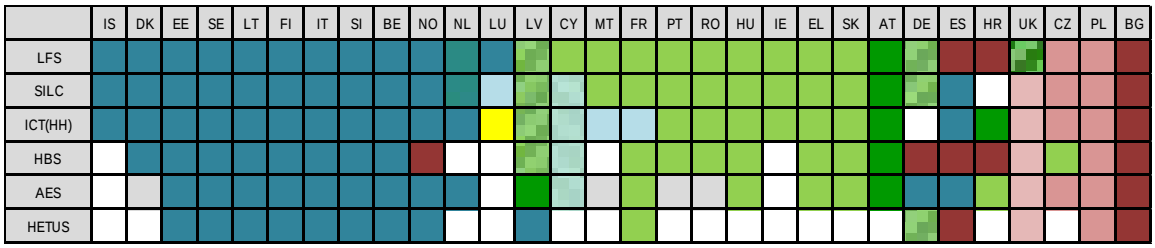

Figure 1. Frames used by 30 ESS members for seven different social surveys
Non-exhaustive legend: **Blue**=population register; **light green**=census; **red/pink**=area-based frames

In addition, a number of cases of surveys "embedded" in other surveys were identified; this practice is perfectly in line with the integration approach.

On the other hand, a number of potential obstacles to integration were found, including a number of differences in terms of target population (e.g. institutionalised persons) and in the definition of the household concept.

4. CONCLUSIONS

To serve as a basis for action, the European panorama needs to be further enriched with detailed information on the frames, such as stratification schemes, procedures for updating the frames, and the precise definition of the household concept.

Interviewer Effects in the Arab World: Evidence from Qatar

Justin Gengler, Social and Economic Survey Research Institute, Qatar

The Arab world is experiencing a critical transition in the availability of systematic and objective public opinion data, yet scholars continue to rely on techniques developed in non-Arab societies to evaluate overall survey quality and estimate total survey error. Interviewers are one of the sources of measurement error in surveys, and researchers have invested significant resources to create methods for detecting and reducing those errors. There are a handful of studies on interviewer effects in surveys conducted in the Middle East and North Africa (MENA), however none—to our knowledge—examines whether the nationality of the interviewer influences respondent answers. This paper asks how interviewer nationality influences responses to survey questions both direct and indirectly via selection effects. Using data from an original survey experiment conducted in 2014 in the Persian Gulf state of Qatar, the study finds strong evidence that interviewer nationality affects survey responses both directly and indirectly. Although this study was conducted in Qatar, we recommend controlling for interviewer characteristics as large-scale survey efforts expand across the MENA region.

Assessing and Adjusting Nonresponse Bias in Small Area Estimation via Bayesian Hierarchical Spatial Models

Chong He, University of Missouri, U.S.A

Nonresponse is a persistent problem in surveys because results from respondents only are subject to nonresponse bias. Many methods have been developed to deal with ignorable (missing at random) nonresponse data. In this paper, we provide a method to assess and adjust nonignorable (not missing at random) nonresponse bias in a small area estimation problem. We propose a bivariate Bayesian hierarchical linear mixed model to estimate both satisfaction rate and response rate. This model uses spatial dependencies among subdomains and auxiliary information from sample units to assess and adjust nonresponse bias. In addition, it explicitly includes a parameter that indicates whether the nonresponse is ignorable or not. The method is used to analyze the 2001 Missouri Deer Hunter Attitude Survey (MDHAS). The result shows that the nonresponse in MDHAS is nonignorable. Hunter age and the number of deer harvested have strong effects on satisfaction and response rates, and spatial dependencies are strong amongst counties of hunters' residences. The estimated satisfaction rates are lower after adjusting for nonresponse bias.

Next Generation Approaches to Mobile Data Collection from Survey Creation to Multiplatform Delivery and Mobile Survey Management

Pamela Hird, NASS/USDA, U.S.A

The National Agricultural Statistics Service (NASS) has been at the forefront of digital data collection for several years. But the advent of mobile technology and cloud services has shifted many of the technology paradigms. How can these new technologies best be employed to improve the data collection process?

In just the past few years, smartphone and tablet operating systems like iOS and Android have matured to the point where mobile technology is in many cases eclipsing the desktop. Meanwhile the software and services layers are being driven forward by advancements in social media platforms like Facebook and Twitter. With so many new tools and techniques at our disposal, mobile data collection is taking great strides forward.

This presentation will examine a full spectrum approach to next generation data collection starting with user centric design for mobile devices, cloud service technologies for survey delivery and offline support, and collaborative web applications for survey creation and analytics.

Participants will learn some practical advantages of cloud services, approaches to survey creation for multiplatform delivery, standardization techniques for mobile data collection and next generation techniques for collaborative survey development. In addition, participants will receive information regarding standardizing mobile surveys and how to manage mobile survey platforms.

Interviewer Age Effects Of the Inglehart Scale. A Comparison of the German and the American cumulative General Social Survey (GSS)

Volker Huefken, Heinrich-Heine-University, Department of Social Science, Germany

Inglehart’s human value orientations are the most widely used comprehensive value measurement. We explore how interviewer age influences responses to the post-materialist value orientations in Germany and the USA. Due to a discrepancy between modern, egalitarian value orientations we expect younger respondents to answer in a socially desirable way and exaggerate their post-materialism vis à vis older interviewers. At the same time, we assume that older respondents underreport their post-materialism to younger interviewers. Data came from the cumulative German General Social Survey (GGSS) 1980-2012

and from the American General Social Survey (GSS) 1972-2012. In the GSS the Inglehart Scale was obtained in four (1993, 1994, 2000, 2010) in the Germany GSS in 18 surveys. Information about the age of the interviewer are also different available, in the GGSS since 1984, and in the GSS since 2004. Under control of respondent education, Age of the respondent, Age of the Interviewer, and cohorts of the respondents binary logit models were estimated. For Germany the results are largely consistent with the expectations. However, the results do not indicate any significant effects across the GSS. Overall, the results show that social desirability may also bias value orientations in Germany. Moreover, interviewer age effects are strongly restricted in the GSS.

Supplementing Official Statistic of Cardiovascular Outcomes from NHANES III Data with Risk-Related Subgroup Trajectories

Edward Ip, Wake Forest School of Medicine, U.S.A

Common cardiovascular risk factors such as abdominal obesity, dyslipidermia, and hypertension often originate in childhood and adolescence. It is therefore important to understand trajectories of risk indicators during earlier life as the collective burden from various cardiovascular risk factors tends to accumulate over time and eventually leads to adverse cardiovascular and other health outcomes later in life. The NHANES III is a nationally representative cross-sectional survey conducted during the period of 1988 to 1994 on children and youths. The survey collected official statistic on many variables including body composition and health outcomes that can be used to characterize cardiovascular risk. Trajectories of risk variables can be derived from NHANES III data by connecting mean values of cross-sectional data. However, without longitudinal data it is difficult to understand how different risk measures covary over time. Additionally, phenotypic classification of individuals can only be analyzed at each time point, and thus the approach could miss important dynamic of change over time. Using data from a contemporaneous cohort study – Project Heartbeat! (N=678) – we supplement the NHANES III official statistic with trajectories of both a high and a low cardiovascular risk group on a collection of anthropometric, lipid, and blood pressure variables. Project Heartbeat! used intensive, longitudinal follow-up assessments of children and adolescents aged 8–18 in three separate cohorts (age 8-12, 12-15, 15-18). For each cohort, principal component functional curve analysis methods were used to summarize salient features of multiple longitudinal measurements in the data to produce trajectories for comparison with those from the NHANES III data. The high risk groups have increased waist circumference, body mass index, and percent body fat as well as higher low-density lipoprotein cholesterol and triglyceride levels, and lower high-density lipoprotein cholesterol. The risk profiles also exhibit patterns of convergence and divergence across the high and low risk groups as a function of age. In most cases, the NHANES III trajectories lie between the trajectories of the high and low risk groups. These observations have clinical and public health implications in identifying groups of children at high risk of CVD for earlier interventions.

Survey Feedback – on the Interplay Between Business Registers and Surveys

Fredrik Jonsson, Statistics Sweden, Sweden

Business surveys are usually conducted within a well-defined statistical system, considering a specific frame of objects, e.g. enterprises within a specific country, with the aim of studying change of relevant population parameters, such as the gross national product. Achieving these aims at a low cost, in terms of work load and distributed respondent burden, motivates the idea of designing business surveys with aid of coordinated probability sampling.

A business survey produces information regarding targeted variables, but also information regarding change in background variables normally kept by the business register, such as industry codes. Feeding this information directly back to partially update the register influences subsequent, coordinated survey sampling.

Our aim is to give a brief overview survey feedback, and the mechanisms with which it may distort statistical estimates with systematic errors. Different solutions are conceivable, roughly dividable into three categories: not allowing certain survey feedback to influence the register; allowing all influence of survey feedback; or some hybrid version where different versions of the register can be used at different steps in the statistical estimation process.

Based on recent experience at Statistics Sweden, we discuss the magnitude of the problem and evaluate some suggested solutions empirically. Practical issues when implementing and maintaining a satisfactory solution are also considered, within the broader picture of maintaining a fruitful interplay between business registers and surveys.

A Simulation Technique to Assess the Behaviour of Representativeness and Balance Indicators

Pär Karlsson, Statistics Sweden, Sweden

Responsive designs have been suggested in order to reduce the effect of nonresponse, by means of changing the collection strategy during the data collection. A prerequisite for adapting the strategy is some indicator of the data collection process. Several such indicators have been suggested in the literature, for example the representativeness indicator by Schouten et al (2009) and the various balance/distance indicators suggested by Särndal (2011). These indicators could in theory be calculated after each new data point collected, but most often they will be calculated only at a few time points. Perfect representativeness or balance will only be achieved at the end when we have data from the complete sample. During the data collection process we will observe that the indicators will numerically deviate from the ideal values. The question is how we should interpret this numbers. What is the range of negligible deviations? How will these indicators behave when the fraction of sample with observed data increases? ...

Here I will present a method that will assist in evaluating the behaviour of these indicators by simulation. The main idea is to line up the sample in a queue, which represent the order in which we get the data. The queue is then divided into two parts. The first part of the queue represents those who had responded, and the last part represents the nonresponders. The simulations generate different orderings of the queue. The simulated orderings can be done to mimic assumptions similar to missing completely at random or missing at random, where the ordering of the queue depends on auxiliary variables.

On Calibration Weighting for Nonresponse with Flawed but Correctible Frame Information

Phil Kott, RTI International, U.S.A

Sometimes in survey sampling we have access to a frame variable that is measured with error. For example, the frame may contain an imperfect indicator of whether a housing unit is owned or rented. Although (we will assume) the error in this variable can be corrected on the survey itself, using a corrected-frame values as a calibration variable will generally bias the resulting estimates. We will show how to avoid that source of bias when adjusting for unit nonresponse through calibration weighting. This can be done by treating the flawed-frame variable as a shadow variable to the corrected-frame variable in the weight-adjustment function. In other words, by calibrating on the flawed version of the variable while assuming, more reasonably, that whether or not a sample unit responds is a function of the corrected version. Since only the respondents are reweighted, we only needed to have corrected versions of the respondents' values in the weight-adjustment function. Some simple simulations will show the effectiveness this weighting approach.

Disclosure Control for Frequency Tables Based on Registers

Karin Kraft, Statistics Sweden, Sweden

Frequency tables generated from registers are often large in size, and large numbers of tables are generated for the same population. Relationships between tables are complex, including hierarchies or common cells or margins that appear in several tables. In addition, since they are based on totally enumerated populations, they are exposed to higher disclosure risks than tables based on sample surveys. Statistics Sweden has in recent years implemented standardised solutions for disclosure control of tables. A method for protection of frequency tables based on registers, developed by the Australian Bureau of Statistics (ABS), is currently being tested and evaluated. Tables are confidentialised by adding noise to cell values (perturbation). The same cell will always receive the same noise, regardless of in which table it appears. In contrast to methods that randomly apply noise, the risk that somebody will be able to disclose information by repeatedly requiring the same table is thus avoided. Each object in the micro data is randomly assigned a permanent numeric value called a record key. When a table is requested, the record keys of the units within a cell are combined to create a cell-level key. Via a perturbation lookup table (a fixed, two-dimensional array of numeric values), the cell-level key is used to determine the amount of perturbation applied to that cell. The perturbation look-up table contains the perturbation values that are to be added to the original cell counts. The look-up table is determined by maximising the entropy subject to a number of constraints, for example the magnitude of the perturbation, the variance of the perturbation, and other properties of the noise. Parameter values can be chosen such that other commonly used protection methods can be implemented using the same framework. In this study we compare the method developed by ABS to random rounding of table cells. The two methods are evaluated on risk reduction and on utility. Parameters of the two methods are varied in order to find the best protection and the largest utility for the protected tables.

Towards Efficient Statistical Data Editing

Karin Lindgren, Statistics Sweden, Sweden

Statistics Sweden has since 2004 worked with implementing selective editing in the business statistics production process. A generic software package called SELEKT has been developed for this purpose. Selective editing focuses on the search for suspected errors with large anticipated impact on output statistics and accepts that final data sets might contain a number of errors with no noticeable effect on the statistical estimates. The main intention is to reduce the resource demanding micro editing and thereby the tedious re-contacts with respondents, beneficial for both the national statistical institute and the responding businesses and organizations. SELEKT produces local scores based on suspicion and potential impact for all variables and all domains of study. The local scores are aggregated for each unit creating a global score determining whether the unit should be flagged for manual follow-up or not. The method behind SELEKT promotes a continuous measure of suspicion of a response being in error rather than a dichotomous one. Anticipated vales based on times series for the same unit or cross sectional data for similar units are used in the score functions. During implementation for a specific survey, cut-off threshold are set based on analysis of unedited and edited data from previous survey rounds. Relative pseudo bias, RPB, is a measure of how much bias is introduced by only editing a subset of the data. The bias is analysed relative to the estimated standard error of an estimate.

Harmonisation of Variables Treated in Business Surveys Questionnaires: a Standard Process

Stefania Macchia, ISTAT, Italy

Istat is reorganising the data collection process for business surveys through the implementation of ‘Business Portal’. This project moves from an approach based on a system of several and parallel single processes (the surveys), to a different one which puts the enterprise in the center. This new model

necessitates to accelerate the process of rationalizing and harmonizing surveys questionnaires in order to ensure uniformity of the questionnaires design and of the definitions of the concepts treated, so as to adopt a common language of communication with respondents. This need is emphasized by the fact that especially largest enterprises are often included into of the samples of different surveys, hence it can happen that they are asked to provide the same information more than once. Nevertheless definitions of the same variables could be different in terms of their content, because EU regulations contain different specifications for structural (SBS) and short-term surveys (STS), or in terms of wording, because described differently by each survey responsible, which should be overcome by the harmonisation activity. In this context an activity has been carried out to design a standard process to produce harmonised definitions of the variables collected in business surveys questionnaires. In the meanwhile a common structure to document these definitions, both for statisticians and respondents, has been defined. To set up the standard process, two variables were considered: Number of employees and Turnover. A series of different sources (national and European) have been identified to come to harmonised definitions. These sources have been analysed and re-elaborated with the aim of using simple, specific and unambiguous terms, clear both for statisticians and respondents. The definitions have been organised according to a Modular Structure, suggested as a standard for all variables: the main module contains the real meaning, expressed in a synthetical way, and the other three modules can contain respectively inclusion/exclusion clauses, peculiarities of surveys, references to national laws. The harmonisation process is the pre-requirement to achieve two main aims of the NSIs: i) the improvement of the comparability of collected data ruled by different EU regulations , ii) the strong reduction of redundancies and of respondent burden.

Respondent-Driven Sampling and Random Walks on Directed Networks

Jens Malmroos, Stockholm University, Sweden

Respondent-driven sampling (RDS) is often used to estimate population properties (e.g. sexual risk behavior) in hard-to-reach populations, e.g. injective drug users and homosexual men. RDS combines a modified snowball sampling methodology with an estimation procedure that yields unbiased population estimates under the assumption that the sampling process behaves like a random walk on the social network of the population. Current RDS estimation methodology assumes that the social network of the population is undirected, i.e., all relationships are reciprocal. However, empirical social networks in general also have non-reciprocated relationships. To account for this fact, we develop an estimation method for RDS on the basis of random walks on directed networks, where we derive estimators of the selection probabilities of individuals as a function of the number of outgoing edges of sampled individuals. The proposed estimators are evaluated on artificial and empirical networks and are shown to generally perform better than existing estimators. This is in particular the case when the fraction of non-reciprocated edges in the network is large.

A New Approach to the Box-Cox Transformation

Fernando Marmolejo-Ramos, Stockholm University, Sweden

We propose a new methodology to estimate λ, the parameter of the Box-Cox transformation, as well as an alternative method to determine plausible values for it. The former is accomplished by defining a grid of values for λ and further perform a normality test on the λ -transformed data. The optimum value of λ, say λ^* , is such that the p-value from the normality test is the highest. The set of plausible values is determined using the inverse probability method after plotting the p-values against the values of λ on the grid. Our methodology is illustrated with two real-world data sets. Furthermore, a simulation study suggests that our method improves the symmetry, kurtosis and, hence, the normality of data, making it a feasible alternative to the traditional Box-Cox transformation.

Using Multidimensional Unfolding for Picturing Agreement Between Partitions

Ana Martins, BRU-IUL and ISEL, Portugal

In the present work, we analyse data from EEA – European Environment Agency data (http://www.eea.europa.eu) concerning countries' greenhouse gas emissions which are reported by countries to UNFCCC (United Nations Framework Convention on Climate Change) and to the EU Greenhouse Gas Monitoring Mechanism (EU Member States). We focus our analysis in the sources of greenhouse gas emissions. These emissions are considered in an aggregated form taking into account the global warming potential of each one of the six Kyoto' gases. The emissions' sources are aggregated by sectors which definition is fully consistent with the common reporting format (CRF) set in the guidelines developed by the Intergovernmental panel on climate change (IPCC). The data includes information on 33 countries: EU-28 Member States, Iceland, Liechtenstein, Norway, Switzerland and Turkey. We perform clustering of the 33 countries based on their annual, per capita, greenhouse gas emissions (CO2 equivalent) referred to the following sources: energy, industrial processes, agriculture, LULUCF (land use, land use change and forestry) and waste (only the sectors without missing data were considered). The clustering procedure – the estimation of a mixture model - is applied (separately) to data concerning the years of 2005, 2008 and 2011. The Bayesian information criterion is used to select the number of clusters in each case. We then evaluate and compare the clusters profiles within and between years. For clustering evaluation we resort to the Silhouette index. As to the comparison between the partitions it is provided by two well-known measures of agreement – the Normalized Mutual Information (to measure simple agreement) and the Adjusted Rand index (to measure paired agreement). In order to clarify the relationships between two clustering structures referring to two different years, we specifically propose the use of Multidimensional Unfolding (MDU). MDU enables the spatial representation (typically in a two-dimensional Euclidean space) of the similarity data between two sets of objects (clusters, in particular). In the resulting map, each partition´s group is represented by a point in a way that the distances between points match, as closely as possible, their dissimilarities – the more similar the groups are, the closer are the corresponding points in the map. We compare MDU results based on raw data (counts of the confusion matrices) with MDU results derived from "adjusted counts" which are meant to exclude agreement by chance. The use of MDU analysis overcomes the need of implementing a matching procedure between the clusters of two partitions being compared. Furthermore, the MDU map provides an easy to interpret representation of the degree of agreement between partitions.

Considerations for Conducting Survey Methods Research in an Official Statistics Production Environment: Why is it Hard? How Do You Make it Work?

Jaki McCarthy, US Dept of Ag/National Ag Statistics Service, U.S.A

Survey methods research in official statistics can have far reaching impacts. Simulation studies or secondary analysis using existing data are common because researchers can conduct this type of research without involving the production staff or impacting ongoing survey operations. Alternatively, research is often conducted with stand alone research samples, but this is expensive and resource intensive and may not completely mimic production survey conditions. But survey methods research in the production environment of Federal Statistical agencies is extremely difficult to conduct for a number of reasons. Research will not provide realistic results without including sample units and procedures like those in the survey. However, operational staff are often hesitant to impose additional survey burden on potential sample members. This is particularly so in establishment surveys, where some units may be sampled often and be critical to estimates or survey populations may be small. Operational staff are also justifiably concerned about how research may change survey estimates if included in operational survey samples. Additional hurdles include a lack of perceived benefit of research, researchers' lack of understanding of the production environment (relative to either conducting research or implementing research findings), concern about potential negative impacts of research, and lack of interest and cooperation from operational staff.

Researchers should consider how to overcome these hurdles as a critical part of research study design. This poster will provide some examples from USDA/NASS showing some of the difficulties inherent in conducting survey methods research and approaches to overcoming them in ongoing Federal data collections. Clear communication between researchers and operational staff is key from research design through research implementation. This is often best facilitated by personal relationships and communication via multiple avenues. Researchers also need to fully understand survey operations to ensure that research is designed so that it is implementable and proposed procedures are reasonable. Survey methods research is difficult in a production environment but can provide information available through no other means. But to make this successful, research in a production setting must include these important considerations from beginning to end.

Model-Assisted Survey Regression Estimation with the Lasso

Kelly McConville, Swarthmore College, U.S.A

In the US Forest Service's Forest Inventory and Analysis (FIA) program, as in other government surveys, many auxiliary variables are available for use in model-assisted inference about finite population quantities. Some of this auxiliary information may be extraneous and therefore performing model selection can improve the efficiency of an estimator. However, FIA must produce several estimates each year and performing model selection for each variable would be a time-consuming task. A model-assisted regression estimator using the lasso is presented, along with an extension to logistic regression. By conducting model selection and parameter estimation simultaneously, the lasso estimator is easy to construct, even when many estimates are required. The gains in efficiency of the lasso estimator over the full regression estimator are demonstrated through a simulation study. An application of the lasso estimator to FIA data from the state of Colorado is also presented but the utility of the lasso estimator extends to any survey agency where a large number of finite population quantities are estimated in the presence of auxiliary information.

Establishing a General Framework for the Collection and Use of Paradata in Business Survey Through the ICT Experience

Alessandra Nuccitelli, Istituto Nazionale di Statistica, Italy

Detailed paradata were collected during the 2013 Italian survey on Information and Communication Technology (ICT) usage and on E-Commerce in Enterprises. On the basis of the available paradata, a set of indicators was created for each respondent company (e.g. connection time, number of days in which the connection occurs, number of visited questions, number of violated edit rules). Besides, the resulting indicators were merged by enterprise with statistical information coming from the ICT survey itself and from other sources, in order to evaluate if possible difficulties in completing the questionnaire may be attributed to some structural business characteristics. With reference to each navigational path, mean and percentiles of the proposed indicators indicated the most critical questions and sections of the Web questionnaire and suggested the measures to be taken to improve the accuracy of the answers. The results from this study have been used to better understand the quality of the survey data and to improve the ICT questionnaire, but only to some extent, as it is harmonized at the European level. In any case, this experience can be considered as a first step towards the systematic production and use of paradata in the framework of the new Italian Business Portal, a platform dedicated to the Web-based collection of statistical information from enterprises. In particular, this work has helped to define a set of basic paradata and indicators, which can also be used in real time to change edits and/or to add instructions, and some questions aimed at measuring the respondent burden. As the Business Portal involves very different surveys, standard paradata cannot cover all information needs, but ad hoc paradata can be produced in case of major changes in questionnaires.

Prejudicial vs. Non-Prejudicial Response Bias in Quasi-Experiments: When is (and isn't) Non-Ignorable Survey Response Bias a Threat to Internal Validity?

Jeffery Osgood, West Chester University, U.S.A

In quasi-experiments collecting data via surveys even non-ignorable missing data does not necessarily pose a threat to internal validity. Only if the non-response process for the Treatment group differs from the non-response process for the Comparison group will results be biased. This is not a new finding--but there is not even a term to describe the problematic type of non-ignorable missingness. We have been investigating methods of evaluating high school curriculum effects on college success--and in practical situations have found the missingness-process to be key in determining study validity. For example, in our study it was possible that Treatment students who failed a college math class got mad and wanted to respond to the survey, whereas Comparison students who failed a math class got embarrassed and became less likely to respond. We coined the term "prejudicial" non-response to describe the type of non-ignorable missingness that threatens internal validity.

Producing Official Statistics Through Big Data on Human Behaviour

Luca Pappalardo, University of Pisa, Italy

Every day we produce huge loads of data about ourselves simply by living in our technological world: we click Web pages, post thoughts on Facebook, make calls on our phones, communicate through emails, shop with credit cards. Such enormous amount of data, usually referred as Big Data, represent a powerful social microscope which allows scientists to understand and predict many aspects of human behavior. Can we use this social microscope to understand the socio-economic status of a territory? The quantity and the quality of our actions are clearly linked to the economic condition of the territory where they take place. Consider the case of human mobility. People move for a purpose, from work to leisure and shopping. A well-developed territory provides for a wide range of activities, an advanced transportation network, a higher availability and diversification of jobs, and other elements that foster human movements. As well as a higher mobility diversification of individuals lead to a higher economic well-being as it could nourish economy, establishes economic opportunities and facilitate flows of people and goods. Statistics institutes generally collect economic data by means of surveys, which generally are updated after months preventing an effective "nowcasting" of the current economic condition of a territory. In our work we show that Big Data allow us the definition of up-to-date measures for the economic development of a territory. We analyze two sets of Big Data: a dataset of phone calls made by 20 million users, and a dataset of millions of purchases performed by 1 million users. We first extract from the data some individual measures describing their shopping, mobile and social behavior. Then we explore the associations with several economic indicators, finding that the mobility diversity and the purchase sophistication of individuals resident in the same municipality varies with GDP, per capita income and a poverty index. Being confirmed by different null models that produce no correlations and by several regression models, our results reveal the high potential of Big Data in providing representative, relatively inexpensive and readily available measures as proxies of economic development.

Mathematics and Gender – The Use of Matching Methodologies for the Analysis of PISA 2012 Results for Portuguese Students

Iola Pinto, ISEL, Portugal

The Programme for International Student Assessment – PISA is a worldwide study by the Organisation for Economic Co-operation and Development (OECD) which aims to evaluate education systems in OECD member and non-member nations by testing the skills and knowledge of 15-year-old students. PISA is a triennial survey allowing countries to assess how well their 15-year-old students are prepared

for real-life situations. Since 2000, PISA has been testing students, near the end of compulsory education, in three competence fields: reading, mathematics and science. However, in each cycle, one of these areas is selected as principal domain of the study. Mathematics was selected as the PISA main area for the first time in 2003 and again in 2012. Science was the focus in 2006 and will be the domain of interest again in 2015. The assessment also collects information on students' backgrounds by asking them about themselves, their homes and their school and learning experiences in order to identify the factors that influence student performance. Portugal has participated in all PISA cycles along with 65 other countries/economies that tested their students in order to obtain information based on the performance standards established by the OECD. In PISA 2012 there are around 510,000 students involved and in Portugal the number of students considered for analysis, after data validation, is 5722. The first results presented in PISA 2012 consider countries rankings in each domain, and usual comparisons are made (e.g. by cycle, gender, grades, regions, contents,…). In this work we study the comparative performance of genders in the mathematics domain including the specific results on the following contents: uncertainty; quantity; space and shape; change and relationships. A more detailed comparison within proficiency levels is also considered, namely when referring to top performers and low performers. The use of matching methodologies enable to perform the comparisons taking into account diverse covariates referring to characteristics of students' education, their family and the students and their parents' attitudes towards Mathematics. Balance measures referred to the diverse covariates are obtained illustrating the degree of similarity between the samples from both genders.

Fellegi-Holt Method, an Application of Binary Vector Spaces

Paul Ramirez de la Cruz, Centro de Investigación en Matemáticas, A. C., Mexico

Fellegi-Holt method (FHM) is the international standard for editing and validating official statistics derived from surveys and censuses. The major difficulty that computational implementation of FHM faces is the alleged existence of an excessive amount of redundant implicit rules, resulting in either a large amount of time and effort spent to find all rules, or that the method is applied indirectly without actually deriving all implicit rules. We show that FHM is a particular case of application of binary vector spaces. Hence, there is no need to generate an excessive number of rules, but it is only required to complete a base to generate the vector space of implicit rules, which has a small number of vectors.

An R Package Expert System for Automated Data Description

Paul Ramirez de la Cruz, Centro de Investigación en Matemáticas, A. C., Mexico

Bureaus of national statistics and other related entities produce large amounts of tabular data and descriptive graphs available to users, either in printed or electronic format. Although very important, this information can sometimes be confusing for the occasional user. We present a R package with a basic approach to an expert system that produces an explanatory text of a dataset main features from the set itself or from a tabular summary.

Survey Data Editing at Statistics Spain: Towards an Industrialised Production Process

David Salgado, Statistics Spain, Spain

Statistics Spain is currently undergoing a strong effort to modernise the production process [1]. To this end we are following a stepwise approach concentrating upon critical phases of the process. In particular, data collection and data editing are under intense new developments to optimise resources while maintaing survey quality. Here we shall focus on data editing. By and large, as a general approach, we try to make a tightly joint use of Statistics and principles of Computer Science to develop survey methodological proposals. The goal is to integrate features of computer code maintainability and evolution right into the statistical methodology itself from its very design. In the organization of the production process we exploit the notion of production function [2] as a key concept to reuse and streamline well-delimited sets of tasks. Upon the generic EDIMBUS editing and imputation strategy [3] we are following

a slightly extended strategy which includes editing during data collection and is expressed in terms of editing functions [4]. As main editing functions, on the one hand, we are exploiting the historical data sets of each survey to construct score functions by automatically adjusting time series for each respondent [5]. This is used for input editing. On the other hand, we are following the optimization approach to selective editing for the output editing phase. As a first general result in production for the Industrial Turnover Indices, Retail Trade Indices and Services Sector Activity Indicators surveys we have reduced the respondents' recontact rate up to 20 percentage points. Furthermore, the implementation allows us to both reuse and enhance all editing functions involved.

How to Handle Survey Data from Children Under 12 Years of Age?

Fredrik Scheffer, Statistics Sweden, Sweden

Over the recent years the Unit for cognitive methods at Statistics Sweden have conducted several cognitive tests on questionnaires towards children, mainly in the age span 10-18 years. We have also, to some extent, analyzed quantitative data from our own Living Conditions Survey of Children (Barn-ULF) as well as surveys from The Swedish National Agency for Education and Swedish Schools Inspectorate.

The cognitive, communicative and social ability of children is still under development and this affects the different steps in the answering process. It has become obvious that the younger children, up to the age of twelve, do not have the cognitive ability in par with older children. Our own cognitive tests and analysis as well as survey literature seem to mark a divider at this age.

Children younger than 12 may have, compared to older children and adolescents, e.g. greater difficulty when it comes to negations, they interpret questions literally and are very susceptible to ambiguous questions and answering alternatives. They also, to a higher extent than adults and adolescents, tend to have problems when information is to be retrieved from memory and are as well at a greater risk of satisficing if they are not motivated enough.

In survey society a vast amount of data is collected from young children, mainly fifth graders (11 year olds (approx. 10-12)). This data are often merged with data collected from adolescents, for instance 14-15 years of age. Are the younger children to be treated as a group of their own or are their views comparable with older children's? Does data from the younger children confuse over all estimates when merged with data from older children and adolescents?

The poster will display (quantitatively) a tendency among younger children to use the extreme positive answering alternative as well as (qualitatively) illustrating citations from cognitive interviews.

Mixed-Mode in the Party Preference Survey (PSU)

Fredrik Scheffer, Statistics Sweden, Sweden

During the year of 2014 a project for mixed-mode in the Party preference survey has been carried out. The project task was to conduct and evaluate an experiment on web as a complement to telephone interviews within the scope of the Party preference survey´s so-called "methodological survey" in September 2014.

Among the challenges of the mixed-mode experiment was how to develop an IT tool for communication between the IT systems for telephone interviews and web surveys to make possible a transfer of respondents from web to the interview queue, and the other way around. Another challenge was how to design the main question of the survey, the party preference question, for the web. This included an experimental design of "don't know" options for comparison with telephone interviewing where this option is not visually available for the respondent.

The main subjects for evaluation were:

- The development of a communication tool between the two IT systems.
- Mode effects.
- Efficiency and savings (comparing experimental- and control groups).
- Response rates.
- In the web group; percentage of web answers, telephone interviews and nonresponse.
- Estimates of party preference and election result.
- Differences in item nonresponse and “don´t know”-answers between different modes as well as different question designs (open- or closed ended).

The results show, among other things, that it technically was successful to connect the two computer assisted systems for web surveys and telephone surveys. Additionally the results indicate a potential for increased response rates by offering a web alternative compared to only conducting telephone interviews in the Party preference survey.

The poster will include the design of the experiment, quantitative results such as response rates, estimates of main parameters for the control- and experimental group and estimates of “don’t’ know” answers by mode and by different question designs in the web. We will also show illustrating citations from open questions on preferred modes and respondent views on answering by the web alternative.

DISSEMINATING TO.M.OR.R.O.W.: Towards Multidimensionality, Organization, Reliability on Website Through Informative Statistical Systems

Daria Squillante, ISTAT, Italy

Population ageing enhances the urgency of a deep revision not only of social and economic organizational strategies for “greying societies", but also of statistical and dissemination strategies, to suitably monitor and communicate multidimensional dynamics, consequences and challenges tied to this planetary demographic revolution.

During the last years, the demand for indicators on population ageing has been experimenting a widespread increasing. In fact, on the one hand they seem to be the more and more indispensable for policy makers to program, implement and evaluate the real compliance of their results if compared to public aims; on the other hand, benchmarks are needful for citizens to assess adequacy and effectiveness of policies.

Therefore, beside the importance of individuating suitable indicators, the modernization of appropriate dissemination strategies becomes the more and more crucial too.

To some extent, ICT are already empowering users in interacting, exploiting and personalizing “knowledge”, through the World Wide Web. Consequently, National Institutes of Statistics should opt for the Internet as a prevailing channel of data dissemination too.

Thanks to this poster, the Authors would present the conceptual and methodological process through which ISTAT is accordingly developing the concept and the contents of “Elders.stat”: an innovative data warehouse that organizes statistical data produced by ISTAT on ageing, in order to make them more accessible to policy makers, researchers, journalists, citizens. Data and analysis are organized in a homogeneous and coherent way through themes, topics, and indicators, which users can customize according to their objectives.

The process to realize Elder.stat is based on a precise sequence of conceptual and operational steps, which would be retraced through the poster:

1. Analysis of the national and international statistics on population ageing;
2. Analysis of communicative aspects of national and international strategies to disseminate statistics on elderly: graphic design; contents; style; usability;
3. Individuation of themes and topics through which to analyse the daily life of elderly;
4. Individuation of appropriate indicators for each theme and topic;
5. Reconnaissance and systematization of the ISTAT's patrimony on elderly;
6. Planning of an extension and upgrade of ISTAT's informative supply on elderly.

SURVEYS FOR.WAR.D (Forthcoming Warranted Data)

Daria Squillante, ISTAT, Italy

In the pluriannual Program "Stat2015", designed by the Italian National Institute of Statistics (ISTAT) following both the European Recommendations and the Wiesbaden Memorandum on the renewal of social statistics, the innovation in data collection's techniques constitutes the core and crucial point.

This ambitious program of renewal involves the reengineering, standardization and industrialization of processes, as well as the integration between different sources (sample surveys and administrative sources) and products' innovation. The aim is the improving of the efficiency of the statistical system, guaranteeing statistical quality (in terms of respect of the quality dimensions identified by Eurostat), reducing expenditures and the statistical burden too.

Fundamental socio-economical ISTAT surveys have therefore been already involved in deep methodological changes as far as data capture techniques are concerned, abandoning the PAPI technique ("P*aper and Pencil Interview*") towards CAPI technique ("*Computer Assisted Personal Interview"*), or the more economic CAWI (*Computer Assisted Web Interview*) and CATI (*Computer Assisted Telephone Interview*) techniques.

Starting from this global methodological (r)evolution, this poster concentrates on the conceptual and methodological regeneration of one of the most meaningful European harmonized surveys: the European Union Statistics on Income and Living Conditions (Eu-Silc), went from the PAPI technique to the CATI technique.

Thanks to this poster it will be easy to capture the most relevant methodological innovations tied with:

a) survey design;
b) questionnaire design;
c) individuation and extraction of samples;
d) data quality control;
e) statistical treatment of errors;
f) data imputation;
g) data demand/gap satisfaction.

These assessments are the most important measures to improve the quality control of the data in a CATI survey.

On the Use of Scanner Data in the Consumer Price Index

Kristina Strandberg, Statistics Sweden, Sweden

Scanner data are big files of transactions, identified uniquely by an EAN-code, also called GTIN-code. Every time a consumer buys a product, the code is scanned at check out and the price is registered and added to the transaction file for the particular period. Frequent use of different scanner devices in most business and commercial transactions has prompted statistical offices in several countries to investigate whether scanned data can be useful in the production of statistics. The national Consumer Price Index (CPI) is a survey in the frontline in the use of this data. Most statistical offices still allocate considerable resources on manual price collection in outlets, in particular for food and other daily consumer items.

In 2011 a project was initiated at Statistics Sweden with the objective of replacing manually collected prices with scanner data for daily necessities. Thorough comparisons of the two data collection methods were made to ensure that the quality was good enough for price index calculations. Statistics Sweden learned that manual price collection more often yields errors than scanner data, due to price collector mistakes or erratic price information on shelves and packages.

Statistics Sweden now receives scanner data for daily necessities, aggregated on weekly basis. Turnover and number of packages sold are aggregated to a quantity weighted arithmetic average price per month. We argue for the use of probability sampling of products. The main advantage is that expert replacements can be made for products that leave the ever-changing market. In fact, the problem arises with the introduction of new EAN-codes which not always means a new product. The sample size can be large enough for keeping sampling variance at a negligible size.

For an NSI to benefit from scanner data, it is important to have an open dialogue with retailers and their organisations. Statistics Sweden has made considerable efforts to create a good relationship with people at various positions within the retail organisations. Statistics Sweden obtains scanner data free of charge by agreements written as partnerships.

A New Method for the Analysis of Interviewer Variance, With an Empirical Application

Patric Sturgis, University of Southampton, UK

It is well-known that interviewers tend to inflate standard errors, relative to self-completion modes, through the ways in which they administer survey questions (O' Muircheartaigh and Campanelli, 1994). When a question is asked across respondents in the same idiosyncratic manner by the same interviewer, a within-interviewer dependency is created. For survey analysis, this manifests in the form of greater variance in the population estimator. Traditional methods for the identification of interviewer variance in observational studies rely on estimation of random effects in multi-level models, where the random effect is specified at the interviewer level (e.g. Brunton-Smith and Sturgis 2012). This approach suffers from a number of analytical limitations, notably a lack of flexibility in modeling the effect of interviewer-level characteristics on response variance.

In this paper we set out a new method for the analysis of interviewer variance; the mixed-effects location scale model extends the random part of a standard 2-level mixed effects model by introducing random effects on the level 1 variance (Leckie, 2014). This enables response variance to be re-parameterized in a way that affords a more flexible and causally focused assessment of the factors associated with interviewer-induced response variability.

We apply this approach to data from wave 3 of the UK Household Longitudinal Survey (UKHLS), which we link to a diverse range of interviewer characteristics measured in an independent survey of interviewers. We demonstrate how the modeling strategy can be used to identify individual-level characteristics which are associated, conditionally, with larger interviewer variance components. We also

show how the approach might be used to improve survey quality by identifying interviewers, and interviewer characteristics, that are associated with more variable survey responses.

Small Area Estimators on Probabilistically Linked Data

Tiziana Tuoto, ISTAT, Italy

In Official Statistics, interest for data integration has greatly increased, though the effect of this procedure on the subsequent analyses has been disregarded for a long time. In recent years, however, it is largely recognised that linkage is not an error-free procedure and linkage errors, as false links e missed links, can invalidate standard statistical estimates. In statistical literature, growing attention is devoted to the effect of record linkage errors on the subsequent analyses. Chambers (2009) reviews the original work by Neter et al. (1965) and its extensions by Scheuren and Winkler (1993, 1997) and by Lahiri and Larsen (2005), moreover Chambers (2009) suggests a Best Unbiased Estimator and its empirical version. Finally, Chambers (2009) proposes a maximum likelihood estimator under the assumption of normality with application to linear and logistic regression functions.

Recently, Samart and Chambers (2014) consider the effect of linkage errors on mixed effect models, extending the settings in Chambers (2009) and suggesting linkage errors adjusted estimators of variance effects under alternative methods. Their proposal finds a natural application in the context of longitudinal studies, where repeated measures are taken on the same individuals.

In official statistics, the mixed models is largely exploited for small area estimation to increase detailed information at local level. For instance, administrative data can be used to augment information collected on sample surveys, in order to expand auxiliary information and improve the model fitting. Linkage of external sources to basic statistical registers as well as to sample surveys can be carried out on different linkage scenarios. Di Consiglio and Tuoto (2014) showed a sensitivity analysis for different alternative linkage error scenarios for linear and logistic regression estimators.

In this work, we extend the analysis on the effects of linkage errors on the predictors based on a unit level mixed models for small area estimation. Assuming mis-matches only occurring within the same small area, the effect on predictors would be given besides by the effects on fixed and random component estimation also by erroneours evaluation of covariates averages on unobserved units. We propose an adjustment of the unit level EBLUP in order to account for mismatches of auxiliary variables with the target variables.

Outcomes of a Calling Protocol in a Mobile CATI Survey

Paula Vicente, Instituto Universitário de Lisboa (ISCTE-IUL), Portugal

Mobile phones are increasingly being used to collect data for social, market and official statistics and some say it is just a question of time before they replace fixed phones or computers. Survey organizations spend a lot of effort trying to catch respondents for their surveys which, in the case of mobile CATI surveys, usually involves making several call attempts for a given number before giving it up. This procedure is dependent on workforce availability and deadline for survey period completion, and its efficacy depends on careful planning of call scheduling in order to increase successful contacts with less effort.

The objective of our analysis is to explore the effect of the calling protocol on the likelihood of contact and of interview in a mobile CATI survey. Data comes from a national mobile CATI survey conducted in Portugal in May 2012, targeted at mobile phone users aged 15 or older with the aim of describing the

population of mobile phone users in terms of mobile phone ownership and usage, and attitudes towards mobile phones.

Logistic regression models are fitted to the data in order to estimate the probability of contact and the probability of interview. A number of different predictor variables are considered and we shall distinguish between influences of the calling protocol on first calls and on subsequent calls.

Findings reveal that the time of the call does not significantly affect the likelihood of contact or of interview on the first call attempt, but some days of the week and specific interactions of day*time are better to yield high contact and interview rates. The outcomes also reveal that the call history has a marked effect on the efficiency of callbacks on mobile phone numbers. Specifically, callbacks on weekends are the most effective to yield both contacts and interviews and mobile phone numbers with many call outcomes of busy or voice mail are very difficult to interview.

Estimation of Non-Response in Survey Sampling Using Exponential Technique

Gajendra Vishwakarma, Indian School of Mines, India

In this paper we have suggested a procedure to estimate finite population parameters of the study variable in presence of non-response using auxiliary information. The properties of the proposed estimators have been obtained to the first degree of approximation. These estimators are compared for their precision with Hansen-Hurwitz and usual unbiased estimator. An empirical study is also carried out to judge the merits of the suggested estimators. Both theoretical and empirical findings are encouraging and support the soundness of the proposed procedure for estimation of population parameters.

The Survey Quality Predictor (SQP 2.0)

Wiebke Weber, Universitat Pompeu Fabra, Spain

This poster presents the program Survey Quality Predictor (SQP) 2.0, an online program to get predictions of the quality of survey questions available for free at sqp.upf.edu.

SQP 2.0 consists on a large database of survey questions with quality estimates and predictions. Quality estimates are obtained from Multitrait-Multimethod (MTMM) analyses, while quality predictions are obtained from SQP 2.0. This large database includes all survey questions from the European Social Survey (ESS) Rounds 1 to 6 and survey questions from a large variety of research fields, in many different countries and about many different topics.

Using this program, the users can obtain a prediction of the quality of new or currently available survey questions including reliability, validity and quality coefficients with confidence intervals and standard errors, and suggestions for improving them, in many different languages and for more than 20 countries. The only effort needed is to introduce the survey question and code its formal characteristics. The coding process in SQP 2.0 consists of 30 to 60 formal characteristics of the survey question, depending on its complexity. Examples of such characteristics are: the domain, the concept, the social desirability, the number of points in the answer scale, the presence of instructions for respondents or interviewers, etc.

Thus, SQP 2.0 is a very powerful tool both at the stage of questionnaire design, before data collection, in order to improve the survey questions forms, and at the stage of data analysis, after data collection, in order to correct for measurement errors.

The poster explains what is behind SQP 2.0, a meta-analysis of thousands of reliability and validity estimates obtained through MTMM experiments. Furthermore, it explains what you can do using this program, and how to proceed to do it.

Improvement of the Estimation Methods for the Cut Off Part in the Swedish Intrastat Survey

Frank Weideskog, Statistics Sweden, Sweden

Objectives: The Intrastat survey is a monthly business survey on arrivals and dispatches of commodities within the European Union. It takes the form of a cut-off survey, in which 93 percent of the arrival trade and 97 percent of the dispatch trade should be included (according to the current Intrastat regulations). The remaining part of the trade should be estimated as the cut-off part. From 2015 the exclusion threshold (cut off value) for arrivals has been increased from 450,000 to 900,000 Euros, and the estimation methods for the cut-off part have to be improved. For providers of dispatches the corresponding threshold is unchanged (450,000 Euros).

Methods: The (invoice) value for non-respondents and companies below the thresholds are first estimated on total company levels. In the next step their estimated trade will be distributed on certain commodities and country levels. This division is made using distribution keys, which is preferably decided by a comp any's previous reporting (for companies with history) and is otherwise decided based on a model of reported data from “similar” companies. In the case of “similar companies” a cluster analysis approach is considered to create homogenous estimation groups. The groups are made up of companies that trade with the same commodities. The groups are formed by size (according to the annual trade values) and the companies' NACE codes from the Swedish business register. The distribution key of a group is determined by the responding companies in a group.

Conclusions: Lower aggregation level in the estimations is needed since less reported detail data is available as auxiliary information when increasing the exclusion threshold (from a three digit level of NACE to a two digit level). Also the level of size classes has been considered and evaluated. Comparisons between old and new methods show some interesting results.

On Coordinating Outlier Treatment Approaches With Seasonal Adjustment of Systems of Time Series

Yingfu Xie, Statistics Sweden, Sweden

Seasonal adjustment is an important step in the official statistic production in order to eliminate seasonality in time series and facilitate visualizing the underlying and particular short term developments, which are usually the most interesting information for the statistic users or decision makers. Outlier treatment plays an essential role in SA to avoid distorted estimation of seasonality caused by extreme values, so-called outliers.

Methods to outlier treatment with seasonal adjustment of a single time series are well developed and implemented in the standard seasonal adjustment programs such as X-12-ARIMA or TRAMO/SEATS. In this poster some approaches are introduced and tested to coordinate outlier treatment during seasonal adjustment of systems of time series such as the national accounts, where both aggregates and their component time series are to be seasonally adjusted separately and the requirement of consistency between them is strong. These approaches include combinations of different ways to fix pre-identified outliers in individual time series and coordinate outliers between the aggregates and components time series, along with different critical values for the outlier identification.

The test data include the real Swedish national accounts (from the expenditure approach) and simulated data. Several deliberately chosen artificial outliers are added to the real data to illustrate different

scenarios such as the collapse of the Swedish real estate market and a new financial crisis. Even randomized outliers, in line with bootstrapping, are also used to add to the real data and test the proposed coordinating outlier treatment approaches.

The result is evaluated by several pre-determined criteria, consisting of both theoretical diagnostics such as white-noise test and spectrum test to the residuals, and empirical criteria such as revision errors, forecasting errors, contemporaneous inconsistency, M and Q indicators from X-12-ARIMA program. The evaluation shows that, as expected, there is no approach that over performs all others in term of all criteria. However one approach is shown robust against extreme scenarios and can be recommended to be used in practice.

This investigation and its result may be of interest for those who need to seasonally adjust systems of time series.

Small Area Estimation for Neighborhood-Level Health Indicators Using Semi-parametric Model with Survey and Population Data

Hongijan Yu, UCLA Center for Health Policy Research, U.S.A

Using data from California Health Interview Survey (CHIS), AskCHIS Neighborhood Edition (NE) is a project that produces and disseminates SAE of a number of health indicators at zip codes, cities, legislative districts and counties in California simultaneously. Unique challenges exist for projects of such type and scale. First is lack of information for granular areas where samples are few or non-existent. This could impact the efficacy of canonical SAE models in which area random effects are explicit. Second is conformity among estimations for multiple geography settings. Users expect estimations to be consistent when smaller areas are aggregated to larger ones. In addition, some areas may not follow the geo-boundaries commonly available in survey, such as legislative districts. We thus developed a method using semiparametric geo-additive model to fit census tract auxiliary predictors in penalized spline regression in addition to unit level generalized linear model. We broadened the concept of spatial correlation to include auxiliary SES contextual variables derived from ACS. Modeling them the same way as geo-coordinates enhances the ability to borrow strength at granular level from the “neighbors” through geographic as well as “SES” adjacency. Penalized spline (P-spline) models continuous auxiliary predictors without imposing strict assumptions. Through common form of minimization criteria, fitting of P-spline can be converted to the solutions for coefficients of random effect models which is commonly available in statistical software. The estimated model parameters are then applied to a census type population data with same set of predictors and comprehensive geo-ids. The SAEs are generated through aggregation of predicted values in population data. The population data enables us to “contextualize” the granular areas, as well as generate consistent estimation for all levels of areas, even for the areas with the boundaries not available in survey. The uncertainty of the model is estimated through survey replicate weights. This greatly reduced complexity in taking into account of survey design. PMSE is derived from the Jackknife method by Jiang et al 2002.