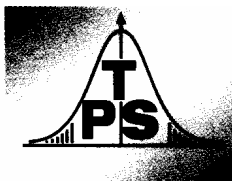




***UNIVERSITY OF ECONOMICS IN KATOWICE
DEPARTMENT OF STATISTICS***



***UNIVERSITY OF ŁÓDŹ
DEPARTMENT OF STATISTICAL METHODS***



POLISH STATISTICAL ASSOCIATION

IN THE 60TH ANNIVERSARY OF THE FOUNDATION OF DEPARTMENT OF STATISTICS

***VII CONFERENCE
SURVEY SAMPLING IN ECONOMIC AND SOCIAL RESEARCH***

Katowice 18-20 September 2011

BOOK OF ABSTRACTS

SPONSORED BY



THE CITY OF KATOWICE

***predictive
SOLUTIONS***



SPSS POLSKA

Katowice 2011

THE AIM OF THE CONFERENCE

In recent years a growing need is observed for statistical data, collected quickly and at low cost. This defines the role of the survey sampling and the importance of improving survey sampling procedures from theoretical and practical perspective. The conference gives an opportunity to present latest developments in this and related fields and to exchange experience on practical applications of survey sampling.

CONFERENCE TOPICS

- Estimation of population parameters based on complex samples
- Statistical inference based on incomplete data
- Small area estimation
- Sample size and cost optimization in survey sampling
- Sampling designs
- Statistical inference using auxiliary information
- Model-based estimation
- Longitudinal surveys
- Practical implementations of sampling methods
- Sampling in statistical quality control
- Sampling in auditing

ORGANIZING INSTITUTIONS:

The University of Economics in Katowice
The University of Łódź
Polish Statistical Association

PROGRAMME COMMITTEE:

Andrzej Barczak, Czesław Bracha, Czesław Domański, Malay Ghosh, Nicholas T. Longford, Zdzisław Hellwig, Jan Kordos (chair), Walenty Ostasiewicz, Jan Paradysz, Jan Steczkowski, Jacek Wesołowski, Janusz L. Wywiał

LOCAL ORGANIZING COMMITTEE:

Janusz L. Wywiał (chair), Józef Kolonko, Wojciech Gamrot, Tomasz Żądło (secretary)

VARIANCE ESTIMATION OF HOT-DECK IMPUTED ESTIMATOR OF CHANGE OVER TIME FROM REPEATED SURVEYS

Yves Berger, Rodolphe Priam

Measuring change over time is a central problem for many users of social, economic and demographic data and is of interest in many areas of economics and social sciences. Smith et al. (2003 JRSS-D) recognised that assessing change is one of the most important challenges in survey statistics. The primary interest of many users is often in changes or trends from one time period to another. A common problem is to compare two cross-sectional estimates for the same study variable taken on two different waves or occasions. These cross-sectional estimates often include imputed values to compensate for item non-response. The estimation of the sampling variance of the estimator of change is useful to judge whether the observed change is statistically significant. Covariances play an important role in estimating the variance of a change. We propose to use a multivariate linear regression approach to estimate covariances. The proposed estimator is not a model-based estimator, as this estimator is valid even if the model does not fit the data. We show how this approach can be used to accommodate the effect of imputation. The regression approach gives design-consistent estimator for the variance of change when the sampling fraction is small and the finite population corrections are negligible.

ESTIMATION OF MEDIAN INCOMES FOR SMALL AREAS : A BAYESIAN SEMIPARAMETRIC APPROACH

Dhiman Bhadra, Malay Ghosh, Dalho Kim

Small area estimation procedures have been frequently applied for the estimation of household and family incomes of small domains. The resulting estimates are often used for various governmental decisions and policies. For example, the median household income for of the 50 states and District of Columbia are used by the U.S Department of Health and Human Services (HHS) in formulating their energy assistant program for low-income families. The HHS procure these estimates annually from the U.S Census Bureau by an informal agreement. The motivation of our work originates from the repetitive nature of the median income estimates. We have viewed the median income values collected annually for a particular state as longitudinal profiles. Also, we have modeled the response-covariate relationship using a non-parametric function estimated by Penalized splines. We have modeled the P-splines using truncated polynomial basis functions with varying degrees and varying number of knots. Analysis has been carried out using a hierarchical Bayesian approach.

USES OF CALIBRATION AND BALANCED SAMPLING FOR THE CORRECTION OF NON-RESPONSE IN SURVEYS

Jean-Claude Deville

Generalized calibration is the simplest and more efficient method for weighting responding units to deal with non-response. In particular, one can use variables of interest in the response model if it is necessary to get unbiasedness. In the case of item non-response, weighting is generally impracticable and imputation must be used. We look for the deterministic (or prediction) imputation methods using parametric models adjusted by some kind of normal equations. It is shown that a convenient choice of the instruments allows an estimation of the model parameters without substantive bias and makes the estimation of totals coherent with weighting methods. Deterministic imputation makes impossible to estimate the total of a function of the variable of interest like a quantile. That the reason for using random imputation where, for numerical variable, a residual is added to the prediction. Unhappily, this procedure adds a parasitical variance to the estimations. The use of a balanced sample for the imputed values allows to reduce, and, sometimes, to annihilate this effect.

THE FIRST POLISH ASSOCIATION OF STATISTICIANS

Czesław Domański

In 2012 we will be celebrating the 100th anniversary of the foundation of the first Polish statisticians' association - Polish Statistical Association. It is important to emphasize on this conference the contribution of Polish statisticians in the development of statistics as a scientific discipline and didactic. This conference also has an anniversary nature as the Department of Statistics, respectively known as the Higher School of Economics, University of Economics, and currently the University of Economics in Katowice, celebrates this year its 60th jubilee of the foundation.

Let me mention here chosen Polish statisticians' achievements that have an international greatness:

1. The implementation of the first national census in 1789, the main originator and executor of it was count Frederick Joseph Moszyński (1737-1817), a member of Polish Parliament.
2. The foundation of the Chair of Statistics in 1811 at the School of Law and Administration in Warsaw, whose leadership was given to the professor of statistics and economics Mr. Wawrzyniec Surowiecki (1769-1827).
3. Tadeusz Pilat (1844-1923), who was University of Lviv professor of statistics and administration, participated in the creation of the International Statistical Institute.

SAMPLING AND COEFFICIENT OF VARIATION: APPROXIMATING EXPONENTIAL DISTRIBUTIONS

Nikolaos Farmakis

The approximation of the distributions near to the Exponential, by a polynomial probability density function (pdf) is studied. We use sampling data as in Farmakis (2003, 2010) and Farmakis & Makris (2011). We utilize the idea of the Coefficient of Variation (Cv). The sampling takes place in a population Π , e. g. city people, websites, lambs produced by a factory. We approximate the pdf of a random variable (rv) X similar to the

$$f(x) = \mu \cdot e^{-\mu x}, x \geq 0 \quad (1)$$

with a polynomial pdf like

$$f(x) = \begin{cases} h \cdot \left\{ 1 - \left(\frac{x}{b} \right)^k \right\}, & x \in [0, b] \\ 0 & x \notin [0, b] \end{cases} \quad k \neq -1, h = \frac{k+1}{b \cdot k} \quad (2)$$

based on sampling, i.e. we use values of X in a sample δ of size n drawn from Π , of size $N > n$. In (2) we calculate the exponent k via the idea of the Cv. Note that both (1), (2) are decreasing functions. Since in (1) is $EX = \frac{1}{\mu}$ & $VarX = \frac{1}{\mu^2}$, we have $Cv=1$,

always. From (2) we take: $k = -2 + \sqrt{3 \cdot \frac{\lambda + 1}{3 - \lambda}}$ with $\lambda = Cv^{-2} = 1$, i.e.

$k = -2 + \sqrt{3} = -0.2679$ & $h = \frac{-2.7327}{b}$. So the polynomial pdf is

$$f(x) = \begin{cases} \frac{-2.7327}{b} \cdot \left(1 - \left(\frac{x}{b} \right)^{-0.2679} \right) & x \in [0, b] \\ 0 & x \notin [0, b] \end{cases} \quad (4)$$

We estimate the extension b, the distribution range, according to the data. Illustrative examples are given.

References

- Farmakis N. (2003) Estimation of Coefficient of Variation: Scaling of Symmetric Continuous Distributions, *Statistics in Transition*, Vol. 6, No 1, pp 83-96.
 Farmakis N. (2010) Coefficient of Variation: Connecting Sampling with some Increasing Distribution Models, *Proc of Stochastic Modelling Techniques and Data Analysis Intern Conference (SMTDA2010)*, Chania, Greece.
 Farmakis N. & Makris G. (2011) Web-sampling: Probabilities of Specific Information Achievement, *Proc of Applied Stochastic Models & Data Analysis, International Conference (ASMDA2011)*, Rome, Italy.
 Kolaczyk E. (2009) Statistical Analysis of Network Data, Springer Science & Business Media, New York.

ESTIMATORS FOR THE HORVITZ-THOMPSON STATISTIC BASED ON ITS POSTERIOR DISTRIBUTION.

Wojciech Gamrot

The Horvitz-Thompson estimator for the finite population total is constructed as a linear combination of sample values of the fixed characteristic under study. The weights of the combination are usually inverses of first-order inclusion probabilities characterizing the sampling scheme used to draw the sample. However sometimes the sampling design is complicated and combinatorial explosion prevents the exact computation of inclusion probabilities. In such a case inclusion probabilities may be estimated from the simulation experiment involving generation of independent sample replications via the same sampling scheme. Meanwhile often some upper and lower bounds on inclusion probabilities may be easily obtained analytically. There are at least two reasons to include this auxiliary information in the estimation of sampling weights. First, it may prevent the appearance of infinite weights in unlucky case of getting the estimated inclusion probability exactly equal to zero. Secondly, it may improve the accuracy of the empirical Horvitz-Thompson estimator. An intuitive way of introducing such auxiliary information is the restricted maximum likelihood principle. In this work some alternative estimators are considered.

ROBUST SMALL AREA ESTIMATION

Julie Gershunskaya, Partha Lahiri

Different methods have been proposed in the small area estimation literature to deal with influential observations. In this paper, we develop a robust estimation technique that is reasonably resistant to influential observations and a parametric bootstrap method to construct confidence intervals of small area means. Using a Monte Carlo simulation study, we compare our proposed method with a few recently proposed robust small area estimators. Empirical evaluation of the estimators is performed using the population data from administrative file.

Key Words: robust estimation, small area, mixture model

PRACTICAL STATISTICAL AND ECONOMIC ASPECTS OF USING SURVEY STUDIES FOR IDENTIFICATION OF THE KEY PLANT CULTIVATION TECHNOLOGY FACTORS.

Anna Imiolek, Janusz Golaszewski¹, Dariusz Zaluski, Zbigniew Nasalski

A survey study was carried out in 2008 in order to determine the key elements in a plant production technology and to calculate unit production costs of growing winter rye (*Secale cereale* L.) for grain. The surveys covered rye grain producers in northeastern Poland, who grow rye on an acreage of over 1 ha. The survey contained questions, divided into several groups, which dealt with the following issues: 1) general characteristics of a farm, 2) rye production technological factors, 3) evaluation of energy consumption (agro-technical evaluation) and 4) structure of outlays. The data on production factors served as predictors in a general linear model, and the grain yield was a dependent variable. In the analysis of variance of the grain yield, type III sums of squares were applied, and the main effects of the factors were assessed. The economic analysis was performed based on direct outlays on production; unit costs and direct margin were calculated and the structure of costs as well as profitability of winter rye production were determined.

THE COMPARISON OF GENERALIZED VARIANCE FUNCTION WITH OTHER METHODS OF PRECISION ESTIMATION FOR POLISH HOUSEHOLD BUDGET SURVEY

Alina Jędrzejczak, Jan Kubacki

Sampling variance or relative variance of a survey estimator can be related to its expected value by a mathematical relationships called generalized variance function (GVF). In the paper, the results of precision estimation using Generalized Variance Function for various income variables from Polish Household Budget Survey for counties (NUTS4) are presented. An attempt was made to compare this method with other precision estimation methods. A starting point was the estimation of Balanced Repeated Replication variances, or bootstrap variances in the situation where using BRR was not applicable. To evaluate the GVF model the hyperbolic function was used. The computation was done using WesVAR and SPSS software and also special procedures prepared for R-project environment. The assessment of estimates consistency for counties was also conducted by means of small area models.

ESTIMATION OF UNEMPLOYMENT IN WIELKOPOLSKA REGION VIA SPATIAL MODEL WITH INFORMATION ABOUT COMMUTES

Tomasz Klimanek

This article presents an attempt to estimate unemployment in Wielkopolska Region (Poland) at the NUTS4 level, with application of the technique of indirect spatial estimation. The article presents the application of estimation method borrowing the strength over space (distances between centroids of the poviats & NUTS4 level areas) and from other data sources than the sample itself. In Poland little work has been done in the field of estimation which makes use of spatial correlation structure. However this kind of estimation should be of a special attention if phenomenon of interest exhibits the clear spatial pattern. We notice clear patterns in the spatial distribution of commuting people. They play an important role at the local and regional labor market and the economic centers of gravity. In the authors' opinion using the information about commuters at the area level should significantly increase the quality of the applied spatial model. Four sources of data were used in this research. The first source was the Labor Force Survey data from all of the four quarters of 2008. The second source was system of numerical maps and information they contain (topology, i.e.: distances, centroids, neighborhood). The Tax Register was the third source of information. It helped to detect the commuting people, based on some tax legislation rules. The last source, but also very important one, was the Social Insurance System. It contains some information about people's status on Labor Market.

REVIEW OF APPLICATION OF ROTATION METHODS IN SAMPLE SURVEYS IN POLAND – THEORY AND PRACTICE

Jan Kordos

The author reviews theory and application of rotation methods in sample surveys in Poland. He begins with reviewing designs of the surveys across time, depending on different objectives, focusing on partial rotation of sub-samples, and considers estimation problems and data quality issues generally. Next he refers to some articles and books published about surveys over time, starting with Wilks (1940), Patterson (1950), Eckler (1955), Woodruff (1963), Bailar (1975), Duncan & Kalton (1987), Kalton & Citro (1993). He mentioned also early Polish papers on rotation methods (Kordos (1966, 1967, 1971, 1982; Lednicki, 1982; Szarkowski & Witkowski, 1994), and concentrates on the Polish household surveys, mainly Household Budget Survey (HBS), Labour Force Survey (LFS), and EU Statistics on Living Conditions and Income (EU-SILC). Special attention is devoted to last research on rotation sampling done by Polish sampling statisticians Ciepiela et al. (2011), Kordos (2002), Kowalczyk (2002, 2003, 2004), Kowalski (2006, 2009), Wesołowski (2010). Concluding remarks are given at the end.

Key words: rotation sampling; sampling on successive occasions; survey across time; panel survey; data quality; sample survey.

References

- Bailar, B. (1975), The effects of rotation group bias on estimates from panel surveys. *JASA*. 70, 23-30.
- Ciepiela, P., Gniado, K., Wesołowski, J. and Wojty, M. (2011), Dynamic K-composite Estimator for an arbitrary rotation scheme, *Statistics in Transition – new series*, Vol. 12, No. 1 (in printing).
- Duncan, G. J., Kalton, G. (1987): Issues of Design and Analysis of Surveys Across Time, *International Statistical Review*, 55, pp. 97-117.
- Eckler, A. R. (1955). Rotation Sampling. *Annals of Mathematical Statistics*, vol. 26, pp. 664-685.
- GUS (1972), Experimental Research of Household Budget Survey by Rotation Method (in Polish), *Biblioteka Wiadomości Statystycznych*, vol. 18, Warszawa.
- Hansen, M.H., W.H. Hurwitz, H. Nisselson, and J. Steinberg (1955). The redesign of the Census Current Population Survey. *J. Amer. Stat. Assoc.* 50, 701-709.
- Jessen, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Ag. Exper. Station Research Bull.* 304, 54-59.
- Kalton, G. i Citro, C.F. (1993), Panel surveys: adding the fourth dimension. *Survey Methodology*, vol. 19, pp. 205-215.
- Kordos, J. (1967), Rotation Method in Sample Surveys (in Polish), *Przegląd Statystyczny*, No. 4, 1967, pp. 373-394.
- Kordos, J. (1971), . Experimental Household Budget Survey by Rotation Method (in Polish), *Przegląd Statystyczny*, No.3- 4, 1971, pp. . 227-246.
- Kordos, J. (1989), Longitudinal Household Surveys in Poland. *Proceedings of the ISI 47th Session. Contributed Papers. Book 4*, Paris, 1989, pp. 123-132.
- Kordos, J. (2002), Estimation Problems and Data Quality in Rotation Samples, In: *Rotierende Stichproben - Datenkumulation und Datenqualität*, *Spektrum Bundesstatistik*, Band 21, Statistisches Bundesamt, 2002, pp. 57-73.
- Kowalczyk, B. (2002). Sampling in Time (in Polish). Ph.D. Thesis, Warsaw, School of Economics, Warsaw, Poland.
- Kowalczyk, B. (2003), Estimation of the population total on the current occasion under second stage unit rotation pattern, *Statistics in Transition*, December 2003, Vol. 6, No. 4, pp. 503—513
- Kowalczyk, B. (2004), Number Index Estimation Using Auxiliary Information in Repeated Rotating Surveys, : *Sampling Designs for Environmental, Economic and Social Surveys: Theoretical and Practical Perspectives*, *Universita degli Studi di Siena*, 23-24 September 2004, Italy.
- Kowalski J. (2006), Rotation in sampling patterns, in review for *JSPI*, 2006.
- Kowalski, J. (2009), Optimal Estimation in Rotation Pattern, *J. Statist. Plann. Infer.* 139 (4), pp. 2429-2436.
- Lednicki, B. (1982), Sampling Design and Estimation Method in Household Budget Rotation Survey (in Polish). *Wiadomości Statystyczne*, No. 9.
- Patterson, H.D. (1950), Sampling on successive occasions with partial replacements of units. *Journal of the Royal Statistical Society, ser. B*, vol. 12, s.241-255.
- Popiński W., (2006), Development of the Polish Labour Force Survey, *Statistics in Transition*, Vol. 7, No. 5, pp.. 1009-1030.
- Rao, J.N.K., Graham, J.E. (1964), Rotation designs for sampling on repeated occasions. *Ann. Math. Statist.* 35, 492-509.
- Szarkowski, A., Witkowski, J. (1994), The Polish Labour Force Survey, *Statistics in Transition*, Vol. 1, No. 4, pp. 467—483.
- Wesołowski, J. (2010), Recursive Optimal Estimation in Szarkowski Rotation Scheme, *Statistics in Transition – new series*, Vol. 11, No.2., pp. 267-285.
- Wilks, S.S. (1940), Representative Sampling and Poll Reliability, *Pub. Opin. Quart.*, 4.
- Woodruff, R.S. (1963), The Use of Rotating Samples in the Census Bureau's Monthly Surveys, *Jour. Am. Stat. Ass.*, vol. 58, pp. 454-467.

ESTIMATION OF NET CHANGES IN THE CONTEXT OF MULTIPURPOSE ROTATING SURVEYS

Barbara Kowalczyk

Most of repeated surveys are of multipurpose nature. They are usually conducted to estimate population parameters on current occasions, net changes, gross changes and total values of several periods simultaneously. The multipurpose nature of repeated surveys poses additional questions to the already difficult theory of rotating surveys. The aim of the paper is to analyze how increase of the efficiency of the estimation on current occasion affects efficiency of net changes estimation in the case of rotating surveys. Various estimators are examined, e.g. estimator using information from previous occasions on the studied variable and regression estimator using auxiliary information on each occasion. In all cases general rotating pattern is taken into account, including overlapping samples, non overlapping samples and the same sample on two successive occasions.

APPLICATION OF LATENT CLASS ANALYSIS TO ESTIMATE FULL POPULATION STRUCTURE FROM INCOMPLETE DATA

Roman Konarski

Knowledge of the joint distribution of population characteristics allows to construct sampling weights and subsequent poststratification. Frequently, however, the full joint population distribution of the relevant characteristics is not known and must be estimated. It is therefore important that the method used to estimate the joint population distribution of these characteristics utilizes the partial information that is available from the census data. We approach the problem from the perspective of missing data and apply the latent class (LC) model to the analysis of contingency tables, in which certain cells of the full table are observable only for some subgroups. As a consequence some variables that constitute the full contingency table are latent variables (missing) for some subgroups of observations, and are observed variables for the remaining observations. The LC model can be estimated by widely available software packages (eg., LEM or LatentGOLD), that allow for multisample analysis. The approach is very flexible, allowing many sources of data and contingency tables of arbitrary dimensions. The application of the proposed methodology is demonstrated with an example of estimating the full contingency table (joint distribution) of four population characteristics: locality type (4 levels), age (6 levels), gender (2 levels), and education (4 levels). Three sources of information are utilized: census data from GUS about the joint population distribution of three variables (locality type, age, and gender), estimated data from GUS about the joint distribution of two variables (gender and education), and survey estimates of the full joint distribution (the full population structure) of the four variables.

USEFULNESS OF PAST DATA IN SAMPLING DESIGN FOR EXIT POLL SURVEYS

Arkadiusz Kozłowski

Exit polling has become an essential part of elections in developed democracies. Its major objective, though not the only one, is to predict winning candidates/committees on the election day, usually just after completing voting. Exit polls provide valuable information about demographic, social, and political characteristics of voters supporting particular parties. Accuracy of exit polls in Poland is sometimes a subject of heavy criticism. The imperfection of polls from the last few years may be a result of poor sample selection of electoral regions. One possibility to improve the quality of sample is to take advantage of information about outcomes of past elections. The aim of the article is to find out whether the knowledge of distribution of votes in each electoral region from the past several elections can be effectively utilized for increasing the quality of the sample of polling stations. Motivation for this research is a supposition that some electoral regions permanently yield distributions of votes that are closer to the overall outcome than the others. Author uses data obtained from the State Electoral Commission (Państwowa Komisja Wyborcza), containing election outcomes at the electoral regions level, to investigate the above topic.

SUCCESSIVE SAMPLING DESIGN IN PRACTICE

Danute Krapavickaite

If there is no list of households and dwellings, population register is used as a sampling frame, and dwelling sample is selected through persons. Equal probability sampling of individuals without replacement is used, and the dwelling of selected persons is included into the sample, until the predetermined dwelling sample size n is obtained. This is a successive sampling design introduced by Bengt Rosen. The second phase stratified sampling design and calibration of weights is used. This sampling strategy may be effectively used for household surveys. It will be compared with other sampling strategies theoretically and empirically.

IMPROVEMENT OF ESTIMATES FOR THE DUTCH STRUCTURAL BUSINESS SURVEY BY SMALL AREA ESTIMATION

Sabine Krieg

Traditionally, statistical offices like Statistics Netherlands prefer design based techniques as the generalized regression (GREG) estimator to produce estimates from survey samples. The estimates based on these techniques are always approximately design unbiased. GREG estimators, however, have relatively large design variances in the case of small sample sizes. Therefore, Statistics Netherlands investigates model based techniques as an alternative. The approach considered in this paper is based on J.N.K. Rao: Small area estimation (Wiley, 2003). By applying a multilevel model, information from other subpopulations is borrowed to improve the accuracy. The approach is applied to the Structural Business Survey, an annual survey about the Dutch business. One important target variable is turnover, which has a skewed distribution. In this research project, an artificial population is created to investigate the properties of the method and to carry out a simulation study. The variable turnover, which is only given for the sample, is replaced by the variable tax-turnover, which is available for the entire population. Since the correlation of tax-turnover and turnover is high, the artificial population is similar to the real population. It is investigated whether the accuracy can be improved by transforming the data. Two-level models are developed both for turnover and for the transformed variable. The accuracy of small area estimates based on both models is compared with each other and with GREG estimates in a simulation study.

POLICY-RELATED SMALL-AREA ESTIMATION

Nicholas T. Longford

A method of small-area estimation with a utility function is developed. The utility characterises a policy planned to be implemented in each area, based on the area's estimate of a key quantity. It is shown that the commonly applied empirical Bayes and composite estimators are inefficient for a wide range of utility functions. An argument is presented for a closer integration of estimation and (regional) policy making. Adaptations for limited budget to implement the policy are explored.

Keywords: Expected loss; exploiting similarity; small-area estimation; shrinkage; utility function.

ADDITIONAL SAMPLES WITH BALANCING OR OVERLAPPING CONDITIONS AND GIVEN INCLUSION PROBABILITIES : THEORETICAL APPROACH AND EXAMPLES IN THE FRAMEWORK OF PISA SURVEYS.

Christine Marc

This paper provides a theoretical frame and methods to solve a problem which occurs as soon as a first sample has been drawn at a given time and that one intends later to draw a 2nd sample in an updated sampling frame, linked in a way with the 1st one, but without any possibility of changing the conditions or results of the drawing of the former sample. The origin of this issue lies in PISA surveys (Programme for international student assessment): the next cycle 2012 will be on the same main topic as in 2003 and it will be necessary to make comparisons between both surveys. One of the ways to perform it is to build the 2012 sample of schools with overlapping conditions with the 2003 sample. But it is also necessary to have the best representativeness for the new sample. This one can be met introducing balancing conditions when the new sample is drawn. Other constraints should be prescribed (fixed size, given inclusion probabilities...). The main tools used are first conditional successive samples and, secondly, balancing techniques. But it will be shown that only approached solutions can be reached, not only from a statistical point of view, but also from a computational one, to obtain numerical solutions. After developing the theoretical approach, results on the French sampling frame of schools will be given.

OPTIMIZATION OF SAMPLE SIZE AND NUMBER OF TASKS PER RESPONDENT IN CONJOINT STUDIES USING SIMULATED DATASETS

Ondřej Vilikus

Broader use of hierarchical Bayesian models to estimate part-worth utilities in conjoint studies has allowed market researchers to get reasonably accurate estimates of individual respondent behavior without torturing respondents by demanding too many tasks or by imposing complicated and unrealistic scenarios. On the other hand it also made it more complicated to estimate before running the survey how many attributes and attribute levels can we test with given sample size and what would be the costs for increasing any of these. In the article I present an approach based on analyzing batches of simulated datasets with given characteristics. The article includes overview of the results for choice-based conjoint studies with usual level of complexity. Search for an optimal combination of sample size and number of tasks per respondent that allows us to achieve required accuracy of our outputs with optimal cost is of main focus but sensitivity of the recommendations with respect to changes in fixed parameters of the datasets is also included.

ROTATION SCHEMES AND CHEBYSHEV POLYNOMIALS

Jacek Wesółowski

In the seminal paper on rotation schemes Patterson (1950) derived recursion for optimal estimator of the mean on a given occasion for patterns with no “holes”. In such schemes a unit leaving the sample on a certain occasion, never returns. Analogous recursion formula for patterns with “holes” had not been available for many years. Nevertheless such schemes have been used in regular surveys, e.g. Current Population Survey in the US with pattern 4-8-4 (see e.g Rao and Graham, 1964, Ciepiela et. al, 2011) or Labour Force Surveys in several European countries with pattern 2-2-2 (e.g. Szarkowski, Witkowski 1994, Popiński, 2006, Wesółowski, 2011). Consequently, non-admissible estimators were used for such surveys. Recently, the recursion problem has been overcome in Kowalski and Wesółowski (2010) under two technical assumptions related to properties of certain polynomial P and solvability of certain linear system. Numerical experiments suggested that these assumptions may be universally satisfied. A very recent progress is that the first assumption is always satisfied. The proof of this fact is heavily based on representation of the polynomial P through Chebyshev polynomials and their miraculous properties. In a special case of 2-2-2 scheme it has been proved recently (Wesółowski, 2010) that both the assumptions are satisfied. Also the question for patterns with gaps of size 1 is completely resolved, see Kowalski (2009).

References

- Ciepiela, P., Gniado, M., Wesółowski, J., Wojtyła, M., Dynamic K-composite estimator for an arbitrary rotation scheme. *Statist. Trans.* (2011) - to appear.
- Kowalski, J., Optimal estimation in rotation patterns. *J. Statist. Plan. Infer.* 139(4) (2009), 2429-2436
- Kowalski, J., Wesółowski, J., Recurrence optimal estimators for rotation cascade patterns with holes (2010) - submitted.
- Patterson, H.D., Sampling on successive occasions. *J. Roy. Statist. Soc. B* 12 (1955), 241-255.
- Popiński, W., Development of the Polish Labour Force Survey. *Statist. Trans.* 7(5) (2006), 1009-1030.
- Rao, J.N.K., Graham, J.E., Rotation designs for sampling on repeated occasions. *Ann. Math. Statist.* 35 (1964), 492-502.
- Szarkowski, A., Witkowski, J., The Polish Labour Force Survey. *Statist. Trans.* 1(4) (1994), 467-483.
- Wesółowski, J., Recursive optimal estimation in Szarkowski rotation scheme. *Statist. Trans.* 11(2) (2010), 267-285.

ON DISTRIBUTION OF HORVITZ-THOMPSON STATISTIC UNDER THE REJECTIVE SAMPLE

Janusz L. Wywiał

There are several sampling designs defined on the basis of the value of a non-negative auxiliary variable. Especially, the sampling designs and scheme with predefined inclusion probabilities proportional to a value of the non-negative auxiliary variable are useful in survey sampling. A review of them is presented by Tillé (2006). The paper deals with the limit properties of the Horvitz-Thompson (H-T) (1952) statistic under the rejective sampling design. On the basis of the papers by Berger and Skinner (2005) and Hájek (1964) we consider the limit distribution of Horvitz-Thompson statistic standardized by its sample variance. Moreover, the variance of the H-T estimator is considered under the assumption that the auxiliary variable value is the observation of the variable under study but with measuring error.

References

- Berger Y. G., C. J. Skinner (2005). A jackknife variance estimator for unequal probability sampling. *Journal of the Royal Statistical Society B* 67, 79-89.
- Hájek J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics* 35, 1491-1523.
- Horvitz D. G., Thompson D. J. (1952). A generalization of sampling without replacement from finite universe. *Journal of the American Statistical Association*, vol. 47, 663-685.
- Tillé Y. (2006). *Sampling Algorithms*. Springer.

ON ACCURACY OF TWO PREDICTORS FOR SPATIALLY CORRELATED LONGITUDINAL DATA

Tomasz Żądło

In the paper considerations are based on the assumption of some longitudinal model which is a special case of the general linear model (GLM) and the general linear mixed model (GLMM). In the model two random components are included under assumptions of spatial autoregressive process and temporal AR(1) process respectively. The accuracy of two predictors is compared in the simulation study for longitudinal data assuming the model. The first predictor under study is the empirical best linear unbiased predictor (EBLUP) derived for this model. The second predictor is derived under the assumption of lack of spatial and temporal correlation. Monte Carlo simulation analysis is prepared using R package.



**INFORMATION ON THE CONGRESS OF POLISH STATISTICS
TO CELEBRATE THE 100TH ANNIVERSARY
OF THE POLISH STATISTICAL ASSOCIATION**

In 2012 we will be celebrating **the 100th anniversary of the Polish Statistical Association (PSA)**, an organization created to integrate specialists involved in public statistical services as well as representatives of the academic community, local and economic government and agencies of state administration interested in the theory and implementation of statistical research. The Association contributes to the development of theoretical, methodological and practical aspects of statistical research and tries to promote statistical knowledge in society. It maintains cooperation with statistical associations in other countries and such organizations as Bernoulli - Society for Mathematical Statistics and Probability, International Society for Quality of Life Research, International Society for Quality-of-Life Studies or International Federation of Classification Societies. Polish Statistical Association is an affiliated member of International Statistics Institute.

To celebrate the 100th anniversary of The Polish Statistical Association we intend to hold the Congress of Polish Statistics on 18 – 20 April 2012 in Poznan and combine this event with the celebration of Polish Statistics Day in 2012. The Congress is organized by: Polish Statistical Association, Central Statistical Office, Poznan University of Economics and Statistical Office in Poznan.

The preliminary programme of the Congress comprises a number of thematic sessions including the anniversary (historical) session, as well as other ones devoted to the methodology of statistical research, regional statistics, population statistics, socio-economic

statistics, the problems of statistical data and the statistics of health, sport and tourism. The Congress will also host two panel discussions on:

- fundamental problems of statistics in the modern world
- the future of statistics

The Organizing Committee would like to invite everyone to join in the unique celebration of Polish statistics and take part in the Congress of Polish Statistics. Updated information about the conference are published on the Congress website: <http://www.stat.gov.pl/pts/kongres2012/english/index.htm>.

We welcome paper proposals on various aspects of statistical research, both theoretical and applied. Paper abstracts should be attached with the participation form and submitted on-line or sent by email at the address of the Congress secretariat by 10.10.2011.

Chairperson of the Organizing Committee of the Congress:

Dr. Elżbieta Gołata, Univ. Prof of UEP: e-mail: elzbieta.golata@ue.poznan.pl

Secretariat of the Organizing Committee of the Congress:

Statistical Office in Poznan,

ul. J. H. Dąbrowskiego 79, 60-959 Poznań: e-mail: kongres2012@stat.gov.pl

tel. + 48 61 27 98 325 (Mo-Fr: 8-15),

+48 61 27 98 343 (Mo: 9-12, Tu: 7-10, We: 7-10),

fax. +48 61 27 98 101

LIST OF PARTICIPANTS

Yves Berger, University of Southampton

Józef Biolik, University of Economics in Katowice

Magdalena Byrska, University of Economics in Katowice

Magdalena Chmielińska, University of Economics in Katowice

Jan Czempas, University of Economics in Katowice

Jean-Claude Deville, Ecole Nationale de la Statistique et de l'Analyse de l'Information, Laboratoire de Statistique d'Enquete

Czesław Domański, University of Economics in Katowice

Leon Dziembala, formerly University of Economics in Katowice

Nikolaos Farmakis, Aristotle University of Thessaloniki

Wojciech Gamrot, University of Economics in Katowice

Jakub Glinka, GFK Polonia

Malay Ghosh, University of Florida

Stanisław Heilpern, University of Economics in Wrocław

Anna Imiolek, University of Warmia and Mazury

Alina Jędrzejczak, University of Łódź

Partha Lahiri, University of Maryland, College Park, USA

Nicholas T. Longford, SNTL and UPF, Barcelona, Spain

Małgorzata Kaczanowicz, formerly University of Economics in Katowice

Tomasz Klimanek, University of Economics in Poznań

Józef Kolonko, University of Economics in Katowice

Roman Konarski, University of Gdańsk

Grzegorz Kończak, University of Economics in Katowice

Jan Kordos, formerly Warsaw School of Economics

Barbara Kowalczyk, Warsaw School of Economics

Arkadiusz Kozłowski, University of Gdańsk

Danute Krapavickaite, Vilnius Gediminas Technical University, Institute of Mathematics and Informatics, Vilnius University

Sabine Krieg, Statistics Netherlands

Jan Kubacki, Statistical Office in Łódź

Magdalena Makuch, University of Economics in Katowice

Christine Marc, National Institute of Statistics (INSEE), Paris.

Dorota Olszewska, Statistical Office in Kraków

Józef Pociecha, University of Economics in Kraków

Dorota Rozmus, University of Economics in Katowice

Małgorzata Szerszunowicz, University of Economics in Katowice

Merijn van Pelt, Statistics Netherlands

Ondřej Vilikus, University of Economics, Prague

Marta Welfe, University of Łódź

Jacek Wesółowski, Central Statistical Office, Warsaw

Ewa Witek, University of Economics in Katowice

Iwona Wrona, Statistical Office in Kraków

Janusz L. Wywiał, University of Economics in Katowice

Tomasz Źądło, University of Economics in Katowice