STATISTICS
IN TRANSITION
new series

An International Journal of the Polish Statistical Association and Statistics Poland

IN THIS ISSUE:

EDITORIAL OFFICE

CONTENTS

From the Editor

Research Articles

The summer issue of the *Statistics in Transition new series* is extended by an additional special section dedicated to the ceremony and celebratory eulogies held on the occasion of awarding the *honoris causa* doctorate to Professor Malay Ghosh, a member of our journal's Editorial Board, by the University of Economics of Katowice.

Eight original *research articles* and one paper from the 2019 Multivariate Statistical Analysis conference (in Łódź, 2019), and one *research communicate* constitute the issue, which cover a wide range of problems concerning statistical theory and methods, along with their applications to a variety of real-life problems, in different contexts.

In the first paper ***Extended Residual coherence with a financial application*** **Xuge Zhang** and **Benjamin Kedem** present the extension of the notion of residual coherence as a graphical tool for selecting potential second-order interaction terms as functions of a single time series and its lags to account for interaction terms of multiple time series. The authors also propose an integrated spectrum as an alternative criterion to facilitate the graphical selection. They employ the regression-based selection method to search significant covariate interactions. The results concerning financial market application suggest that daily increments of implied volatility of the stock market are possibly influenced by products of the daily increments (and their lags) of implied volatility of commodity markets. The authors confirm the need for the further exploration of the essential factors in the relationship between the implied volatility of stock market and certain commodity markets.

The next article, by **Berislav Žmuk**, entitled ***Estimating completion and breakoff functions in a business survey***, starts with the observation about the importance of the business survey length – it should not be too long as the breakoff rate in this case tends to be high, resulting in a low response rate. In this paper, completion and breakoff times are observed and compared across different questionnaire and respondent characteristics. A regression modelling approach has been adopted to estimate the completion and breakoff functions to help a researcher determine which respondents completed a questionnaire and which broke it off too quickly or too slowly. By omitting such respondents, a researcher is able to obtain the relevant estimates more efficiently. In addition, the completion and breakoff functions offer a better insight into the completion and breakoff development rates, allowing the researcher to make a better-

informed decision as to whether the survey requires any modifications or not. While the regression diagnostic results have shown that the estimated completion and breakoff functions are of good fit, the estimates are valid only for the observed survey. The author also emphasizes the limitation of the proposed approach, which is the need of a pilot study.

The article ***Credibility of disability estimates from the 2011 population census in Poland*** by **Elżbieta Gołata** and **Grażyna Dehnel** discusses the problem of statistics on disability – a phenomenon considered by the authors as one of the most important due to, inter alia, aging of the population. Data on disability are collected through numerous statistical surveys and censuses. The main objective of the study is to assess the quality of estimates relating to the number of disabled persons obtained on the basis of the 2011 census data. The study aims at identifying discrepancies between the estimates, and determining the size and source of these discrepancies taking into account such aspects as the measurement methods, the definitions and criteria of disability, the voluntary nature of the question, and the quality of the information on disability obtained from various sources. The analysis shows that a high degree of non-response results from the voluntary nature of the question, causing a major bias in data and 14% decrease in the number of people with disabilities. The introduction of stricter criteria of disability assessment also discouraged many potential applicants from applying for a disability benefit, which led to a fall in the actual size of this subpopulation and its estimates in the year 2011 in the legal sense. The authors conclude that new administrative regulations affect the system of disability assessment and decisions made by assessors, which is reflected in statistical data.

In the next paper, ***Interviewer allocation through interview–reinterview nested design for response error estimation in sample surveys***, **Fidan Mahmut Fahmi, H. Öztaş Ayhan** and **İnci Batmaz** consider the problem of quantification of non-sample errors in surveys compared to sampling errors. These errors are challenging due to their complex nature but avoiding them results in biased survey estimates. The authors applied nested experimental design in interview-reinterview surveys relating to the time use and life satisfaction of academicians at Middle East Technical University, Turkey. They investigated response errors concerning the respondent, the interviewer or their interaction. The results show that response variances are usually revealed in the questions hard to quantify, especially feelings, such as satisfaction level, therefore showing higher interviewer effect. The authors also present the considerable reduction of the associated response variances accounting for the total variance in the main survey as a result of the specific raising awareness training provided to the interviewers immediately after the pilot survey analysis. They also conclude that a researcher must keep in mind respondent fatigue and memory limitations.

Praveen Kumar Tripathi, Rijji Sen and **S. K. Upadhyay** in their article *A Bayes algorithm for model compatibility and comparison of ARMA(p;q) models* present a Bayes analysis of an autoregressive-moving average model and its components based on exact likelihood and weak priors for the parameters where the priors are defined so that they incorporate stationarity and invertibility restrictions naturally. They use a Gibbs-Metropolis hybrid scheme to draw posterior based inferences for the models under consideration and examine the compatibility of the models with the data using the Ljung-Box-Pierce chi-square-based statistic. The paper also compares different compatible models through the posterior predictive loss criterion in order to recommend the most appropriate one. For a numerical illustration of the above, data on the Indian gross domestic product growth rate at constant prices are considered. Differencing the data once prior to conducting the analysis ensured their stationarity. Retrospective short-term predictions of the data are provided based on the final recommended model. The considered methodology is expected to offer an easy and precise method for economic data analysis. A short-term retrospective prediction based on the final chosen model conveys that the proposed model can be used, in general, except when there is abrupt fluctuation in the data from those of previous years.

In the next article *Developing calibration estimators for population mean using robust measures of dispersion under stratified random sampling* by **Ahmed Audu, Rajesh Singh** and **Supriya Khare** two modified, design-based calibration ratio-type estimators are presented. They were developed under stratified random sampling using information on an auxiliary variable in the form of robust statistical measures, including Gini's mean difference, Downton's method and probability weighted moments. The quality properties of the proposed estimators are checked up to the terms of first-order approximation by means of Taylor's series approximation. Their theoretical results were supported by a simulation study conducted on four bivariate populations and generated using normal, chi-square, exponential and gamma populations. According to the results of the study, the estimators proposed under both calibration schemes are not only robust but more efficient than the usual ratio estimator in stratified sampling, making them applicable in real life situation when data is somewhat affected by the presence of extreme values.

Tomasz Bąk's article *Spatial sampling methods modified by model use* addresses the vibrant field of spatial sampling methods and adaptive sampling methods as one of the dynamic trends in the sampling theory. The author analyses five of widely known spatial sampling methods while designing the experiment using artificial data to including statistical model in the sampling procedure. As in the case of adaptive methods, it serves to modify drawing probabilities during sampling. The experiment resulted in the improvement of the quality of method by the model modification.

The author reaches a conclusion that model modification can also be used for other design-based methods, and that from a theoretical point of view, the presented solution can easily be translated into other methods and represents quite an effective method, which expands the range of design based methods. The author also states that the real-time observation and analysis of the sample represent an interesting direction in the development of sampling methods. The sampling modification gives a possibility to adjust the sampling method to the analysed population and its different characteristics.

Łukasz Wawrowski and **Maciej Beręsewicz** in the last research article entitled ***Small area estimates of the low work intensity indicator at voivodeship level in Poland*** argue that model-based small area estimation can be used to obtain direct estimates of labour market statistics at low levels based on the survey the EU Statistics on Income and Living Conditions. The authors estimated the low work intensity indicator for the spatial domains in Poland between years 2005 and 2012 applying several models, including Rao and You (1994), Fay and Diallo (2012), and Marhuenda, Molina and Morales (2013). They also proposed Bootstrap MSE for the discussed methods. All this allowed to obtain more reliable (in the sense of CV) estimates in previously unpublished domains.

<div align="center">**Other articles**</div>

Among *other articles* there is a conference paper from the Multivariate Statistical Analysis conference held in Łódź in 2019, entitled ***A dynamic MST-deltaCoVaR model of systemic risk in the European insurance sector*** by **Aneta Denkowska** and **Stanisław Wanat**. The authors analyse the contribution of each of the 28 largest European insurance companies, including those appearing on the G-SIIs list, to systemic risk and aims to determine whether the most important contribution to systemic risk is made by companies with the highest betweenness centrality or the highest degree in the obtained MST. Using time series analysis they show that in the period from 2005 to 2019 for each of the companies there is an obvious relation between their contribution to systemic risk and the structure of the network of connections (MST). During the entire period, the contribution of each company remains at the same level, save for the clearly apparent period during which the deltaCoVaR decreases and, consequently, the contribution to the systemic risk increases, and this happens at the very centre of the subprime crisis, October 17th, 2008. As the deltaCoVaR changes, the APL ratio increases. The authors emphasized that for the entire period under study, it reaches its maximum exactly on December 5th, 2008. Also, the authors proposed a relationship between the contribution to systemic risk and the minimum spanning tree structure described by topological network indicators to be used in the construction of models whose task is to predict the possibility of systemic risk.

Research Communicates

Brij Behari Khare, Ashutosh and **Piyush Kant Rai** in their *research communicate* entitled *A comparative study of a class of direct estimators for domain mean with a direct ratio estimator for domain mean using auxiliary character* present the theoretical aspects of the proposed class of direct estimators for domain mean with the use of a single auxiliary character. The results for MSE supported the superiority of the proposed estimators theoretically as compared to the direct ratio estimator. The authors' finding prove the proposed estimators outperform the direct ratio estimator for domain mean using a single auxiliary character in the case of two studied populations and their analysed domains considered from Sarndal et al. (1992). They recommend that the class of direct estimators proposed in this article for the estimation of domain mean using proper auxiliary information have substantial utility in the domain estimation methodology as compared to the existing direct ratio estimator under the condition that a sufficient member of units fall in the domain concerned.

Section dedicated to Malay Ghosh's *honoris causa* ceremony

It is with great satisfaction and pride that we are closing this issue with a section dedicated to the ceremony of awarding *honoris causa* to Professor Malay Ghosh, our invaluable colleague and member of the journal's team of scientific advisors. The section contains following items: *Introduction* by **Her Magnificence Rector Celina Olszak** and **V-Rector Wojciech Dyduch – reviews** by **Yves G. Berger** and **Ralf Münnich – presentation/laudation** by **Janusz L. Wywiał** and congratulations by **invited speakers:** Debashis Ghosh, Jr; Robert Tomanek; Dominik Rozkrut; Waldemar Tarczyński; Krzysztof Jajuga; Włodzimierz Okrasa; Nicholas T. Longford; and **congratulatory letters** from Graham Kalton; Carl-Eric Särndal; Czesław Domański; Elżbieta Gołata; Tomasz Zjawiony; Adam Weintrit.

The lecture given by the laureate was focused on *Small area estimation: its evolution in five decades* – "a topic that was near and dear to me for more than three decades", said Malay Ghosh at the outset, drawing extensively from the Invited Paper published in Special Issue of Statistics in Transition new series, organized by Partha Lahiri (SiTNs, Vol. 21 (4)).

Włodzimierz Okrasa
Editor

Submission information for Authors

Statistics in Transition new series (SiT) is an international journal published jointly by the Polish Statistical Association (PTS) and Statistics Poland, on a quarterly basis (during 1993–2006 it was issued twice and since 2006 three times a year). Also, it has extended its scope of interest beyond its originally primary focus on statistical issues pertinent to transition from centrally planned to a market-oriented economy through embracing questions related to systemic transformations of and within the national statistical systems, world-wide.

The SiT-*ns* seeks contributors that address the full range of problems involved in data production, data dissemination and utilization, providing international community of statisticians and users – including researchers, teachers, policy makers and the general public – with a platform for exchange of ideas and for sharing best practices in all areas of the development of statistics.

Accordingly, articles dealing with any topics of statistics and its advancement – as either a scientific domain (new research and data analysis methods) or as a domain of informational infrastructure of the economy, society and the state – are appropriate for *Statistics in Transition new series.*

Demonstration of the role played by statistical research and data in economic growth and social progress (both locally and globally), including better-informed decisions and greater participation of citizens, are of particular interest.

Each paper submitted by prospective authors are peer reviewed by internationally recognized experts, who are guided in their decisions about the publication by criteria of originality and overall quality, including its content and form, and of potential interest to readers (esp. professionals).

Manuscript should be submitted electronically to the Editor:
sit@stat.gov.pl,
GUS/Statistics Poland,
Al. Niepodległości 208, R. 296, 00-925 Warsaw, Poland

It is assumed, that the submitted manuscript has not been published previously and that it is not under review elsewhere. It should include an abstract (of not more than 1600 characters, including spaces). Inquiries concerning the submitted manuscript, its current status etc., should be directed to the Editor by email, address above, or w.okrasa@stat.gov.pl.

For other aspects of editorial policies and procedures see the SiT Guidelines on its Web site: http://stat.gov.pl/en/sit-en/guidelines-for-authors/

Editorial Policy

The broad objective of *Statistics in Transition new series* is to advance the statistical and associated methods used primarily by statistical agencies and other research institutions. To meet that objective, the journal encompasses a wide range of topics in statistical design and analysis, including survey methodology and survey sampling, census methodology, statistical uses of administrative data sources, estimation methods, economic and demographic studies, and novel methods of analysis of socio-economic and population data. With its focus on innovative methods that address practical problems, the journal favours papers that report new methods accompanied by real-life applications. Authoritative review papers on important problems faced by statisticians in agencies and academia also fall within the journal's scope.

***

Abstracting and Indexing Databases

Statistics in Transition new series is currently covered in:

Databases indexing the journal:

- BASE – Bielefeld Academic Search Engine
- CEEOL – Central and Eastern European Online Library
- CEJSH (The Central European Journal of Social Sciences and Humanities)
- CNKI Scholar (China National Knowledge Infrastructure)
- CNPIEC – cnpLINKer
- CORE
- Current Index to Statistics
- Dimensions
- DOAJ (Directory of Open Access Journals)
- EconPapers
- EconStore
- Electronic Journals Library
- Elsevier – Scopus
- ERIH PLUS (European Reference Index for the Humanities and Social Sciences)
- Genamics JournalSeek
- Google Scholar
- Index Copernicus
- J-Gate
- JournalGuide
- JournalTOCs
- Keepers Registry
- MIAR
- Microsoft Academic
- OpenAIRE
- ProQuest – Summon
- Publons
- QOAM (Quality Open Access Market)
- ReadCube
- RePec
- SCImago Journal & Country Rank
- TDNet
- Technische Informationsbibliothek (TIB) - German National Library of Science and Technology
- Ulrichsweb & Ulrich's Periodicals Directory
- WanFang Data
- WorldCat (OCLC)
- Zenodo

Extended residual coherence with a financial application

Xuze Zhang,[1] Benjamin Kedem[2]

ABSTRACT

Residual coherence is a graphical tool for selecting potential second-order interaction terms as functions of a single time series and its lags. This paper extends the notion of residual coherence to account for interaction terms of multiple time series. Moreover, an alternative criterion, integrated spectrum, is proposed to facilitate this graphical selection. A financial market application shows that new insights can be gained regarding implied market volatility.

Key words: interaction, residual coherence, nonlinear, time series, volatility index.

1. Introduction

Nonlinear phenomena in random processes have attracted much attention going back to the work of Wiener (1958) concerning random nonlinear oscillators excited by a random input, random shot effect as input for testing nonlinear circuits, and more generally concerning a class of nonlinear polynomial functionals to model input-output relationships in nonlinear systems. In Wiener's words, he was interested in "methods of handling the spectrum," which motivates the use of higher order spectra dealt with by quite a few authors including Brillinger (1965), Brillinger and Rosenblatt (1967), Hinich (1979), Nikias and Mendel (1993), and Elgar et al. (1998). The excellent review paper by Sanaullah (2013) provides numerous additional references about applications of nonlinear techniques based on higher order spectra. Inherent in all nonlinear systems is the problem of assessing the degree and extent of nonlinearity, which can be approached by the detection of nonlinear components or *interactions* (Tick (1961), Elgar et al. (1998)).

In this paper, the detection of nonlinear second-order interactions is done by an extension of *residual coherence* introduced in Khan, Katzoff and Kedem (2014) and applied in mortality forecasting. Residual coherence is a nonlinear variation of the well-known measure of linear coherence. The method is then applied to two volatility indices, the Chicago Board Options Exchange Volatility Index (VIX), and the Russell 2000 Volatility Index (RVX).

---

[1]Department of Mathematics and Institute for Systems Research, University of Maryland, USA. E-mail: xzhang51@umd.edu. ORCID: https://orcid.org/0000-0002-8672-8515.

[2]Department of Mathematics and Institute for Systems Research, University of Maryland, USA. E-mail: bnk@umd.edu. ORCID: https://orcid.org/0000-0002-7945-8713.

2. Extensions of residual coherence

2.1. Preliminaries

The coherence between two time series $(X(t), Y(t))$ measures the extent of linear relationship between them in the frequency domain. Provided all auto- and cross-spectra exist, it is defined as

$$\gamma_{XY}(\lambda) = \frac{|f_{XY}(\lambda)|^2}{f_{XX}(\lambda)f_{YY}(\lambda)}$$

(see Koopmans (1974)) where f_{XX} and f_{YY} are the spectra of $X(t)$ and $Y(t)$, respectively, and f_{XY} is the cross-spectral density of $X(t)$ and $Y(t)$. This is widely used in detecting connections and clustering of time series. Relevant works include Sun, Miller and D'Esposito (2004), Maharaj and D'Urso (2010) and Euan, Sun and Ombao (2019), among many others. When the relationship is nonlinear, it is frequently analyzed by bispectra, trispectra, or higher-order spectra. For example, a bispectral method for detecting lag processes was proposed by Hinich (1979). Lagged coherence and residual coherence were first introduced in Kedem-Kimelfeld (1975) and Khan, Katzoff and Kedem (2014), respectively, to detect and select potential interaction effects as input to nonlinear systems, based on an orthogonal decomposition in Kimelfeld (1974) without involving bispectrum or higher-order spectra.

Let $Y(t)$ be the output of a system of which the input consists of linear and quadratic filters of $X(t)$ plus noise $\varepsilon(t)$,

$$Y(t) = L[X(t)] + \sum_{k=1}^{\infty} L_{u_k}[\tilde{X}_{u_k}(t)] + \varepsilon(t)$$

where $\tilde{X}_{u_k}$ is a lag process defined as $\tilde{X}_{u_k}(t) = X(t)X(t-u_k) - E[X(t)X(t-u_k)]$. For simplicity, assume that $Y(t)$ and $X(t)$ are zero-mean real valued jointly stationary processes and that all relevant auto- and cross-spectra exist. Then, for sufficiently large n, $Y(t)$ can be approximated by

$$Y^*(t) = G_1(t) + \sum_{k=1}^{n} G_{2,k}(t) + \varepsilon(t) \tag{1}$$

where $G_1(t)$ is a linear filter of $X(t)$, and as in Kedem-Kimelfeld (1975), $G_{2,k}(t)$ is a sum of a linear filter of $X(t)$ and a linear filter of $\tilde{X}_{u_k}(t)$, such that $G_{2,k}(t) \perp G_1(t)$ for $k = 1, \ldots, n$.

Kedem-Kimelfeld (1975) showed that if there is prior knowledge that $Y(t)$ takes on a simpler form

$$Y(t) = L[X(t)] + L_u[\tilde{X}_u(t)] + \varepsilon(t), \tag{2}$$

then it can be rewritten as a sum of two orthogonal processes, $G_1(t)$ and $G_2(t;u)$ plus noise $\varepsilon(t)$,

$$Y(t) = G_1(t) + G_2(t;u) + \varepsilon(t). \tag{3}$$

Then the lag process, or interaction, $\tilde{X}_u(t)$ that minimizes $E\varepsilon^2(t)$ can be selected by finding the lag u that maximizes the lagged coherence $S_2(\lambda;u)$ over all frequencies $\lambda \in [-\pi, \pi]$

such that

$$S_2(\lambda;u) = \frac{f_{G_1G_1}(\lambda) + f_{G_2G_2}(\lambda;u)}{f_{YY}(\lambda)}.$$

However, as mentioned in Kedem-Kimelfeld (1975), there might not exists a u that maximizes $S_2(\lambda;u)$ over all frequencies. One way to resolve this issue is to define the residual coherence as

$$RC(u) = \sup_\lambda \frac{f_{G_2G_2}(\lambda;u)}{f_{YY}(\lambda)}$$

and find u that maximizes $RC(u)$. This is shown to be useful for interaction selection in Khan, Katzoff and Kedem (2014) and Kedem (2016).

2.2. Lagged coherence and residual coherence for more than two orthogonal components

Consider the model

$$Y(t) = \sum_{k=1}^n L_{k,u_k}[X_{k,u_k}(t)] + \varepsilon(t). \tag{4}$$

The goal is to select $X_{k,u_k}(t)$ from a certain family of processes $\{X_{k,u_k}(t) : u_k = 1,2,\dots\}$ for $k = 1,\dots,n$. Assume that all relevant series are jointly stationary and all relevant auto- and cross-spectra exist. This reduces to (2) when $n = 2$ and $L_{1,u_1}[X_{1,u_1}(t)]$ is the linear filter of $X(t)$. For $n > 2$, we shall extend the orthogonal decomposition (3)

$$Y(t) = \sum_{k=1}^n G_k(t;u_1,\dots,u_k) + \varepsilon(t)$$

where all G_k's for $k = 1,\dots,n$ are mutually orthogonal, to account for more orthogonal components, given by

$$G_k(t;u_1,\dots,u_k) = \sum_{j=1}^k \int_{-\pi}^\pi e^{it\lambda} A_{j,k-j+1}(\lambda) dZ_{X_{j,u_j}}(\lambda)$$

for $k = 1,\dots,n$, where the A's are non-zero and Z's are the corresponding spectral measures.

The A's can be obtained by using the orthogonal conditions among G_k's such that

$$A_{k,1}(\lambda) = \overline{\left[\frac{\sum_{j=1}^k c_{k,j}(\lambda) f_{X_{j,u_j}Y}(\lambda)}{\sum_{j=1}^k c_{k,j}(\lambda) f_{X_{j,u_j}X_{k,u_k}}(\lambda)}\right]} \qquad k = 1,\dots,n$$

and

$$A_{j,k-j+1}(\lambda) = c_{k,j}(\lambda) A_{k,1}(\lambda) \qquad j = 1,\dots,k, \qquad k = 1,\dots,n$$

where

$$c_{k,k}(\lambda) = 1, c_{k,j}(\lambda) = \frac{\boldsymbol{F}_{k,j}(\lambda)}{\boldsymbol{F}_k(\lambda)}, \boldsymbol{F}_k(\lambda) = (f_{i,j}(\lambda))_{(k-1)\times(k-1)},$$

$f_{i,j} \equiv f_{X_{j,u_j} X_{i,u_i}}$ and $\boldsymbol{F}_{k,j}(\lambda)$ is equivalent to $\boldsymbol{F}_k(\lambda)$, of which jth column is replaced by

$$\boldsymbol{f}_k(\lambda) \equiv -[f_{1,k}(\lambda), \ldots, f_{k-1,k}(\lambda)]^T.$$

More details are provided in Appendix.

Subsequently,

$$f_{G_k G_k}(\lambda; u_1, \ldots, u_k) = |A_{k,1}(\lambda)|^2 \left[\sum_{j=1}^{k} c_{k,j}(\lambda) f_{X_{j,u_j} X_{k,u_k}}(\lambda) \right]$$

$$S_k(\lambda; u_1, \ldots, u_k) = \frac{\sum_{j=1}^{k} f_{G_j G_j}(\lambda; u_1, \ldots, u_j)}{f_{YY}(\lambda)} \tag{5}$$

$$RC(u_1, \ldots, u_k) = \sup_{\lambda}[S_k(\lambda; u_1, \ldots, u_k) - S_m(\lambda; u_1, \ldots, u_m)] \quad 1 \le m \le k$$

for $k = 1, \ldots, n$. Note that $S_k(\lambda; u_1, \ldots, u_k)$ and $RC(u_1, \ldots, u_k)$ depend only on $u_{m+1}, \ldots, u_k$ once $u_1, \ldots, u_m$ are determined. The estimates of the above quantities are obtained based on the estimates of the relevant auto- and cross-spectra. Also, to avoid confusion, we denoted the residual coherence in (5) by $RC_{(m+1):k}(u_{m+1}, \ldots, u_k)$ when $u_1, \ldots, u_m$ are determined.

2.3. Selection Criteria

In this section, lagged coherence and residual coherence are examined and an alternative criterion is proposed. Take $n = 2$ and fix u_1, then it reduces to the case in Kedem-Kimelfeld (1975). It illustrates that if there exists a u_2 that maximizes $S_2(\lambda; u_2)$ for all λ, then such u_2 minimizes $E\varepsilon^2(t)$ in

$$E\varepsilon^2(t) = \int_{-\pi}^{\pi} f_{\varepsilon\varepsilon}(\lambda)d\lambda = \int_{-\pi}^{\pi} f_{YY}(\lambda)[1 - S_2(\lambda; u_2)]d\lambda.$$

Indeed, the quantity we wish to maximize is $\int_{-\pi}^{\pi} f_{G_2 G_2}(\lambda; u_2)d\lambda$ since

$$\int_{-\pi}^{\pi} f_{YY}(\lambda) S_2(\lambda; u_2)d\lambda = \int_{-\pi}^{\pi} [f_{G_1 G_1}(\lambda) + f_{G_2 G_2}(\lambda; u_2)]d\lambda$$

based on (5). Such criterion works even if such u_2 does not exist so that this can be an alternative to residual coherence. This criterion can be readily extended to a more general case. Suppose there is prior knowledge for the inclusion of first m processes, i.e. $u_1, \ldots, u_m$ are fixed, then we define the integrated spectrum

$$IS_{(m+1):n}(u_{m+1}, \ldots, u_n) \equiv \int_{-\pi}^{\pi} \sum_{k=m+1}^{n} f_{G_k G_k}(\lambda; u_k, \ldots, u_n)d\lambda$$

and find $u_{m+1}, \ldots, u_n$ that maximizes $IS_{(m+1):n}(u_{m+1}, \ldots, u_n)$.

Once all u's are determined, the regression-based selection method proposed in Khan, Katzoff and Kedem (2014) and Kedem (2016) is used to select significant terms within the processes selected by the graphical method and this is illustrated in both Section 3 and

Section 4. Relevant regression for time series that the selection entails can be found in Kedem and Fokianos (2002).

3. Simulation

In this section, a simulation is performed with $n = 4$ and $u1, u2$ fixed to validate and compare the two criteria, residual coherence and integrated spectrum. The steps are as follows:

1. Generate $\{x_1(t)\}_{t=1}^{1010}$ from an AR(1) process $X_1(t) = 0.4X_1(t-1) + u_1(t)$ and $\{x_2(t)\}_{t=1}^{1010}$ from an AR(1) process $X_2(t) = 0.2X_2(t-1) + u_2(t)$, where u's are white noise $N(0,1)$.

2. Obtain

$$y(t) = 0.4x_1(t) + 0.3x_2(t) + 0.4x_1(t-2)x_2(t-1) + 0.3x_1(t)x_2(t-4) + \varepsilon(t), \quad (6)$$

where ε's are white noise $N(0,1)$ and t=11,...,1010 so that all relevant series have length 1000.

3. This is the model (4) with $n = 4$ and known $X_1(t)$, $X_2(t)$. We considered selecting $X_{3,u_3}(t)$ and $X_{4,u_4}(t)$ from the family $\{X_1(t+h)X_2(t) : h = -9, -8, \ldots, 0, \ldots, 9\}$. In fact, this family can be made larger and the choice here only serves as an example. Then, we estimated all relevant auto- and cross-spectra using Tukey-Hamming kernel with window size 10 for frequencies $\lambda_k = -\pi + k\pi/1000$, $k = 0, \ldots, 2000$. Subsequently, we estimated $RC_{3:3}(u_3)$, $IS_{3:3}(u_3)$, $RC_{4:4}(u_4)$ and $IS_{4:4}(u_4)$ for $u_3, u_4 = -9, -8, \ldots, 0, \ldots, 9$. Note that

$$\widehat{IS}_{3:3}(u_3) = \sum_{k=1}^{2000} \pi \hat{f}_{G_3 G_3}(\lambda_k; u_3)/1000$$

$$\widehat{IS}_{4:4}(u_4) = \sum_{k=1}^{2000} \pi \hat{f}_{G_4 G_4}(\lambda_k; u_3, u_4)/1000$$

and $RC_{4:4}(u_4)$, $IS_{4:4}(u_4)$ only depend on u_4 once u_3 is fixed.

The results are shown by Figure 1, which indicates that the process $X_1(t-1)X_2(t)$ is the optimal choice for the third input. It is also observed from Figure 1 that $X_1(t+4)X_2(t)$ is another potential input since the bars that correspond to $u_3 = 4$ are the second highest ones in both graphs. With $u_3 = -1$ fixed, u_4 can be determined by $\widehat{RC}_{4:4}(u_4)$ and $\widehat{IS}_{4:4}(u_4)$, as shown by Figure 2, and both graphs indicate that $u_4 = 4$ is the optimal choice, which accords with the original model (6).

With u_3 and u_4 determined, we select significant covariates from the selected processes $X_1(t-1)X_2(t)$ and $X_1(t+4)X_2(t)$ using the regression-based method in Khan, Katzoff and Kedem (2014) and Kedem (2016). We selected four lag terms from each input, i.e. $x_1(t), \ldots, x_1(t-3), x_2(t), \ldots, x_2(t-3), x_1(t-1)x_2(t), \ldots, x_1(t-4)x_2(t-3), x_1(t)x_2(t-4), \ldots, x_1(t-3)x_2(t-7)$ and regressed $y(t)$ on all the selected covariates. We then performed stepwise selection based on AIC and it is observed from Table 1 that the selected
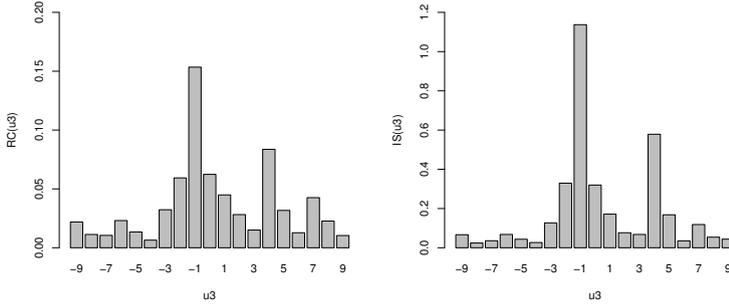
Figure 1. $\widehat{RC}_{3:3}(u_3)$ (left) and $\widehat{IS}_{3:3}(u_3)$ (right) for $u_3 = -9,\ldots,9$.
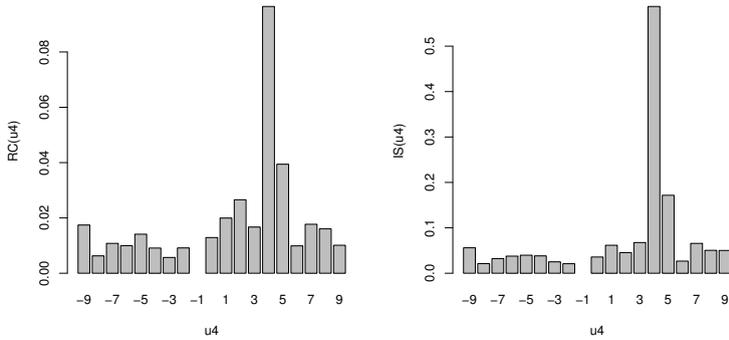
Figure 2. $\widehat{RC}_{4:4}(u_4)$ (left) and $\widehat{IS}_{4:4}(u_4)$ (right) for $u_4 = -9,\ldots,9$. Note that the bars that correspond to $u_4 = -1$ are set to be 0 since $u_3 = -1$.

Table 1. Regression result of the selected model

	Estimate	SE	p-value
Intercept	-0.0264	0.0322	0.4125
$x_1(t)$	0.3876	0.0292	0.0000
$x_2(t)$	0.2907	0.0303	0.0000
$x_2(t-2)$	0.0571	0.0307	0.0629
$x_2(t-3)$	-0.0555	0.0307	0.0708
$x_1(t-2)x_2(t-1)$	0.3729	0.0277	0.0000
$x_1(t)x_2(t-4)$	0.2569	0.0275	0.0000

model is similar to (6) since the significant ($\alpha = 0.05$) covariates are identical to the ones in (6) and the estimated coefficients correspond to the ones in (6), which validates the method.

   Remark: It seems that regression-based selection criteria of interaction terms can be applied directly, thus bypassing the need for our graphical method. However, we rationalize the use of our spectral graphical selection for the following reason. The number of potential covariates in the initial model might be too large, which could result in conflicting selections and possible inconsistencies depending on the model selection method. Our graphical method identifies potentially useful interactions which can then be taken into account and reduce significantly the number of covariates fed into any model selection method, thus rendering the selection more manageable.

4. An Application to Volatility Index

The Volatility Index of a certain underlying asset gives the expectation of the corresponding market volatility in a certain future period. The first and most famous one, VIX, was introduced by Whaley (1993). The underlying asset for VIX is the S&P 500 index so that it reflects the implied volatility of the stock performance of large capitalization companies. For the implied volatility of small capitalization stocks, we have chosen RVX. These two volatility indices shall be considered here as indicators for the stock market. For commodity markets, two important volatility indices, the Crude Oil Exchange Traded Funds Volatility Index (OVX) and the Gold Exchange Traded Funds Volatility Index (GVX) were used. The two-year (2018-2019) daily data of these four series were taken from the Federal Reserve Economic Data Website (https://fred.stlouisfed.org/).

   The above methods were applied to analyze the relationships between the volatility indices of the stock and commodity markets. This section is divided into two parts, one investigates the influence of OVX and GVX on VIX and the other examines the influence on RVX.

   Before the analysis, the four series were pre-processed to render them approximately stationary. That was achieved by first-order differencing of the original series and centring at zero. Figure 3 depicts the four series before and after processing.
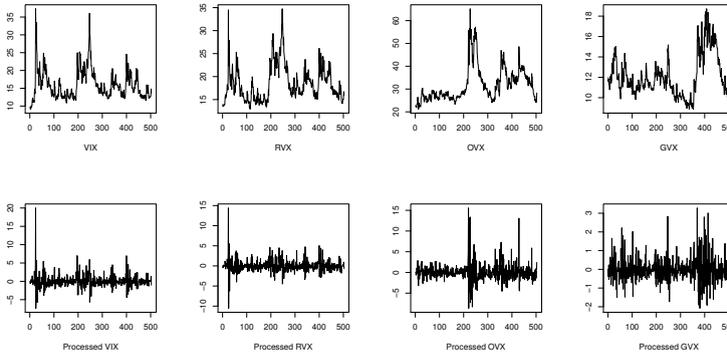
Figure 3. VIX, RVX, OVX and GVX series before and after processing

Table 2. Regression result of $y(t)$ on $x_1(t)$ and $x_2(t)$

	Estimate	SE	p-value
$x_1(t)$	0.1486	0.0347	0.0000
$x_2(t)$	0.9395	0.1075	0.0000

4.1. VIX, OVX and GVX

Consider the processed VIX series as the output and the processed OVX and GVX series as the input. Denote the processed VIX as $y(t)$, processed OVX as $x_1(t)$, and processed GVX as $x_2(t)$. The results of linear regression of $Y(t)$ on $x_1(t)$ and $x_2(t)$ in Table 2 indicate that it is reasonable to include these two series as input since their coefficients are significant. Note that the intercept is omitted since all three series were centred at zero.

Then, the goal is to find the third input based on the cross products of $x_1(t)$ and $x_2(t)$. This resembles the simulation problem so that we perform the same analysis as we did in Section 3. We first select the third input from the family of processes $\{X_1(t+h)X_2(t) : h = -9, -8, \ldots, 0, \ldots, 9\}$ and the estimated *RC*'s and *IS*'s are shown in Figure 4 and it is observed that both criteria indicate that $u_3 = 4$ is the optimal choice. The u_4 is checked with $u_3 = 4$ fixed and Figure 5 shows that none of the bars is particularly prominent so that we stop at the third input. In correspondence to Section 3, we selected four lag terms from each of the three input series and performed a stepwise selection. The final model selected by AIC is shown in Table 3. It includes the two significant (α=0.05) interaction terms $x_1(t)x_2(t-4)$ and $x_1(t-1)x_2(t-5)$.

4.2. RVX, OVX and GVX

We repeated the analysis in Section 4.1 with VIX replaced by RVX. We still consider the processed OVX and GVX as input and try to detect possible significant interactions. In
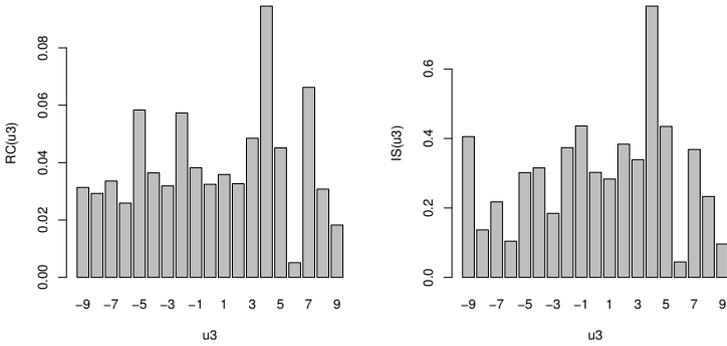
Figure 4. $\widehat{RC}_{3:3}(u_3)$ (left) and $\widehat{IS}_{3:3}(u_3)$ (right) for $u_3 = -9,\ldots,9$.
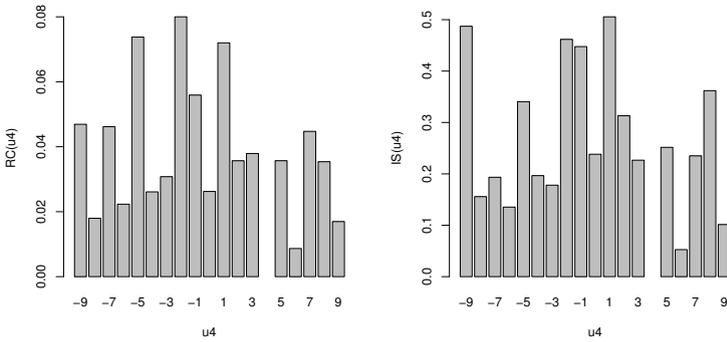
Figure 5. $\widehat{RC}_{4:4}(u_4)$ (left) and $\widehat{IS}_{4:4}(u_4)$ (right) for $u_4 = -9,\ldots,9$. Note that the bars that correspond to $u_4 = 4$ are set to be 0 since $u_3 = 4$

Table 3. Regression result of the selected model

|	Estimate	SE	p-value
Intercept	-0.0007	0.0717	0.9927
$x_1(t)$	0.1677	0.0360	0.0000
$x_2(t)$	0.9039	0.1133	0.0000
$x_2(t-1)$	-0.2504	0.1078	0.0206
$x_1(t)x_2(t-4)$	0.1344	0.0498	0.0072
$x_1(t-1)x_2(t-5)$	-0.1010	0.0508	0.0473
$x_1(t-2)x_2(t-6)$	-0.0914	0.0504	0.0707
$x_1(t-3)x_2(t-7)$	0.0817	0.0501	0.1037
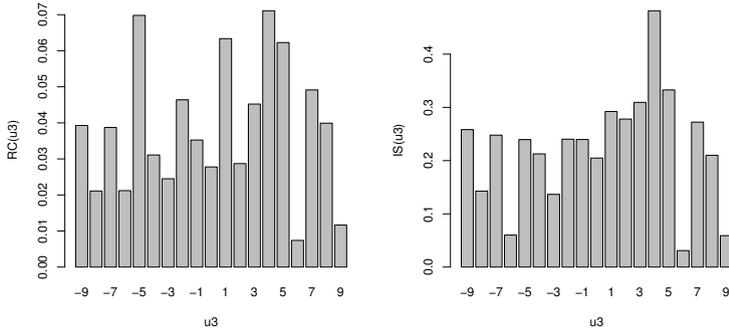
Figure 6. $\widehat{RC}_{3:3}(u_3)$ (left) and $\widehat{IS}_{3:3}(u_3)$ (right) for $u_3 = -9, \ldots, 9$.
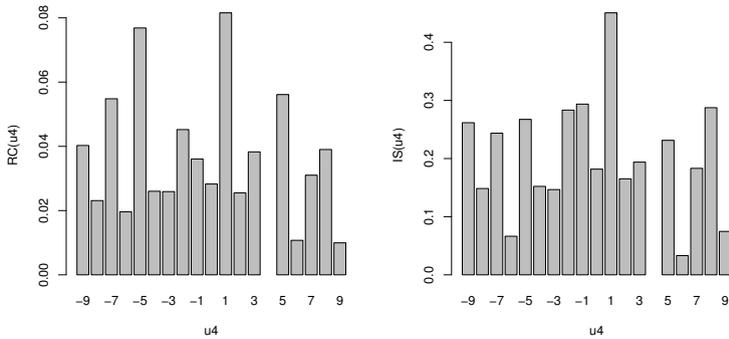

Figure 7. $\widehat{RC}_{4:4}(u_4)$ (left) and $\widehat{IS}_{4:4}(u_4)$ (right) for $u_4 = -9, \ldots, 9$. Note that the bars that correspond to $u_4 = 4$ are set to be 0 since $u_3 = 4$

Table 4. Regression result of the selected model

	Estimate	SE	p-value
Intercept	0.0035	0.0626	0.9549
$x_1(t)$	0.1253	0.0314	0.0001
$x_2(t)$	0.8296	0.0994	0.0000
$x_2(t-1)$	-0.2130	0.0954	0.0261
$x_1(t)x_2(t-4)$	0.1191	0.0440	0.0071
$x_1(t-1)x_2(t-5)$	-0.1114	0.0448	0.0132
$x_1(t-3)x_2(t-7)$	0.0632	0.0432	0.1438
$x_1(t)x_2(t-1)$	0.1049	0.0512	0.0410
$x_1(t-1)x_2(t-2)$	-0.1313	0.0497	0.0085

Figure 6, both bar plots indicate that the optimal choice for u_3 is 4 while the bar plot of $\widehat{RC}_{3:3}(u_3)$ indicates that we might need to consider 1 and -5 as well. Therefore, we checked for u_4 and Figure 7 shows that no bar stands out in the graph of $\widehat{RC}_{4:4}(u_4)$ while the bar of $u_4 = 1$ is prominent in the graph $\widehat{IS}_{4:4}(u_4)$. Therefore, we took $X_1(t+1)X_2(t)$ as the fourth input.

We selected the lag terms as in Section 3 and 4.1 and the result of stepwise regression based on AIC is shown in Table 4. Four significant ($\alpha = 0.05$) interaction terms, $x_1(t)x_2(t-4)$, $x_1(t-1)x_2(t-5)$, $x_1(t)x_2(t-1)$ and $x_1(t-1)x_2(t-2)$, are detected where the first two are from $X_1(t+4)X_2(t)$ and the last two are from $X_1(t+1)X_2(t)$.

5. Conclusion

Residual coherence and integrated spectrum proposed in this paper are graphical devices which point to possible significant interactions based on the result of Sections 3 and 4. Significant interactions could produce one or more than one prominent bars in the bar plots of $RC_{k:k}(u_k)$ and $IS_{k:k}(u_k)$ as functions of the kth input interaction.

When there are multiple prominent bars, one could consider u_{k+1} for more possible significant interactions. Once the input processes are determined, one can employ the regression-based selection method proposed in Khan, Katzoff and Kedem (2014) and Kedem (2016) to search for significant covariate interactions.

In addition, it is observed from the analysis in Section 4 that the cross product interaction $X_1(t+4)X_2(t)$ of the first order differences of OVX and GVX has significant influence on the first order differences of VIX and RVX. This suggests that daily increments of implied volatility of the stock market are possibly influenced by products of the daily increments (and their lags) of implied volatility of commodity markets. The process $X_1(t+4)X_2(t)$ might be an essential factor in the relationship between the implied volatility of stock market and certain commodity markets and therefore further exploration is warranted.

Acknowledgements

References

Brillinger, D. R., (1965). An introduction to polyspectra. *The Annals of mathematical statistics*, pp. 1351–1374.

Brillinger, D. R., Rosenblatt, M., (1967). Asymptotic theory of estimates of kth-order spectra. *Proceedings of the National Academy of Sciences of the United States of America*, 57(2), p. 206.

Elgar, S., Vanhoff, B., Aguirrre, L. A., Freitas, U. S., Chandran, V., (1998). Higher-order spectra of nonlinear polynomial models for chua's circuit. *International Journal of Bifurcation and Chaos*, 8(12), pp. 2425–2431.

Euan, C., Sun, Y., Ombao, H., (2019). Coherence-based time series clustering for statistical inference and visualization of brain connectivity. *Annals of Applied Statistics*, 13(2), pp. 990–1015.

Hinich, M. J., (1979). Estimating the lag structure of a nonlinear time series model. *Journal of the American Statistical Association*, 74 (366a), pp. 449–452.

Kedem, B., (2016). Coherence consideration in binary time series analysis. *Handbook of Discrete-Valued Time Series*, p. 311.

Kedem, B., Fokianos, K., (2002). *Regression models for time series analysis*. Wiley, Hoboken.

Kedem-Kimelfeld, B., (1975). Estimating the lags of lag processes. *Journal of the American Statistical Association*, 70(351a), pp. 603–605.

Khan, D., Katzoff, M., Kedem, B., (2014). Coherence structure and its application in mortality forecasting. *Journal of Statistical Theory and Practice*, 8(4), pp. 578–590.

Kimelfeld, B., (1974). Estimating the kernels of nonlinear orthogonal polynomial functionals. *The Annals of Statistics*, pp. 353–358.

Koopmans, L. H., (1974). *The spectral analysis of time series*. New Yorkn NY: Academic Press.

Maharaj, E. A., D'Urso, P., (2010). A coherence-based approach for the pattern recognition of time series. *Physica A: Statistical mechanics and its Applications*, 389(17), pp. 3516–3537.

Nikias, C. L., Mendel, J. M., (1993). Signal processing with higher-order spectra. *IEEE Signal processing magazine*, 10(3), pp. 10–37.

Sanaullah, M., (2013). A review of higher order statistics and spectra in communication systems. *Global Journal of Science Frontier Research, Physics and Space Science*, 13(4).

Sun, F. T., Miller, L. M., D'Esposito, M., (2004). Measuring interregional functional connectivity using coherence and partial coherence analyses of fmri data. *Neuroimage*, 21(2), pp. 647–658.

Tick, L. J., (1961). The estimation of "transfer functions" of quadratic systems. *Technometrics*, 3(4), pp. 563–567.

Whaley, R. E., (1993). Derivatives on market volatility: Hedging tools long overdue. *The journal of Derivatives*, 1(1), pp. 71–84.

Wiener, N., (1958). *Nonlinear problems in random theory*. M.I.T. Press, Cambridge.

APPENDIX

Since all G_k's are mutually orthogonal, fix k, then $\forall h$,

$$EG_k(t+h;u_1,\ldots,u_k)\overline{G_j(t;u_1,\ldots,u_j)} = 0 \qquad j = 1,\ldots,k-1$$

$$\Rightarrow \int_{-\pi}^{\pi} e^{ih\lambda} \sum_{l=1}^{k} A_{l,k+l-1}(\lambda) f_{X_{l,u_l}X_{j,u_j}}(\lambda)d\lambda = 0 \qquad j = 1,\ldots,k-1$$

$$\Rightarrow \sum_{l=1}^{k} A_{l,k+l-1}(\lambda) f_{X_{l,u_l}X_{j,u_j}}(\lambda)d\lambda = 0 \qquad j = 1,\ldots,k-1$$

$$\Rightarrow \boldsymbol{F}_k(\lambda)\boldsymbol{A}_k(\lambda) = \boldsymbol{f}_k(\lambda)A_{k,1}(\lambda)$$

where $\boldsymbol{A}_k(\lambda) \equiv [A_{1,k}(\lambda), A_{2,k-1}, \ldots, A_{k-1,2}]^T$. Then, by Cramer's rule,

$$A_{j,k-j+1}(\lambda) = \frac{\boldsymbol{F}_{k,j}(\lambda)}{\boldsymbol{F}_k(\lambda)} A_{k,1}(\lambda) = c_{k,j}(\lambda)A_{k,1}(\lambda)$$

for $j = 1,\ldots,k-1$. Based on the orthogonality and the uniqueness of Fourier transform, we also have

$$EG_k(t+h;u_1,\ldots,u_k)\overline{Y(t)} = EG_k(t+h;u_1,\ldots,u_k)\overline{G_k(t;u_1,\ldots,u_k)}$$

$$\sum_{j=1}^{k} A_{k,k-j+1}(\lambda) f_{X_{j,u_j}Y}(\lambda) = \overline{A_{k,1}(\lambda)} \sum_{j=1}^{k} A_{k,k-j+1}(\lambda) f_{X_{j,u_j}X_{k,u_k}}(\lambda)$$

$$A_{k,1}(\lambda) \sum_{j=1}^{k} c_{k,j}(\lambda) f_{X_{j,u_j}Y}(\lambda) = |A_{k,1}(\lambda)|^2 \sum_{j=1}^{k} c_{k,j}(\lambda) f_{X_{j,u_j}X_{k,u_k}}(\lambda)$$

$$A_{k,1}(\lambda) = \overline{\left[\frac{\sum_{j=1}^{k} c_{k,j}(\lambda) f_{X_{j,u_j}Y}(\lambda)}{\sum_{j=1}^{k} c_{k,j}(\lambda) f_{X_{j,u_j}X_{k,u_k}}(\lambda)} \right]}$$

Therefore, all A's for G_k are solved.

Estimating completion and breakoff functions
in a business web survey

Berislav Žmuk[1]

ABSTRACT

A business web survey should be of an appropriate length. On the one hand it should include all the questions which are important to the researcher, but on the other hand, it should not be too long as the breakoff rate in this case tends to be high, resulting in a low response rate. In consequence, the researcher is forced to invest more time and money in order to reach a sample size which would enable an appropriately performed statistical analysis. In this paper, completion and breakoff times are observed and compared across different questionnaire and respondent characteristics. A regression modelling approach has been adopted to estimate the completion and breakoff functions to help a researcher determine which respondents completed a questionnaire and which broke it off too quickly or too slowly. By omitting such respondents, a researcher is able to obtain the relevant estimates more efficiently. In addition, the completion and breakoff functions offer a better insight into the completion and breakoff development rates, allowing the researcher to make a better-informed decision as to whether the survey requires any modifications or not.

Key words: breakoff function, business web survey, completion function, questionnaire.
JEL: C12, C20, C83.

1. Introduction

The most common way of conducting a research in the modern business world is through a survey. However, conducting a survey is not so each task as it may look at the first sight. Namely, there are always more questions which researchers want to include into the questionnaire. On the other hand, business or enterprises do not have unlimited time and resources to be spent on taking part in a survey. In addition, if the survey questionnaire is too long, respondents in enterprises could simply break off at a certain point. In this way, the research would not get all answers from such respondents. It has to be emphasized that the survey questionnaire could be considered

---

[1] University of Zagreb, Faculty of Economics and Business, Department of Statistics, Trg J. F. Kennedyja 6, HR-10000 Zagreb, Croatia. E-mail: bzmuk@efzg.hr. ORCID: https://orcid.org/0000-0003-3487-1376.

too long not only because it has too many questions, but because it has too difficult questions for the respondents as well.

Consequently, the question is how to make a survey of an optimal length. As indicators of an optimal survey length, a researcher could observe completion and breakoff times. If the completion time is too long, a research should reduce the number of questions and lower the level of their complexity. If the breakoff time is too short, the same should be done. However, when the completion time is observed it is assumed that respondents answer all given questions, whereas when the breakoff time is observed the time when they stopped providing answers is observed. In this way, by observing breakoff times, some impact on overall response rates could be given as well.

There are three main outcomes related to the survey response of respondents. So, a respondent can make a decision not to participate in the survey. In that case, survey nonresponse appears (Kish, 1995, Groves, 2006). While survey nonresponse is something unwanted for a researcher, the most wanted situation for a researcher is when a respondent completes the survey by answering all survey questions. The third possible outcome is that a respondent starts with the survey but gives up at some point in the survey. In that case, one can speak about survey breakoff. There are some other types of breakoff like unit and item nonresponse (Peytchev, 2009) but due to limitation of the paper length it will not be discussed here. Also, because survey nonresponse is not the focus of this paper, it will not be observed in more details.

The goal of the paper is to introduce a way how completion and breakoff times could be used in the process of survey preparation. Consequently, in the paper completion and breakoff functions are going to be estimated. Some authors already tried to estimate impact of different items on completion and breakoffs in web surveys. However, they focused only on point estimation and comparison between different surveys. Crawford, Couper and Lamias (2001) estimated the likelihood of breakoff by sharing different information to respondents about the survey length. However, that information was not true for all respondents. Galesic (2006) used three surveys of different length in her analysis. It came out that the 30-minute survey had 40% higher breakoff risk than at a 10-minute survey. By reducing the length of the survey to 20 minutes, the breakoff risk was 20% higher than at a 10-minute survey. Göritz (2006) has shown that surveys with incentives have in average 27% higher probability of being complete. Peytchev (2011) compared breakoffs for respondents in consecutive surveys. He used multinomial regression approach to estimate odds ratios for different previous survey outcomes and for different characteristics of respondents (gender, race, years in school). Similarly, Blumenberg et al. (2018) also compared breakoff rates by comparing different characteristics of respondents by applying logistic regression. Vehovar and Cehovin (2014) conducted meta-analysis and found that 80% of all breakoffs occur on the first introductory pages. Mittereder (2019) used survival analysis for predicting

when respondents will breakoff from a web survey. The analysis is based on comparing survey results from two different years.

In order to be able to estimate the completion and breakoff functions, the paradata from a business web survey conducted on a sample of Croatian enterprises is used. Still, in addition to the research question whether the completion and breakoff functions potentially could be used for finding optimal survey time, another research question is whether they could be used in the process of detecting respondents who completed the questionnaire too fast (speeders) or too slowly. Accordingly, two research hypotheses have been defined. The first research hypothesis is that optimal survey lengths, estimated by the completion functions, are different for different questionnaire designs and for respondents of different characteristics. The second research hypothesis contains the assumption that the completion and breakoff functions can be used to detect too fast and too slow respondents.

The paper is organized as follows. After a brief introduction and an overview of previous research, data used in the analysis are presented. In addition, in the second chapter methods applied in the analysis are introduced as well. In the third chapter the completion times and the completion function are analysed, whereas in the fourth chapter focus is given to breakoff times and breakoff function estimation. Discussion is provided in the fifth chapter, whereas the sixth chapter concludes.

2. Data and methodology

Data for the analysis are taken from the business web survey which was conducted in 2016. The target population were active enterprises which were registered at the Commercial Court in Croatia until July 1, 2016 and which have a public available e-mail address. The responses were collected in the period from October 4 to December 31, December 2016. In that period, in addition to the initial invitation to participate in the survey research, a total of two reminders were sent.

The topic of the survey was statistics methods use in enterprises. The questionnaire mostly consisted of about 20 close-ended questions. However, some of the close-ended questions were binary questions, some questions were Likert scale based, and some were single select multiple choice questions. Also, a couple of open-ended questions were added mostly as a support to respondents if they wanted to emphasize something. According to that, this web survey could be observed as a medium long and complex questionnaire (1ka, 2020a,b, Žmuk, 2017).

For the purpose of the analysis in this paper, instead of responses, completion and breakoff times will be the focus. The completion time is defined here as the time which a respondent needed to fully complete the questionnaire. The time is started to be measured when a respondent opened the questionnaire and it is stopped when the

respondent selects to submit his answers. On the other hand, the breakoff time is measured from the point when a respondent opened the questionnaire to the point when the last action of the respondent in the questionnaire in the used survey software is registered. In order to collect data LimeSurvey software is used.

Initially different questionnaire versions were prepared. Different questionnaire versions were randomly associated to the population units. However, it has to be emphasized that different questionnaire versions were allocated to approximately the same number of the population units. Still, due to different response rates at different questionnaire versions, a different number of responses were collected in each questionnaire version.

The questionnaire versions were different according to the fact whether the pictures were included in the questionnaire or not and what kind of pictures were there. The first questionnaire version did not include any pictures, whereas other two questionnaire versions included pictures. In one, the so-called "positive", and in the second one "negative" pictures are implemented. Positive pictures included some positive information like picture with positive business trend or a table full of advanced statistical books. On the other hand, the questionnaire version with negative pictures included pictures where something "negative" is shown like a negative business trend or a table with small number of basic statistical books.

The questionnaire versions were different according to the number of questions presented to a respondent per questionnaire screen. In the first questionnaire version all questions were immediately shown to respondents. In the second questionnaire version questions were grouped into logical sections. The third questionnaire version presents only one question per questionnaire screen at a time.

The completion and breakoff times are observed according to the main characteristic of the participating enterprises as well. In this way, enterprises according to their legal form, size and main activity are inspected. According to the Enterprises Act, a distinction between joint stock enterprises, limited liability enterprises and simple limited liability enterprises is made (Narodne novine, 2011). Furthermore, according to the Accounting Act the enterprises are stratified by their size on small, medium and large enterprises (Narodne novine, 2015). Also, enterprises are observed according to their main activities. According to National Classification of Economic Activities (Narodne novine, 2007) overall 21 areas of enterprises activities are recognized. However, for the purpose of the analysis those areas are merged and, in this way, the number of main activities groups is reduced to four (industrial enterprises, trade enterprises, service enterprises, other enterprises).

The completion and breakoff times will be analysed separately but the same approaches and methods will be used in both cases. First of all, data cleaning will be performed. Namely, it is possible that some respondents needed too much time to

complete the questionnaire. For example, they started the survey but were distracted by a phone call, e-mails, other colleagues and similar. In this way, the completion time is longer that in reality was. Also, it is possible that the survey software wrongly measured times. For example, that the survey is completed in zero seconds. Such cases should also be omitted from the further analysis.

After the data is cleaned, descriptive statistics methods will be used to observe the completion and breakoff times. Because there are a lot of unique completion and breakoff times (measured in seconds), the descriptive statistics analysis will be applied on grouped data. In order to be able to compare results the same time groups will be used for completion and for breakoff times analysis.

The structure of respondents at different questionnaire versions will be compared by using the chi-square test for equality of three or more population proportion results. The null hypothesis of the chi-square test is that proportions are statistically equal across all observed populations. On the other hand, in the alternative hypothesis the assumption that not all proportions across the different populations are equal is incorporated. The empirical chi-square value is calculated as follows:

$$\chi^2 = \sum_{j=1}^{k} \frac{(m_j - e_j)^2}{e_j} + \sum_{j=1}^{k} \frac{(n_j - m_j - e_j^C)^2}{e_j^C} \tag{1}$$

where χ^2 is the empirical chi-square value, m_j is the number of units in the j-th population with certain characteristics, e_j is the expected number of units in the j-th population with certain characteristics, n_j is the size of the j-th population, e_j^C is the expected number of units in the j-th population without certain characteristics. The expected number of units in the j-th population with certain characteristics is calculated using the following equation:

$$e_j = n_j \cdot \frac{\sum_{j=1}^{k} m_j}{\sum_{j=1}^{k} n_j} \tag{2}$$

where e_j is the expected number of units in the j-th population with certain characteristics, n_j is the size of the j-th population, m_j is the number of units in the j-th population with certain characteristics.

In the final stage of analysis, simple linear regression analysis will be applied. The estimated simple linear regression model by ordinary least squares method has the following form:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x \tag{3}$$

where $\hat{y}$ is estimated value of the dependent variable, x is independent variable, $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated parameters. In the analysis dependent variable is going to be cumulative proportion of respondents. Cumulative proportion is defined as the sum of respondents with the survey time lower than a certain limit divided by the total number

of respondents. On the other hand, independent variable is the time in which a respondent finished the survey. However, a respondent could finish the survey so that he completes it or break off at the certain point in the survey. Therefore, one regression analysis is conducted where the completion time is independent variable, whereas in the second regression analysis independent variable is the breakoff time. If it were be necessary, the observed variables would be transformed.

3. Analysis of complete responses

3.1. Descriptive statistics of completion times

The observed business web survey covered 1,433 enterprises overall. From that number the survey system registered 780 completed surveys. However, the completion times varied too much with some really strange results. Therefore, further analysis excluded respondents who needed zero (0) seconds to complete the web survey (137 cases) and respondents who needed more than 1,800 seconds (24 cases). In this way, the final number of observed respondents is 619. The distribution of respondents according to their completion time is shown in Table 1.

Table 1. Distribution of respondents according to completion time, all respondents

Time, in min.	Respondents	Percentage	Cumulative percentage
0-5	387	63%	63%
5-10	173	28%	90%
10-15	33	5%	96%
15-20	16	3%	98%
20-30	10	2%	100%
Total	619	100%	-----

Source: Author.

According to the results from Table 1, more than half of respondents needed less than five minutes to complete the survey. Furthermore, 90% of respondents completed it in less than 10 minutes.

Table 2 Distribution of respondents according to completion time, different questionnaire designs regarding presented pictures in the questionnaire

Time, in min.	Positive pictures			Negative pictures			Without pictures		
	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.
0-5	125	60%	60%	126	67%	67%	136	62%	62%
5-10	64	30%	90%	47	25%	92%	62	28%	90%
10-15	12	6%	96%	9	5%	96%	12	5%	95%
15-20	7	3%	99%	3	2%	98%	6	3%	98%
20-30	2	1%	100%	4	2%	100%	4	2%	100%
Total	210	100%	-----	189	100%	-----	220	100%	-----

Source: Author.

While Table 1 observes respondents no matter which questionnaire version they are given, in Table 2 respondents' completion times are observed according to the presented pictures in the questionnaire. In this way, respondents who completed questionnaire with positive pictures, with negative pictures or without pictures are observed separately. However, the results are quite similar to the results at the overall level. Therefore, in all three questionnaire versions about 90% of respondents completed their questionnaire version in less than 10 minutes.

Table 3. Distribution of respondents according to completion time, different questionnaire designs regarding the number of questions presented to respondent per questionnaire screen

Time, in min.	One question per screen			Group of questions			All questions		
	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.
0-5	151	55%	55%	111	53%	53%	125	95%	95%
5-10	94	34%	88%	73	35%	88%	6	5%	99%
10-15	16	6%	94%	16	8%	95%	1	1%	100%
15-20	8	3%	97%	8	4%	99%	0	0%	100%
20-30	8	3%	100%	2	1%	100%	0	0%	100%
Total	277	100%	-----	210	100%	-----	132	100%	-----

Source: Author.

In Table 3 the distribution of respondents according to the completion time is given again but now the completion times are observed according to the different number of questions presented to a respondent per questionnaire screen. If questionnaire versions with one question per screen and with a group of questions are observed, the completion times are in line with completion times at the overall level. However, in the case when all questions are immediately presented to respondents, almost all respondents (99%) needed less than 10 minutes to complete the questionnaire.

Table 4. Distribution of respondents according to completion time, different questionnaire designs regarding the legal form of enterprises

Time, in min.	Joint stock enterprises			Limited liability enterprises			Simple limited liability enterprises		
	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.
0-5	5	42%	42%	359	62%	62%	23	72%	72%
5-10	5	42%	83%	160	28%	90%	8	25%	97%
10-15	2	17%	100%	30	5%	95%	1	3%	100%
15-20	0	0%	100%	16	3%	98%	0	0%	100%
20-30	0	0%	100%	10	2%	100%	0	0%	100%
Total	12	100%	-----	575	100%	-----	32	100%	-----

Source: Author.

Table 4 presents the distribution of respondents according to the completion time when different legal forms of enterprises are taken into account. It has to be emphasized that vast majority of respondents (575 respondents or 93%) were limited liability enterprises and because of that their results are almost the same as at the overall level. In the case of joint stock and simple limited liability enterprises, all respondents managed to complete the questionnaire under 15 minutes time.

Table 5. Distribution of respondents according to completion time, different questionnaire designs regarding the size of enterprises

Time, in min.	Small enterprises			Medium enterprises			Large enterprises		
	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.
0-5	373	62%	62%	9	60%	60%	5	71%	71%
5-10	166	28%	90%	5	33%	93%	2	29%	100%
10-15	32	5%	96%	1	7%	100%	0	0%	100%
15-20	16	3%	98%	0	0%	100%	0	0%	100%
20-30	10	2%	100%	0	0%	100%	0	0%	100%
Total	597	100%	-----	15	100%	-----	7	100%	-----

Source: Author.

In Table 5 the size of enterprises is taken into account. Similarly to previous categories, there is also one dominating category here. So, small enterprises have share of 96% in the total number of observed enterprises. All respondents from medium enterprises managed to complete the questionnaire in less than 15 minutes, whereas all respondents from large enterprises managed to complete the questionnaire in less than 10 minutes.

Table 6. Distribution of respondents according to completion time, different questionnaire designs regarding the main activity of enterprises

Time, in min.	Industrial enterprises			Trade enterprises			Service enterprises			Other enterprises		
	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.	Res.	Per.	Cum. per.
0-5	109	62%	62%	72	57%	57%	188	65%	65%	18	69%	69%
5-10	48	27%	89%	45	35%	92%	76	26%	91%	4	15%	85%
10-15	11	6%	95%	4	3%	95%	15	5%	96%	3	12%	96%
15-20	5	3%	98%	4	3%	98%	7	2%	99%	0	0%	96%
20-30	3	2%	100%	2	2%	100%	4	1%	100%	1	4%	100%
Total	176	100%	-----	127	100%	-----	290	100%	-----	26	100%	-----

Source: Author.

Table 6 observes respondents' completion time according to the main activity of the enterprises.

It turned out that 92% of respondents from trade enterprises completed the questionnaire in less than 10 minutes. Such respondents from service enterprises were 91%, from industrial enterprises 89% and from other enterprises 85%.

3.2. Comparison of respondents' proportions according to their completion time

In the previous chapter completion time distributions of respondents according to different characteristics have been just reported. In this chapter the differences between completion time distributions of respondents according to different characteristics will be inspected. In order to do that the chi-square test for equality of three or more population proportions is applied. It has been inspected whether the proportions of respondents are equal at the same completion time level across the observed characteristics or not.

Table 7. Chi-square test for equality of three or more population proportion results, completion time observed, different questionnaire designs regarding presented pictures in the questionnaire

Time, in min.	Comm. prop.	Exp. res. – positive pictures	Exp. res. – negative pictures	Exp. res. – no pictures	Emp. chi-square	p-value
0-5	0.6252	131.29	118.16	137.54	2.238	0.3267
5-10	0.2795	58.69	52.82	61.49	1.563	0.4577
10-15	0.0533	11.20	10.08	11.73	0.189	0.9098
15-20	0.0258	5.43	4.89	5.69	1.232	0.5401
20-30	0.0162	3.39	3.05	3.55	0.936	0.6262
Total	-----	210	189	220	-----	-----

Source: Author.

In Table 7 the chi-square test for equality of three or more population proportion results between respondents with different presented pictures is shown. The results suggest that at the significance level of 5% the null hypothesis cannot be rejected. This conclusion is valid for all five observed completion time categories. In other words, there is no statistically significant difference in the proportion of respondents in each of the five observed time categories between questionnaires with positive pictures, with negative pictures and without them.

Table 8. Chi-square test for equality of three or more population proportion results, completion time observed, different questionnaire designs regarding the number of questions presented to respondent per questionnaire screen

Time, in min.	Comm. prop.	Exp. res. – one question per screen	Exp. res. – group of questions	Exp. res. – all questions	Emp. chi-square	p-value
0-5	0.6252	173.18	131.29	82.53	74.271	<0.0001
5-10	0.2795	77.42	58.69	36.89	45.673	<0.0001
10-15	0.0533	14.77	11.20	7.04	7.758	0.0207
15-20	0.0258	7.16	5.43	3.41	4.855	0.0883
20-30	0.0162	4.47	3.39	2.13	5.571	0.0617
Total	-----	277	210	132	-----	-----

Source: Author.

According to the results given in Table 8, at the significance level of 5%, there is a statistically significant difference in respondents' proportions for different questionnaire designs regarding the number of questions presented to a respondent per questionnaire screen at completion time categories 0–5 minutes, 5–10 minutes and 10–15 minutes. However, it seems that there is no statistically significant difference in respondent's proportions in the last two completion time categories.

Table 9. Chi-square test for equality of three or more population proportion results, completion time observed, different questionnaire designs regarding the legal form of enterprises

Time, in min.	Comm. prop.	Exp. res. – joint stock enterprises	Exp. res. – limited liability enterprises	Exp. res. – simple limited liability enterprises	Emp. chi-square	p-value
0-5	0.6252	7.50	359.49	20.01	3.424	0.1805
5-10	0.2795	3.35	160.70	8.94	1.264	0.5316
10-15	0.0533	0.64	30.65	1.71	3.378	0.1847
15-20	0.0258	0.31	14.86	0.83	1.257	0.5334
20-30	0.0162	0.19	9.29	0.52	0.778	0.6778
Total	-----	12	575	32	-----	-----

Source: Author.

Results from Table 9 suggest that there is no statistically significant difference, at the significance level of 5%, in respondent populations according to all observed completion time categories between joint stock enterprises, limited liability enterprises and simple limited liability enterprises.

Table 10. Chi-square test for equality of three or more population proportion results, completion time observed, different questionnaire designs regarding the size of enterprises

Time, in min.	Comm. prop.	Exp. res. – small enterprises	Exp. res. – medium enterprises	Exp. res. – large enterprises	Emp. chi-square	p-value
0-5	0.6252	373.25	9.38	4.38	0.278	0.8702
5-10	0.2795	166.85	4.19	1.96	0.223	0.8943
10-15	0.0533	31.83	0.80	0.37	0.448	0.7992
15-20	0.0258	15.43	0.39	0.18	0.605	0.7389
20-30	0.0162	9.64	0.24	0.11	0.375	0.8292
Total	-----	597	15	7	-----	-----

Source: Author.

According to the chi-square test for equality of three or more population proportion results given in Table 10, it can be concluded that at the significance level of 5% the null hypothesis cannot be rejected for all five observed completion time categories. So, there is no difference in respondent distribution according to completion time categories between small, medium and large enterprises.

Table 11. Chi-square test for equality of three or more population proportion results, completion time observed, different questionnaire designs regarding the main activity of enterprises

Time, in min.	Comm. prop.	Exp. res. – industrial enterprises	Exp. res. – trade enterprises	Exp. res. – service enterprises	Exp. res. – other enterprises	Emp. chi-square	p-value
0-5	0.6252	110.04	79.40	181.31	16.26	3.025	0.3878
5-10	0.2795	49.19	35.49	81.05	7.27	6.048	0.1093
10-15	0.0533	9.38	6.77	15.46	1.39	3.491	0.3219
15-20	0.0258	4.55	3.28	7.50	0.67	0.930	0.8181
20-30	0.0162	2.84	2.05	4.68	0.42	0.926	0.8192
Total	-----	176	127	290	26	-----	-----

Source: Author.

Finally, in Table 11 the respondents' distributions are compared for industrial, trade, service and other enterprises. It can be concluded that, at the significance level of 5%, there is no statistically significant difference in the respondents' distributions between those four groups of enterprises in all five observed completion time categories.

3.3. Regression modelling of the completion function

In order to estimate the completion function a regression modelling approach is used. In order to keep things straightforward as much as possible, a simple linear regression model is applied. In the regression model the cumulative proportion of respondents is observed as dependent variable, whereas time, given in seconds, is observed as independent variable.
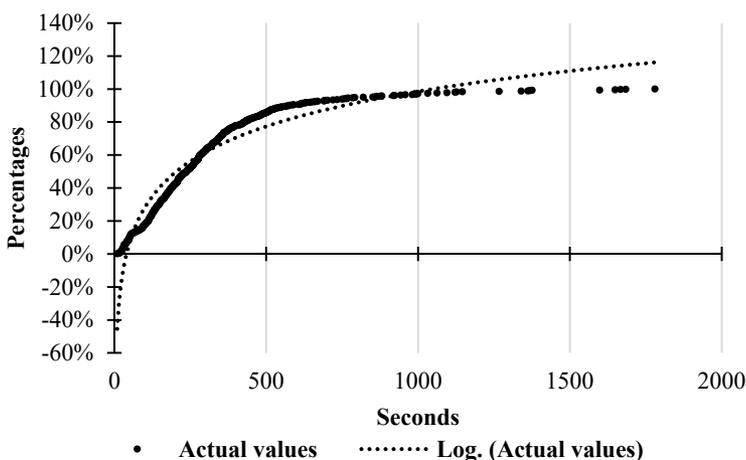
Figure 1. Cumulative distribution of respondents according to completion time and the estimated regression line, all respondents

Source: Author.

According to Figure 1 it is obvious that when all respondents are observed, the relationship between the cumulative proportion of respondents' variable and the time variable is not linear. However, if logarithmic values of the time variable are used, the resulting function is quite similar to the actual distribution of the actual values. In order to be able to compare the results for different characteristics of respondents, in the further analysis only regression models with logarithmic values of the time variable as independent variables are observed.

Still, it has to be emphasized that some limitations appeared due to the data transformation. First of all, by transforming values into logarithmic values the variables completion time and breakoff time are not anymore given in seconds. Therefore, the results are not going to have such intuitive interpretation unless they are transformed back to seconds. Furthermore, the logarithmic transformation could have impact on the data distribution (Feng et al., 2014).

Table 12. Regression analysis results, dependent variable cumulative proportion of respondents, independent variable ln time (in seconds), completion time observed, results for all observed categories of questionnaire designs

Questionnaire designs	No. of res.	No. of unique times	R square	Reg. stand. error	Intercept		Variable ln time	
					Estim.	p-value	Estim.	p-value
Total	619	397	0.9260	0.0807	-1.1253	<0.0001	0.3056	<0.0001
Pictures								
Positive pictures	210	170	0.9081	0.0903	-1.2829	<0.0001	0.3281	<0.0001
Negative pictures	189	162	0.9238	0.0796	-1.1201	<0.0001	0.3062	<0.0001
Without pictures	220	183	0.9350	0.0745	-1.0048	<0.0001	0.2863	<0.0001
Questions								
One question per screen	277	223	0.9327	0.0771	-2.0480	<0.0001	0.4489	<0.0001
Group of questions	210	183	0.9692	0.0508	-2.0379	<0.0001	0.4496	<0.0001
All questions	132	89	0.9139	0.0870	-0.8060	<0.0001	0.3200	<0.0001
Legal form								
Joint stock enterprises	12	11	0.7297	0.1712	-1.0967	0.0094	0.2945	0.0008
Limited liability ent.	575	381	0.9260	0.0808	-1.1110	<0.0001	0.3028	<0.0001
Simple limited liab. ent.	32	31	0.9506	0.0642	-1.4502	<0.0001	0.3757	<0.0001
Size								
Small enterprises	597	389	0.9260	0.0807	-1.1407	<0.0001	0.3078	<0.0001
Medium enterprises	15	15	0.7981	0.1390	-1.0548	0.0004	0.2963	<0.0001
Large enterprises	7	7	0.8855	0.1144	-0.5891	0.0276	0.2623	0.0016
Main activity								
Industrial enterprises	176	154	0.9260	0.0797	-1.0487	<0.0001	0.2910	<0.0001
Trade enterprises	127	116	0.8998	0.0944	-1.1304	<0.0001	0.3032	<0.0001
Service enterprises	290	225	0.9216	0.0831	-1.1628	<0.0001	0.3131	<0.0001
Other enterprises	26	26	0.9614	0.0590	-0.8494	<0.0001	0.2650	<0.0001

Source: Author.

In Table 12 the main regression analysis results for the estimated linear regression models are shown. Each row in Table 12 is related to one linear regression model. According to coefficient determination (R square) values, it can be concluded that all estimated regression models are highly representative. Furthermore, all estimated parameters of the estimated linear regression models are statistically significant at the significance level of 5%.

Table 13. Speeding and completion times estimates based on the regression models, completion time observed, results for all observed categories of questionnaire designs

Questionnaire designs	Speeding time				Completion time			
	Sec.	Min.	No. of faster res.	Per. of faster res.	Sec.	Min.	No. of slower res.	Per. of slower res.
Total	40	0.66	41	6.62%	1,047	17.45	16	2.58%
Pictures								
Positive pictures	50	0.83	13	6.19%	1,052	17.53	5	2.38%
Negative pictures	39	0.65	14	7.41%	1,017	16.94	6	3.17%
Without pictures	33	0.56	17	7.73%	1,099	18.32	4	1.82%
Questions								
One question per screen	96	1.60	4	1.44%	889	14.82	16	5.78%
Group of questions	93	1.55	2	0.95%	861	14.34	13	6.19%
All questions	12	0.21	1	0.76%	283	4.71	7	5.30%
Legal form								
Joint stock enterprises	41	0.69	1	8.33%	1,235	20.58	0	0.00%
Limited liability enterprises	39	0.65	38	6.61%	1,066	17.77	15	2.61%
Simple limited liability enterprises	47	0.79	2	6.25%	680	11.33	1	3.13%
Size								
Small enterprises	41	0.68	40	6.70%	1,047	17.46	16	2.68%
Medium enterprises	35	0.59	1	6.67%	1,027	17.12	0	0.00%
Large enterprises	9	0.16	0	0.00%	427	7.12	0	0.00%
Main activity								
Industrial enterprises	37	0.61	11	6.25%	1,143	19.04	3	1.70%
Trade enterprises	42	0.69	10	7.87%	1,126	18.77	2	1.57%
Service enterprises	41	0.68	18	6.21%	1,001	16.68	9	3.10%
Other enterprises	25	0.41	2	7.69%	1,075	17.91	1	3.85%

Source: Author.

Based on linear regression results given in Table 12, too fast and too slow respondents could be identified. According to estimated linear regression models, too fast respondents are those respondents who completed the survey in less time when the cumulative proportion of respondents in the regression model is equal to 0%. On the other hand, too slow respondents are those respondents who completed the survey in longer time when the cumulative proportion of respondents in the regression model is equal to 100%.

Table 13 presents limit times when a respondent is considered to be too slow or too fast. When all respondents are observed, the limit for speeders is 40 seconds and for too slow respondents 1,047 seconds (about 17 and half minutes). In this way, 41 speeders and 16 too slow respondents are detected who can be omitted from the analysis to get more representative survey results. In Table 13 such results are given for other regression models as well. The results have shown that the most respondents should be omitted at the small enterprises and no respondents at the large enterprises.

4. Analysis of breakoffs

4.1. Descriptive statistics of breakoff times

While in Chapter 4 completion times were observed, in this chapter the breakoff times are the focus. The analysis will be conducted oin analogous way as before. In this way, the direct comparison of the results between those two analyses is possible. Overall 219 respondents started the survey but at some point they broke off.

Table 14. Distribution of respondents according to breakoff time, all respondents

Time, in min.	Breakoffs	Percentage	Cumulative percentage
0-5	203	93%	93%
5-10	11	5%	98%
10-15	3	1%	99%
15-20	0	0%	99%
20-30	2	1%	100%
Total	219	100%	-----

Source: Author.

In Table 14 distribution of all respondents according to the breakoff time is given. As it was expected, the respondents broke off much quicker than the respondents could complete the questionnaire. About 93% of respondents broke off in less than 5 minutes, whereas 98% of respondents broke off in less than 10 minutes.

Table 15. Distribution of respondents according to breakoff time, different questionnaire designs regarding presented pictures in the questionnaire

Time, in min.	Positive pictures			Negative pictures			Without pictures		
	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.
0-5	71	95%	95%	68	92%	92%	64	91%	91%
5-10	2	3%	97%	4	5%	97%	5	7%	99%
10-15	1	1%	99%	1	1%	99%	1	1%	100%
15-20	0	0%	99%	0	0%	99%	0	0%	100%
20-30	1	1%	100%	1	1%	100%	0	0%	100%
Total	75	100%	-----	74	100%	-----	70	100%	-----

Source: Author.

In Table 15 distributions of respondents according to their breakoff time for different questionnaire designs regarding presented pictures in the questionnaire are given. The results are quite similar to the results at the overall level.

Table 16. Distribution of respondents according to breakoff time, different questionnaire designs regarding the number of questions presented to respondent per questionnaire screen

Time, in min.	One question per screen			Group of questions			All questions		
	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.
0-5	144	94%	94%	57	90%	90%	2	100%	100%
5-10	8	5%	99%	3	5%	95%	0	0%	100%
10-15	1	1%	99%	2	3%	98%	0	0%	100%
15-20	0	0%	99%	0	0%	98%	0	0%	100%
20-30	1	1%	100%	1	2%	100%	0	0%	100%
Total	154	100%	-----	63	100%	-----	2	100%	-----

Source: Author.

According to the results in Table 16, about 99% of respondents who got questionnaire version with one question per questionnaire screen broke off in less than 10 minutes, whereas 95% of respondents who got questionnaire version with a group of questions per questionnaire screen broke off in the same period. Unfortunately, there were only two breakoffs in the case of the questionnaire when all questions were presented at once.

Table 17. Distribution of respondents according to breakoff time, different questionnaire designs regarding the legal form of enterprises

Time, in min.	Joint stock enterprises			Limited liability enterprises			Simple limited liability enterprises		
	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.
0-5	5	83%	83%	188	93%	93%	10	100%	100%
5-10	1	17%	100%	10	5%	98%	0	0%	100%
10-15	0	0%	100%	3	1%	99%	0	0%	100%
15-20	0	0%	100%	0	0%	99%	0	0%	100%
20-30	0	0%	100%	2	1%	100%	0	0%	100%
Total	6	100%	-----	203	100%	-----	10	100%	-----

Source: Author.

Table 17 presents the distribution of respondents according to the breakoff time when different legal forms of enterprises are taken into account. The vast majority of respondents (203 respondents or 93%) were limited liability enterprises. Consequently, their results are the same to the results at the overall level. In the case of joint stock enterprises all respondents have the breakoff time lower than 10 minutes, whereas respondents from simple limited liability enterprises broke off in less than 5 minutes.

Table 18. Distribution of respondents according to breakoff time, different questionnaire designs regarding the size of enterprises

Time, in min.	Small enterprises			Medium enterprises			Large enterprises		
	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.
0-5	188	92%	92%	11	100%	100%	4	100%	100%
5-10	11	5%	98%	0	0%	100%	0	0%	100%
10-15	3	1%	99%	0	0%	100%	0	0%	100%
15-20	0	0%	99%	0	0%	100%	0	0%	100%
20-30	2	1%	100%	0	0%	100%	0	0%	100%
Total	204	100%	-----	11	100%	-----	4	100%	-----

Source: Author.

According to the results in Table 18, all respondents from medium and large enterprises broke off in less than 5 minutes. About 98% respondents from small enterprises broke off in 10 minutes.

Table 19. Distribution of respondents according to breakoff time, different questionnaire designs regarding the main activity of enterprises

Time, in min.	Industrial enterprises			Trade enterprises			Service enterprises			Other enterprises		
	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.	Break.	Per.	Cum. per.
0-5	68	96%	96%	46	94%	94%	80	89%	89%	9	100%	100%
5-10	2	3%	99%	1	2%	96%	8	9%	98%	0	0%	100%
10-15	1	1%	100%	1	2%	98%	1	1%	99%	0	0%	100%
15-20	0	0%	100%	0	0%	98%	0	0%	99%	0	0%	100%
20-30	0	0%	100%	1	2%	100%	1	1%	100%	0	0%	100%
Total	71	100%	-----	49	100%	-----	90	100%	-----	9	100%	-----

Source: Author.

In Table 19 respondents' breakoff times are observed according to the main activity of the enterprises. It has been shown that 99% of respondents from industrial enterprises broke off the questionnaire in less than 10 minutes. Such respondents from service enterprises were 99%, and from trade enterprises 96%. All respondents from other enterprises broke off in less than 5 minutes.

4.2. Comparison of respondents' proportions according to their breakoff time

In this chapter the distributions of respondents according to their breakoff time levels are compared by the chi-square test for equality of three or more population proportions. Unfortunately, due to lack of data it was not possible to conduct the chi-square test for 15–20 minutes time level.

Table 20. Chi-square test for equality of three or more population proportion results, breakoff time observed, different questionnaire designs regarding presented pictures in the questionnaire

Time, in min.	Comm. prop.	Exp. break. – positive pictures	Exp. break. – negative pictures	Exp. break. – no pictures	Emp. chi-square	p-value
0-5	0.9269	69.52	68.59	64.89	0.667	0.7165
5-10	0.0502	3.77	3.72	3.52	1.555	0.4596
10-15	0.0137	1.03	1.01	0.96	0.003	0.9986
15-20	-----	-----	-----	-----	-----	-----
20-30	0.0091	0.68	0.68	0.64	0.948	0.6224
Total	-----	75	74	70	-----	-----

Source: Author.

In Table 20 the chi-square tests for equality of three or more population proportion results where breakoff time is observed, for different questionnaire designs regarding presented pictures in the questionnaire are observed. At the significance level of 5%, the null hypothesis cannot be rejected at all observed time levels. So, the structure of breakoff is the same at all three questionnaire designs across all given time levels.

Table 21. Chi-square test for equality of three or more population proportion results, breakoff time observed, different questionnaire designs regarding the number of questions presented to respondent per questionnaire screen

Time, in min.	Comm. prop.	Exp. break. – one question per screen	Exp. break. – group of questions	Exp. break. – all questions	Emp. chi-square	p-value
0-5	0.9269	142.75	58.40	1.85	0.765	0.6820
5-10	0.0502	7.74	3.16	0.10	0.124	0.9397
10-15	0.0137	2.11	0.86	0.03	2.138	0.3433
15-20	-----	-----	-----	-----	-----	-----
20-30	0.0091	1.41	0.58	0.02	0.453	0.7972
Total	-----	154	63	2	-----	-----

Source: Author.

According to Table 21, at the significance level of 5%, there is no statistically significant difference in the proportion of respondents' breakoff at all observed time for different questionnaire designs regarding the number of questions presented to a respondent per a questionnaire screen.

Table 22. Chi-square test for equality of three or more population proportion results, breakoff time observed, different questionnaire designs regarding the legal form of enterprises

Time, in min.	Comm. prop.	Exp. break. – joint stock enterprises	Exp. break. – limited liability enterprises	Exp. break. - simple limited liability enterprises	Emp. chi-square	p-value
0-5	0.9269	5.56	188.17	9.27	1.567	0.4569
5-10	0.0502	0.30	10.20	0.50	2.238	0.3266
10-15	0.0137	0.08	2.78	0.14	0.240	0.8870
15-20	-----	-----	-----	-----	-----	-----
20-30	0.0091	0.05	1.85	0.09	0.159	0.9235
Total	-----	6	203	10	-----	-----

Source: Author.

Table 22 reveals that there is no statistically significant difference in the proportion of respondents' breakoffs according to the observed time categories between respondents from enterprises with a different legal form.

Table 23. Chi-square test for equality of three or more population proportion results, breakoff time observed, different questionnaire designs regarding the size of enterprises

Time, in min.	Comm. prop.	Exp. break. – small enterprises	Exp. break. – medium enterprises	Exp. break. – large enterprises	Emp. chi-square	p-value
0-5	0.9269	189.10	10.20	3.71	1.269	0.5301
5-10	0.0502	10.25	0.55	0.20	0.852	0.6532
10-15	0.0137	2.79	0.15	0.05	0.224	0.8942
15-20	-----	-----	-----	-----	-----	-----
20-30	0.0091	1.86	0.10	0.04	0.148	0.9285
Total	-----	204	11	4	-----	-----

Source: Author.

Results from Table 23 are in line with all the previous results. Again, it can be concluded that there is no statistically significant difference in respondent's proportions according to the observed time levels. So, this conclusion is valid for enterprises of different size as well.

Table 24. Chi-square test for equality of three or more population proportion results, breakoff time observed, different questionnaire designs regarding the main activity of enterprises

Time, in min.	Comm. prop.	Exp. break. – industrial enterprises	Exp. break. – trade enterprises	Exp. break. – service enterprises	Exp. break. – other enterprises	Emp. chi-square	p-value
0-5	0.9269	65.81	45.42	83.42	8.34	3.730	0.2921
5-10	0.0502	3.57	2.46	4.52	0.45	4.933	0.1767
10-15	0.0137	0.97	0.67	1.23	0.12	0.334	0.9536
15-20	-----	-----	-----	-----	-----	-----	-----
20-30	0.0091	0.65	0.45	0.82	0.08	1.465	0.6904
Total	-----	71	49	90	9	-----	-----

Source: Author.

Finally, in Table 24 distributions of respondents' breakoffs are compared for enterprises of different main activity. The chi-square results lead to the conclusion that there is no statistically significant difference in respondents' proportions at all observed breakoff time levels.

4.3. Regression modelling of the breakoff function

In order to estimate the breakoff function linear regression modelling is applied. According to Figure 2, where cumulative distribution of all respondents according to breakoff time is shown, the relation between these two variables is not linear. Because of that, in the linear regression model cumulative proportion of respondents is observed as dependent variable, whereas the logarithm of time is observed as independent variable.
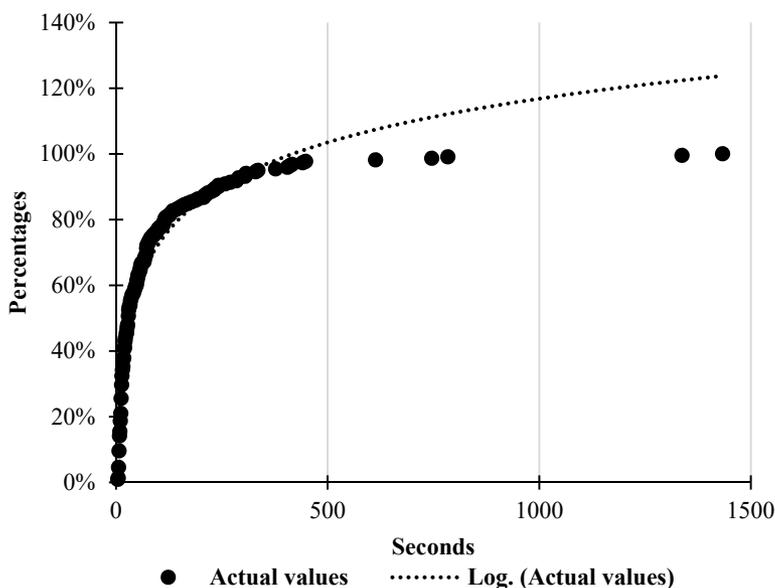
Figure 2. Cumulative distribution of respondents according to breakoff time, all respondents
Source: Author.

When actual and regression are compared in Figure 2, it can be concluded that the regression function follows the actual values quite well. The results in Table 25 confirmed that the estimated regression model fits well the actual values.

Table 25. Regression analysis results, dependent variable cumulative proportion of respondents, independent variable ln time (in seconds), breakoff time observed, results for all observed categories of questionnaire designs

Questionnaire designs	No. of. break.	No. of unique times	R square	Reg. stand. error	Intercept		Variable ln time	
					Estim.	p-value	Estim.	p-value
Total	219	110	0.9467	0.0568	-0.1597	<0.0001	0.1923	<0.0001
Pictures								
Positive pictures	75	54	0.9409	0.0671	-0.2235	<0.0001	0.2095	<0.0001
Negative pictures	74	53	0.9285	0.0728	-0.1701	<0.0001	0.1989	<0.0001
Without pictures	70	50	0.9829	0.0358	-0.2421	<0.0001	0.2050	<0.0001
Questions								
One ques. per screen	154	74	0.9061	0.0794	-0.1106	0.0008	0.1941	<0.0001
Group of questions	63	56	0.9727	0.0464	-0.3839	<0.0001	0.2139	<0.0001
All questions	2	2	1.0000	0.0000	-1.0019	-----	0.4666	-----
Legal form								
Joint stock ent.	6	6	0.9779	0.0519	-0.3996	0.0065	0.2280	0.0002
Limited liability ent.	203	104	0.9434	0.0585	-0.1466	<0.0001	0.1900	<0.0001
Simple lim. liab. ent.	10	10	0.8510	0.1239	-0.4455	0.0192	0.2786	0.0001
Size								
Small enterprises	204	104	0.9425	0.0598	-0.1524	<0.0001	0.1905	<0.0001
Medium enterprises	11	11	0.9810	0.0438	-0.3273	<0.0001	0.2523	<0.0001
Large enterprises	4	4	0.9429	0.0945	-0.2907	0.2225	0.2201	0.0290
Main activity								
Industrial enterprises	71	50	0.9752	0.0432	-0.2922	<0.0001	0.2217	<0.0001
Trade enterprises	49	37	0.9256	0.0781	-0.1828	<0.0001	0.2008	<0.0001
Service enterprises	90	56	0.9564	0.0587	-0.1731	<0.0001	0.1930	<0.0001
Other enterprises	9	9	0.9006	0.1026	-0.4171	0.0133	0.2773	0.0001

Source: Author.

In Table 25, except for all respondents, main regression results for different categories of respondents are given as well. Almost all estimated regression model estimated regression coefficients are statistically significant at the significance level. Only in the regression model where large enterprises are observed, the constant term in not statistically significant at 5%. However, this is mainly due to fact that the regression model was estimated based on only four respondents.

Table 26. Speeding and completion times estimates based on the regression models, breakoff time observed, results for all observed categories of questionnaire designs

Questionnaire designs	Speeding breakoff time				Completion breakoff time			
	Sec.	Min.	No. of faster break.	Per. of faster break.	Sec.	Min.	No. of slower break.	Per. of slower break.
Total	2	0.04	0	0.00%	417	6.94	7	3.20%
Pictures								
Positive pictures	3	0.05	0	0.00%	343	5.72	4	5.33%
Negative pictures	2	0.04	0	0.00%	358	5.97	4	5.41%
Without pictures	3	0.05	0	0.00%	428	7.13	1	1.43%
Questions								
One question per screen	2	0.03	0	0.00%	306	5.09	9	5.84%
Group of questions	6	0.10	3	4.76%	644	10.74	3	4.76%
All questions	9	0.14	0	0.00%	73	1.22	0	0.00%
Legal form								
Joint stock enterprises	6	0.10	0	0.00%	464	7.73	0	0.00%
Limited liability enterprises	2	0.04	0	0.00%	418	6.97	7	3.45%
Simple limited liability enterprises	5	0.08	0	0.00%	179	2.99	1	10.00%
Size								
Small enterprises	2	0.04	0	0.00%	423	7.05	7	3.43%
Medium enterprises	4	0.06	0	0.00%	193	3.21	0	0.00%
Large enterprises	4	0.06	0	0.00%	352	5.86	0	0.00%
Main activity								
Industrial enterprises	4	0.06	0	0.00%	340	5.67	2	2.82%
Trade enterprises	2	0.04	0	0.00%	362	6.03	3	6.12%
Service enterprises	2	0.04	0	0.00%	436	7.26	2	2.22%
Other enterprises	5	0.08	0	0.00%	166	2.76	1	11.11%

Source: Author.

In the same way as at the estimated completion functions, estimated breakoff functions can be used to identify respondents who break off too quickly or too slowly. However, according to Table 26, only respondents who had the questionnaire version where the group of questions were presented to a respondent per questionnaire screen had speeders. On the other hand, some too slow breakoff respondents were identified at some questionnaire designs.

5. Discussion

In the paper the completion and breakoff functions of a business web survey are estimated. The estimations were conducted at the overall level, for all respondents together, but also according to certain characteristics of respondents. The estimates for the completion functions can be found in Table 12, whereas the estimates for the breakoff functions are given in Table 25. Due to paper length limits estimated functions are graphically presented only at the overall level for all respondents. In this way, in Figure 1 the completion function for all respondents is given, whereas in Figure 2 the breakoff function is shown.

According to the first research hypothesis, the optimal survey lengths, estimated by the completion and breakoff functions, are different for different questionnaire designs and for respondents of different characteristics. The optimal survey length here is defined as the time in which all respondents should have completed the survey. In this way, the optimal survey length is observed as the longest time in which respondents should have completed the survey. The longest time can be calculated by observing the estimated completion functions and calculate times when 100% as value of dependent variable is given.

Optimal survey lengths are given in Table 13. According to the results, the optimal survey length, when all respondents are taken into account, is 1,047 seconds or 17.45 minutes. If only results in Table 13 are observed, it can be concluded that for some questionnaire designs and for some characteristics of respondents the optimal survey length is not so different, whereas in some cases the difference is considerable. The difference in the optimal survey length seems to be rather small in the case of questionnaire designs where different pictures were presented to the respondents and in the case when enterprises are observed according to their main activity. In other cases, there is always one category at which the optimal survey length is considerably lower than in other categories. Because of that, researches have to take into account different questionnaire designs and characteristics of respondents during the process of developing the questionnaire design. In support of this process, the results from Chapter 3.2., where distributions of respondents according to the completion times are observed, should be consulted as well. In this way, generally speaking, the first research hypothesis can be accepted.

The estimated completion and breakoff functions could be used to detect respondents who were too fast or too slow. Too fast respondents did not have enough time for full cognitive perception of questions. Also, respondents could break off before they read the introductory survey page. On the other hand, too slow respondents probably just opened the survey and worked something else. This could happen in business surroundings very often because employees who are participating in the

survey could be interrupted, for example, by e-mail, phone call or other work colleagues. It could happen that a respondent left the survey open and then quickly provides answers, also. In both cases, in cases of too fast and too slow respondents, such respondents should be omitted from the analysis because their answer cannot be declared valid. If such answers would not been omitted, they could have certain impact on survey results and, consequently, wrong conclusions could be made.

In order to detect respondents who were too fast or too slow, the completion and breakoff functions should be used to calculate the limit times. The limit times are calculated by taking into account that the cumulative percentage of respondents is 0% and 100% respectively. In this way, two times are obtained in the completion and breakoff functions. If respondent's survey time is lower than the lower limit, the conclusion is that this respondent was too fast. On the other hand, if the respondents' survey time was longer than the upper limit, the conclusion is that those respondents were too slow. Those limits calculated based on the completion and breakoff functions are presented in Table 13 and Table 26, respectively. Except to omit too fast and too slow respondents which answers could have significant impact on the survey results, the results of this analysis could be used for further improvement of the survey questionnaire design. Finally, it can be concluded that the second research hypothesis can be accepted.

6. Conclusion

The response rates in web surveys tend to be very low. Because of that, researchers should invest more effort to reach some appropriate response rate levels. One of the ways to increase response rates is to carefully design the questionnaire and its length. However, the question is how to know whether the questionnaire is too long or too complicated for respondents or not.

In the paper the completion and breakoff functions are proposed to be used to determine the optimal survey length. Those functions are estimated by observing cumulative proportion distribution of respondents according to their completion and breakoff times in a web business survey conducted on a sample of enterprises in Croatia. In order to keep things as simple as possible and therefore easily interpretable at the same time, a simple linear regression approach to the completion and breakoff functions estimation was used.

After the completion and breakoff functions have been estimated, the possibilities of their use are shown. It has been illustrated how the completion function can be used to estimate the optimal survey length or to estimate time in which respondents, even those who are inexperienced or unfamiliar with the survey, should complete the survey. In the paper additional possible use of the completion and breakoff functions is

presented as well. So, the use of the completion and breakoff functions in detecting too fast and too slow respondents is recommended as well.

However, while the regression diagnostic results have shown that the estimated completion and breakoff functions are of good fit, the estimates are valid only for the observed survey. So, the main limitation of the proposed approach is that a pilot study is needed to be able to estimate the completion and breakoff functions. In further research a way of estimating standardized completion and breakoff functions which could be used in business web survey should be found.

References

1ka, (2020a). Osnovna priporočila: Kaj pomeni kompleksnost ankete? (in Slovenian). Available at: https://www.1ka.si/d/sl/spletne-ankete/osnovna-priporocila/kaj-pomeni-kompleksnost-ankete [02 June 2020].

1ka, (2020b). Osnovna priporočila: Kako dolga naj bo moja anketa? (in Slovenian). Available at: https://www.1ka.si/d/sl/spletne-ankete/osnovna-priporocila/kako-dolga-naj-bo-moja-anketa [02 June 2020].

Blumenberg, C., Zugna, D., Popovic, M., Pizzi, C., Barros, A.J.D., Richiardi, L., (2018). Questionnaire Breakoff and Item Nonresponse in Web-Based Questionnaires: Multilevel Analysis of Person-Level and Item Design Factors in a Birth Cohort, Journal of Medical Internet Research, Vol. 20, No. 12., pp. 1–11, DOI: 10.2196/11046.

Crawford, S. D., Couper, M. P., Lamias, M. J., (2001). Web surveys: Perceptions of burden, Social Science Computer Review, Vol. 19, No. 2, pp. 146–162.

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., Tu, X. M., (2014). Log-transformation and its implications for data analysis, Shanghai archives of psychiatry, Vol. 26, No. 2, pp. 105–109.

Galesic, M., (2006). Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey, Journal of Official Statistics, Vol. 22, No. 2, pp. 313–328.

Groves, R. M., (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys, Public Opinion Quarterly, Vol. 70, No. 5, pp. 646–675.

Göritz, A. S., (2006). Incentives in Web Studies: Methodological Issues and a Review, International Journal of Internet Science, Vol. 1, No. 1, pp. 58–70.

Kish, L., (1995). Survey Sampling, New York: John Wiley & Sons.

Mittereder, F. K., (2019). Predicting and Preventing Breakoff in Web Surveys, dissertation, University of Michigan, pp. 1–762, available at: https://isr.umich.edu/wp-content/uploads/2019/04/Dissertation_FinalSubmit_V03.pdf.

Narodne novine, (2007). The National Classification of Economic Activities. Official Gazette, 16(58).

Narodne novine, (2011). The Enterprises Act. Official Gazette, 20(152).

Narodne novine, (2015). The Accounting Act. Official Gazette, 24(78).

Peytchev, A., (2009). Survey Breakoff, Public Opinion Quarterly, Vol. 73, No. 1, pp. 74–97.

Peytchev, A., (2011). Breakoff and Unit Nonresponse Across Web Surveys, Journal of Official Statistics, Vol. 27, No. 1, pp. 33–47.

Žmuk, B., (2017). Impact of Questionnaire Length and Complexity on Survey Time: Comparison of Two Business Web Questionnaire Versions, Metodološki zvezki - Advances in Methodology and Statistics, Vol. 14, No. 2, pp. 49–74.

Credibility of disability estimates from the 2011 population census in Poland

Elżbieta Gołata[1]**, Grażyna Dehnel**[2]

ABSTRACT

The problem of disability is perceived as one of the most serious social issues faced by the contemporary society. The number of people with disability is consistently rising for a variety of reasons, including the aging of the population. Data on disability are collected through numerous statistical surveys, among which censuses are the most wide-scale ones. In the period between the 2002 and 2011 censuses (the last two censuses conducted in Poland), a 14% decrease in the number of people with disabilities was observed. However, it should be emphasised that significant modifications were introduced to the methodology of the last census. Population census 2011 was the first census in Poland combing administrative data sources and the survey sampling method. The main objective of the study is to assess the quality of estimates relating to the number of disabled persons, obtained on the basis of the 2011 census data. It is a comparative study aimed at identifying the similarities and discrepancies between the estimates, and determining the size and source of these discrepancies. The analysis takes into account such aspects as the measurement methods, the definitions and criteria of disability, the voluntary nature of the question, and the quality of the information on disability obtained from various sources.

Key words: disability, health condition, demographic processes, quality of a statistical survey.
JEL: I15, I18, J11, J14.

1. Introduction and motivation

The measurement of disability is a particular challenge for statisticians. Results of ad hoc surveys tend to indicate higher proportions of disabled people than census-based estimates (Loeb, 2016b; Mont, 2007; US Census Bureau, 2017; WHO & The World Bank, 2011). Data from censuses are usually an important source of information about disability, especially in countries which do not conduct regular surveys on this

---

[1] Poznań University of Economics and Business. Department of Statistics, Poznań, Poland.
  E-mail: elzbieta.golata@ue.poznan.pl. ORCID: https://orcid.org/0000-0001-5570-9576.
[2] Poznań University of Economics and Business. Department of Statistics, Poznań, Poland.
  E-mail: grazyna.dehnel@ue.poznan.pl. ORCID: https://orcid.org/0000-0002-0072-9681.

topic. The interest in this problem was motivated by the discrepancy between expectations and the actual estimates of the number of disabled persons in the 2011 census in Poland (NSP 2011). Given the continuing aging of Poland's population, it was reasonable to expect a higher number and percentage of disabled persons. However, according to the actual census results, the number of disabled persons was put at 4,697,000, which means a decline of over 750,000 compared to the 2002 census, a decrease in the share of disabled persons from 14.3% to 12.2%. This gave rise to criticism levelled against the approach used in 2011 census, in particular the fact that replies to the question about disability were voluntary (Dz.U. Nr 47 poz. 277, 2010). This solution was adopted in view of the sensitive nature and the topic. It raised reservations and triggered controversy within the scientific community, especially given the large number of refusals in the survey (Slany, 2014).

Any such assessment is further complicated by the multiplicity of definitions of disability, regulations used for purposes of administrative registers or social assistance in Poland (Antczak, Grabowska, & Polańska, 2018; Dehnel & Klimanek, 2016). There are also differences between approaches adopted in surveys conducted by international organisations (Altman, 2016; Molden & Tøssebro, 2010; Mont, 2007; UN, 2008b; Van Oyen, Bogaert, Yokota, & Berger, 2018). Depending on the survey type, the definition and criteria used in identifying people with disabilities, the population of disabled persons in Poland could range from 4.9 to 7.7 million.

The main aim of the study described in this article is to assess the quality of estimates of the number of disabled persons obtained on the basis of data from the 2011 census in Poland. It is a comparative study aimed at identifying similarities and discrepancies between estimates, and determining the size and source of these discrepancies.

The first part contains an overview of definitions of disability used in various surveys including references to the literature and results obtained. The overview comprises definitions and classifications used in population census, ad hoc survey modules and administrative registers. Another aspect addressed in this respect is the question of the quality of disability information obtained from various sources. In particular, a number of reasons for this multiplicity of definitions are identified, which prevents direct comparability between different surveys, although they do have their social justifications.

The next part is devoted to the presentation of international initiatives aimed at ensuring the validity of estimates, as well as their reliability and comparability. Particular reference is made to recommendations concerning the measurement of disability in censuses based on the results of the Washington Group on Disability Statistics (UN, 2008b). The presentation includes methods of measurement as well as similarities and differences between various approaches.

The reliability of the data collected during the 2011 census is assessed in two ways. Firstly, metadata and characteristics of the 2011 census are compared with other surveys described previously. Secondly, methods of demographic analysis are applied to assess the census results by comparing them with those obtained in the 2002 census. Unfortunately, the comparability of results produced in both surveys is limited by the fact that different definitions were used in both cases. However, an attempt was made to provide detailed explanations for specific discrepancies.

The article ends with a discussion of the results.

2. Measurement of disability in statistical surveys

According to the first World Report on Disability published by the World Bank and the World Health Organisation, "more than billion in the world live with some form of disability" (WHO & The World Bank, 2011). The problem of disability is becoming increasingly widespread and is now estimated to affect about 15% of the world population. A better knowledge of the needs and problems faced by disabled persons is the key to providing them with effective help. We are also becoming increasingly aware of the fact that most of us, at some point in our lives, will experience some form of disability. Given the ubiquity and scale of this phenomenon, it is more and more frequently addressed in discussions and activities undertaken not only at the local and national level, but is also tackled globally.

Disability can be approached from different perspectives. There are two approaches in the literature: the medical and social view (Dehnel & Klimanek, 2016; US Census Bureau, 2017; WHO, 2002). Some studies also distinguish a functional approach or use other concepts such as the biological model (Antczak et al., 2018). When analysing disability research, it is useful to refer to the recommendations of WHO (2002). However, even they do not dispel all the existing doubts. According to the medical model proposed by WHO, disability is defined as "a feature of the person, directly caused by disease, trauma or other health condition, which requires medical care provided in the form of individual treatment by professional" (WHO, 2002). Under the social model, disability is viewed not as an attribute of an individual but as a social problem created by an unaccommodating physical environment, which demands a political response. WHO (2002) experts believe that "disability is a complex phenomenon that is both a problem at the level of a person's body, and a complex and primarily social phenomenon". For this reason they stress that disability involves an interaction between features of the person and characteristics of the environment in which the person lives (Figure 1). Because "some aspects of disability are almost entirely internal, while others are almost entirely external", the appropriate approach to disability at the individual level should combine both social and medical responses.

The approach proposed by WHO as the International Classification of Functioning, Disability and Health (ICF) (UN, 2008b; US Census Bureau, 2017; WHO, 2002), is universal and can be used to describe and measure disability for purposes of many sectors (medicine, economy, social policy). This approach ensures comparability of results obtained in different surveys not only between sectors but also at the international level. The basic idea behind this classification is that "every human being can experience a decrement in health and thereby experience some disability". This means that disability is defined by assessing the person's health in the context of their relationship with the environment, taking into account three levels of limitations: (i) Body Functions and Structures, (ii) Activity (iii) Participation. For example, Antczak et al. (2018) distinguish limitations of body functions and abilities (e.g. a blind person cannot see); limitations of activity (the same person may experience difficulties with moving, preparing meals, self-care, etc.); limited participation in social life.
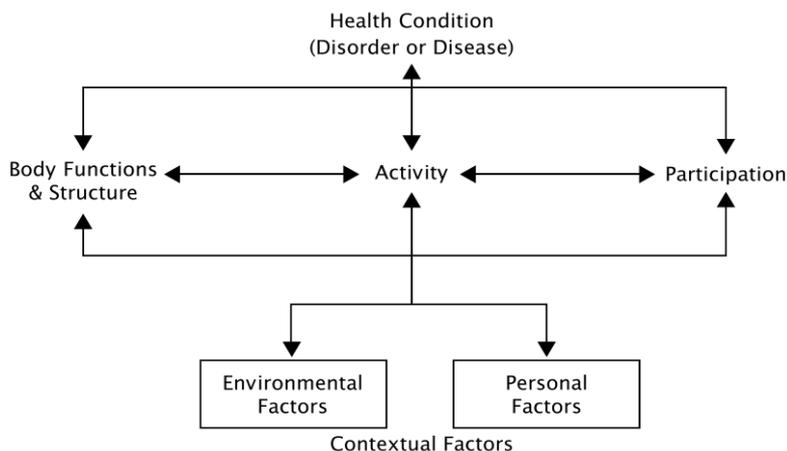
Figure 1. The International Classification of Functioning, Disability and Health Conceptual Model of Disability

Source: WHO 2002.

Definitions and classifications of disability according to the functional model are matched by specific methods of measurement proposed by the Washington Group on Disability Statistics (UN WG). The UN WG was created in 2001 as a result of the International Seminar on the Measurement of Disability, which sought to propose universal measurement tools that could ensure international comparability. This led to the development of a Short Set of Questions (UN WG, 2006). During the Global Disability Summit 2018, the World Bank Group, together with other participants, announced the Summit's Charter for Change, containing a list of 10 pledges (WBG,

2018a, 2018b) aimed at accelerating global measures for the equalization of opportunities for disabled persons and counteracting their social exclusion. The list of commitments was created to support the goals of the 2030 Agenda for Sustainable Development adopted during the UN Summit (UN, 2015b). The charter included a commitment to gather comparable data according to best practices and world standards, with special emphasis on the short set of questions developed by UN WG. Adopted in 2006, the set was recommended by the UN for the census rounds in 2010 and 2020 (UN, 2008b; UN WG, 2006; US Census Bureau, 2017). The proposed set contains the following six questions:

1) Do you have difficulty seeing, even if wearing glasses?
2) Do you have difficulty hearing, even if using a hearing aid?
3) Do you have difficulty walking or climbing steps?
4) Do you have difficulty remembering or concentrating?
5) Do you have difficulty (with self-care such as) washing all over or dressing?
6) Using your usual (customary) language, do you have difficulty communicating, for example understanding or being understood?

Each question can be answered with four replies: (i) No – no difficulty, (ii) Yes – some difficulty, (iii) Yes – a lot of difficulty, (iv) Cannot do it all. These replies can be used to establish the degree of ability limitations from mild to severe. According to the recommendations of UN WG, the population of disabled persons includes all those who indicated the presence of difficulties in at least one of the core functional domains (questions 1–6) by choosing options (iii) or (iv) (UN, 2008b, 2008a, 2015a; US Census Bureau, 2017).

The recommendations of UN WG were taken into account in the recommendations prepared before the census rounds in 2010 and 2020. It was agreed that set of questions was an appropriate tool for measuring disability of persons aged 5 and older. However, given the limited scope of a census, as a survey designed to collect information about multiple domains, only the first four questions were included in the census questionnaire; the full set was to be used in ad hoc surveys devoted specifically to disability. It is also emphasized that because the concept of disability can be differently understood by respondents, it is crucial that the survey questionnaire should be formulated carefully in order to ensure correct identification of the population. The WG also recommended that questions about disability should be put individually to each respondent and control questions about the presence of disabled persons in the household should be avoided, such as "Is there a disabled person in the household?" (US Census Bureau, 2017).

The WG method of measurement has been evaluated and compared to other approaches like the Model Disability Survey (Sabariego et al., 2015). The discussion conducted (Madans, Mont, & Loeb, 2015) exemplifies the crucial role played by the

definition and the choice of the measurement method. The gap in the measurement methodology was filled by Loeb (2016a) and Meltzer (2016), who published a list of challenges that need to be addressed with respect to the measurement of disability among children and the proposal of a census module devoted to disability, which was developed by Crialesi, De Palma and Battisti (2016). Another crucial problem involved in the measurement of disability is the correct identification of environmental and contextual factors, which is discussed by Altman and Meltzer (2016). Problems of measuring disability among people living in group quarters and their impact on the comparability of international estimates were addressed by Cambois, Jagger, Nusselder, Van Oyen and Robine (2016).

Loeb (2016b) notes that about 30 countries reported using the short set of 6 WG questions in the 2010 census round. In order to determine the impact of the tool used to measure disability on the final estimates, WG researchers conducted a voluntary survey involving about 120 countries, asking the respondents to indicate the type of disability model used and the exact wording of the questions (Table 1). A clear distinction was made between the medical model focusing on impairments and types of disability and the social model emphasizing activity limitations, including the WG short set of questions. The survey involved countries where disability data were collected in the census (26 countries) as well as those where a sample survey was used (25 countries). The response rate was 54%. It turned out that out of the countries where disability data came from the census only Aruba (6 questions) and Israel (4 questions) used the tool according to the WG recommendations. Turkey also used 6 similar questions in the 2011 census, but they were not identical to those proposed by the WG. The estimated shares of disabled people varied considerably, ranging from under 1% (the Dominican Republic) to 12.9% (Peru).

Estimated percentages based on sample surveys were generally higher but also showed a great degree of variation, with values ranging from 1.4% in Togo, 2.0% in Yemen or 2.6% in Lesotho, to 12.5% in the Netherlands, 13.8% in Poland, 14.3% in Canada, 14.8% in Israel and 16.6% in New Zealand (Loeb, 2016b). In the European Health Interview Survey (EHIS) conducted in Poland 9 questions about activity limitations were used. The approach adopted in Thailand, Poland, Hungary and the Netherlands was similar to the one proposed by the WG, however, the question wording was not identical. Estimates obtained in these countries, except for Thailand, are also believed to be similar to those expected under the WG approach. Based on this analysis, (Loeb, 2016b) notes that the use of the definition of disability based on information about impairments resulted in obtaining the lowest estimate of the share. Estimates of the share of disabled people based on the WG approach are regarded as moderate, except for the value for Israel (1.4%).

Table 1. Models of disability used in the 2010 round of censuses and in sample surveys

Disability model	Census	Sample survey
Medical model – loss of ability (impairments)	8	5
Social model – activity limitations	11	13
WG short set of 6 questions	7	6

Source: Based on (Loeb, 2016b).

Unlike censuses, sample surveys can cover a subpopulation of a certain age. Surveys with the highest estimates of disability covered subpopulations of a specified age: in Poland – aged 15 and older, in the Netherlands – aged 12 and older or aged 55 and older, in Hungary – aged 15 and older, in Israel – aged 20 and older (Loeb, 2016a). As the share of disabled people in the population increases with age, so the way the target population is defined may affect the final estimates. Results of the analysis lead Loeb (2016b) to conclude that there are considerable differences between approaches, definitions and measurement methods, which render international problematic, but there are also good reasons to question the usefulness of estimates obtained in each country for national purposes.

In both rounds of the EHIS that took place in Poland (2009 and 2014), the same standard question was used[3], as recommended by Eurostat. Data obtained in this way are supposed to enable the estimation of the level of disability that is comparable between European countries and to estimate the Healthy Life Years indicator (HLY) (Bogaert, Van Oyen, Beluche, Cambois, & Robine, 2018; EHLIS, 2015; Van Oyen et al., 2018). However, the Eurostat guidelines do not refer to the WG recommendations. The WG short set of questions is not used in the disability module of the Polish Labour Force Survey (LFS), which is conducted according to the recommendations of Eurostat (GUS, 2012). However, according to the UN recommendations for census rounds in 2010 and 2020 (UN, 2008a, 2015a), disability should be viewed in the light of the ICF model and measured using the WG short set of questions. The UN definition of disabled people includes persons who are more likely than the general population to experience limitations in the performance of certain tasks or when trying to participate in activities associated with their social roles.

---

[3] The question was formulated as follows (GUS, 2016): Do you experience a health-related limitation in your ability to perform typical activities of daily life that has lasted for at least 6 past months?
- Yes, a serious limitation.
- Yes, but not very serious.
- No, I have not experienced any limitations.

3. The definition of disability used in official statistical surveys in Poland

A detailed description of disability surveys in Poland, including information about their frequency and the scope of published results, and, above all, methods of measurement, can be found in the paper by Antczak, Grabowska & Polańska (2018). A detailed presentation of the method of measuring disability in Polish censuses can be found in the paper by Dehnel and Klimanek (2016). With regard to the 2002 census, the authors point out that the medical definition of disability, focusing on the degree of impairment, was replaced by the definition involving limitations in the performance of basic activities regarded as typical for a given age (Dehnel & Klimanek, 2016). The definition of a disabled person in the 2011 census included an additional note that the duration of the experienced limitation should be at least 6 months (GUS, 2011a). In the revised definition the number of degrees of activity limitations was extended from "complete" and "serious" in the 2002 census, to "complete", "serious" and "moderate" in the 2011 census.

According to the definition used in the censuses a person was regarded as disabled if they could present an appropriate decision issued by an authorised body or, in the absence of such a document, if they experienced activity limitations (Antczak et al., 2018; Dehnel & Klimanek, 2016; GUS, 2013a). This means that the definition used in the censuses reflects two views of disability: the formal indication confirmed by a legal decision (disability in the legal sense), and the subjective indication of a person who experiences activity limitations (disability in the biological sense). Therefore, when describing disability in Poland, one has to take into account the existing regulations in this respect, which can have a significant influence on the final estimates. The act on social and occupational rehabilitation and the employment of disabled persons, which has been in effect since 1997, despite numerous amendments, retains the same definition. In the act disabled persons are defined as those whose physical, psychological or mental condition creates a permanent or temporary limitation in the performance of social roles, and, in particular, limits their ability to work[4]. The unchanged definition of a disabled person seems to guarantee the comparability of the population of people classified as disabled in the legal sense in the 2002 and 2011 census. However, the presence of disability in Poland is assessed by different institutions and for different purposes and not all statements of disability can be used to claim disability discounts or allowances (GUS, 2011b, 2016). Although the legal definition of disability is the same, the two systems of disability assessment existing

---

[4] „Niepełnosprawnymi są osoby, których stan fizyczny, psychiczny lub umysłowy trwale lub okresowo utrudnia, ogranicza bądź uniemożliwia wypełnianie ról społecznych, a w szczególności ogranicza zdolności do wykonywania pracy zawodowej".

in Poland[5] make it difficult to obtain reliable information about the number of people that actually have official decisions confirming disability, which is reflected by census data. It should be noted that in the surveys conducted by Statistics Poland, including the LFS[6], the EU Statistics on Income and Living Conditions (EU-SILC) and in statistical reporting, disabled persons are identified only on the basis of official decisions, while in the European Health Interview Survey, like in the census, both kinds of disability are taken into account (the legal and biological sense). Moreover, the two categories of disability are not exclusive (Table 2).

Table 2. Categories of disability according to the definition used in 2011 Census

BIOLOGICAL	LEGAL
disabled persons EXCLUSIVELY in the BIOLOGICAL SENSE, i.e. those who did not have an official decision/statement of disability but who felt that their ability to perform basic activities typical for their age was completely, seriously or moderately limited.	disabled person EXCLUSIVELY in the LEGAL SENSE, i.e. those who had a valid statement of disability but did not report any limitations in the ability to perform basic activities typical for their age
	disabled persons in the LEGAL SENSE, i.e. those who had a valid statement of disability issued by an authorised body: - by ZUS for purposes of disability allowances, - by district and provincial disability evaluation boards for other purposes

LEGAL AND BIOLOGICAL
disabled persons in the LEGAL AND BIOLOGICAL SENSE i.e. those who had a valid statement of disability and reported a completely, seriously or moderately limited ability to perform basic activities typical for their age

Source: (GUS, 2013b).

Each survey is based on different definitions and classifications. This limits the possibility of making comparisons. For 2011, there are virtually no comparable data about the number of disabled persons. Only in the case of disability in the legal sense is it possible to compare census data with those collected in the LFS. According to the 2011 Census, the number of disabled persons in the legal sense was equal to almost 3,131 thousand, while according to the LFS, it was 3,505.5 thousand (Table 3). This

---

[5] For purposes of disability allowances, disability assessment is conducted by the Social Insurance Institution and for other purposes – by district and provincial disability evaluation boards.

[6] With the exception of the ad hoc module conducted as part of the LFS in the 2nd quarter of 2011, which took into account disability in the legal and biological sense (GUS, 2012) according to the Commission Regulation (EU) No 317/2010 of 16 April 2010.

means that the census count was lower by 374 thousand. In addition, the subpopulation of disabled persons in the LFS, identified within the total population aged 15 and older, included people aged 16 and older, who had received statements about the degree of disability or work disability. Consequently, the estimated count in the census should be decreased by subtracting disabled persons under the age of 16, i.e. 129,950 persons.

The second point that needs to be emphasized is the result of comparing the LFS data with data from the EHIS for 2009 and 2014. Estimates of the number of disabled persons for both years are higher than the LFS estimate: by about 18% (630 thousand) in 2009, and by about 10% (333 thousand) in 2014 (Table 3). As already pointed out, in official publications (GUS, 2011b, 2016), estimates from both surveys were based on respondents' declarations. The results are therefore not fully comparable. However, this is more the case with the scale of the phenomenon, and less so with respect to the structure of the distribution. According to GUS, the discrepancy between the results could be due to different objectives of each survey, which may have affected respondents' answers.

Table 3. Disabled persons in the legal sense in Poland on the basis of selected sources

Survey year	Source of data			
	EHIS (thousand)	EHIS (thousand) (aged 16+)	LFS* (thousand)	Census (thousand)
2009	4155.3	3971.3	3505.5	
2011			3341.3 (3359**)	3131.5
2014	3801.5	3607.0	3272.0	

Note: * midyear data, **2nd quarter of 2011.

Source: (GUS, 2011b, 2016), http://www.niepelnosprawni.gov.pl/p,81,bael, 2011 Census.

4. Disability in census 2002 and 2011

Comparing disability from the last two censuses, it can be noticed that the decline in the number and percentage of disabled persons is not the same across different categories (Table 4). This can be observed with regard to the total number of disabled persons, but is due to the decline in the number of disabled persons in the legal sense. This was mainly the result of the complicated rules that were the basis for disability decisions, which discouraged many people from applying for a disability allowance, and, consequently, led to a decrease in this category (Dehnel & Klimanek, 2016; GUS, 2011b).

In contrast, the number of persons classified as disabled persons exclusively in the biological sense rose to about 1.5 million in 2011, i.e. by about 56% compared to the 2002 census. Their share increased from 2.63% to 4.07%. Consequently, the share of disabled in the biological sense, in both categories (exclusively biological as well as

biological and legal) rose by about 10% from 3.8 million in 2002 to over 4.2 million in 2011 (10.95% of the total population). However, a decrease in the number of disabled persons exclusively in the legal sense by nearly 70% corresponds to a decline in their share from 4.25% to just 1.24% and has an effect on the overall estimate showing a decline in the scale of disability.

Given the obligatory nature of the questions about disability in 2002 census (in contrast to the 2011 census, where these questions were asked only in the survey part), those results are generally regarded as reliable, putting the number of disabled persons at about 5.5 million, i.e. 14% of the total population. There is a lot of variation in the distribution of disabled persons by age. In the age group 0-3, the share of disabled persons is 1-2%, which increases to 3% for people aged 20. There is a marked growth in the share of disabled persons around the age of 40, when it rises from 8% to 30% over the following 15-year interval. In the group of people aged 75 and older the share of disabled persons is close to 50%.

Table 4. Disability according to censuses in 2002 and 2011

Disability category	NSP 2002		NSP 2011	
	number of persons	share of the population	number of persons	share of the population
Total	5 455 914	14.27	4 697 048	12.20
Legal	4 449 685	11.64	3 131 459	8.13
Exclusively legal	1 624 568	4.25	479 453	1.24
Biological	3 831 346	10.02	4 217 598	10.95
Exclusively biological	1 006 229	2.63	1 565 592	4.07
Biological and legal	2 825 117	7.39	2 652 006	6.89

Source: NSP 2002, NSP 2011.

The percentage of disabled persons increases with age almost exponentially (Figure 1). This is particularly true for men up to the age of 60. Attempts at modelling the share of disabled by age show a very good fit for the exponential function and, obviously, the second or possibly the third degree polynomial[7]. Nonetheless, one can observe evident changes in the scale of disability between the age of 50 and 70. There is a clear difference between the trends for men and women, which is not apparent up to the age of about 53. From the age of 50 to 60, the share of disabled persons in the male population is considerably higher and increases by over 50% (from 0.288 at 53 to 0.432 at 59). In the group of women of this age the share of disabled persons increases by only 13%. This dramatic increase in the share of disabled men over a relatively short period of time, followed by a period of relative stability and a decline to the level observed for women can be attributed to various causes. Certainly, it should be linked to the higher

---

[7] Coefficient of determination is equal to 0.92 for exponential functions and 0.98 for polynomials.

incidence of cardiovascular diseases, malignant tumours, as well as accidents and injuries in men. Another pattern observed in demographic research is the higher death rate for men compared to women of this age, the so-called excess male mortality[8] (Fihel, 2011; Szukalski, 2018). Another cause of this higher level of male disability may be associated with men's interest in pension-related benefits and their desire to obtain official confirmation of their health condition. A reverse trend can be observed for the population after the age of 80: the share of disabled women grows faster, with an increase of over 60%.

When analysing the share of disabled people by age based on data from the 2011 census (Figure 2, Table 5), one thing worth noting is that it is lower than the level observed in 2002. The difference becomes evident from the age of about 35 and is maintained up to the oldest age groups. The growth in the share of disabled persons after the age of 40 is clearly less abrupt than that observed in 2002. A similarly weaker increase can be seen in the pre-retirement age.
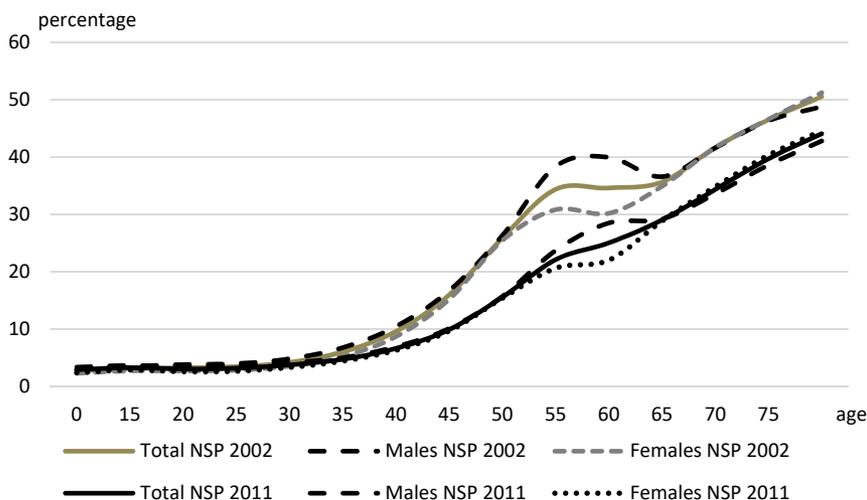
Figure 2. Disabled persons per 100 population by sex and age, Poland, NSP 2002 and NSP 2011

Source: NSP 2002, NSP 2011

Table 5. Share of disabled persons by sex and age in the censuses of 2002 and 2011

Age	NSP 2002			NSP 2011		
	Total	Men	Women	Total	Men	Women
Total	14.3	13.9	14.7	12.2	11.6	12.7
0-14	2.7	3.0	2.3	2.9	3.4	2.4
15-19	3.1	3.5	2.7	3.3	3.7	2.9
20-24	3.3	3.8	2.7	3.0	3.4	2.6

---

[8] Excess mortality is measured by the male/female ratio of death rates, the probability of dying or other life table parameters, such as life expectancy by age.

Table 5. Share of disabled persons by sex and age in the censuses of 2002 and 2011 (cont.)

Age	NSP 2002			NSP 2011		
	Total	Men	Women	Total	Men	Women
25-29	3.4	4.0	2.8	3.2	3.6	2.7
30-34	4.2	4.8	3.6	3.8	4.2	3.3
5-39	6.0	6.8	5.3	4.7	5.0	4.4
40-44	9.6	10.5	8.7	6.6	6.9	6.3
45-49	16.0	16.7	15.2	9.9	10.1	9.7
50-54	25.9	26.4	25.5	15.6	15.7	15.4
55-59	34.4	38.4	30.9	22.1	23.7	20.6
60-64	34.6	40.0	30.2	25.0	28.5	22.0
65-69	35.7	36.6	34.9	29.0	29.2	28.9
70-74	41.7	41.6	41.7	34.4	33.6	34.9
75-79	46.5	46.4	46.6	39.7	38.6	40.4
80 +	50.5	48.8	51.3	44.1	42.8	44.6

Source: NSP 2002, NSP 2011.

When data from both censuses are compared, one is struck by the higher share of disabled people in the youngest age groups, up to the age of 20. In contrast, the percentage of disabled persons in the population aged 40 and older is lower by 10 pp and drops by as much as 40% in the 50-54 age group.

It is the above differences in the estimates of disability based on two consecutive censuses that motivated our attempt to look for an explanation and evaluate the quality of estimates based on the 2011 census.

5. Assessment of the quality of disability estimates obtained in NSP 2011 – the use of the aging algorithm from demographic projections

When one compares the results of the 2002 and 2011 censuses it is important to keep in mind two significant differences between them. The first one is the change from the traditional method of conducting a census in 2002 to the mixed-mode approach adopted in 2011. The so-called short census form contained data obtained from administrative registers, while information provided in the ad hoc modules attached to the long questionnaire was collected in a 20% sample survey. The second difference is the voluntary nature of responses to the question about disability. This was the consequence of the provision of the Polish constitution (Article 51), which prohibits the imposition of an obligation to reveal information about one's health. The decision was motivated by the sensitive nature of such information. The question was only put to adult respondents who agreed to answer it, while information about children could only be provided by their parents or caretakers. Over 1.3 million respondents exercised their right to refuse to answer the question about disability. Dehnel and Klimanek

(2016) argued that there are good reasons to believe that this group included disabled people. The quality of estimates may also have been affected by the high rate of nonresponse in the survey.

The assessment of disability estimates obtained in the 2011 census poses a challenge not only because of a certain degree of ambiguity and changes of definitions but also for many other reasons. The main one is the lack of other sources of information about disability that could be used as a reference point in comparative analyses concerning the year of the census. Other problems stem from the use of different methods of census organisation, different sampling schemes, sample sizes and estimation methods.

The group of methods used to assess the quality of censuses exploiting existing data sources includes, among others, demographic analyses based on data from previous censuses, comparisons with administrative registers and with existing surveys, e.g. such as those focusing on household budgets or the labour force. Obviously, such methods do not eliminate the crucial problem due to the difference between the traditional (NSP 2002) and mixed-mode census (NSP 2011), but similarity of estimates obtained from independent surveys is the best evidence of their reliability and quality.

Accordingly, our assessment of the disability estimates from the 2011 census is made in reference to the data collected in 2002. The analysis was conducted by applying the cohort component method, which is used for constructing population projections. The method is based on the idea of a longitudinal study in which particular generations are tracked over intercensal periods. Life table parameters for successive single year of age, especially survival probabilities, were used to age the population into the future. As a result, a population projection was obtained, broken down by sex and age for the year of the next census. It was supplemented by a projection and aging of the number of births, accounting for disability. This study was based on unit-level data from both censuses shared by Statistics Poland for research purposes under a special agreement[9].

The applicability of the above method for the purpose of assessing the quality of census-based estimates of disability by category is seriously limited by the lack of information about survival rates for subpopulations of disabled persons[10]. As regards life table parameters for males and females, it was assumed that the survival probability for healthy and disabled persons is the same. Assuming a closed population, one could therefore expect that estimation results for 2011 should be higher than those actually

---

[9] Under the agreement, access to the sample survey data on disability from the 2011 census was granted to researchers via a computer located in the Statistical Office in Poznań. The data could be processed using the SAS software.

[10] Life tables published by the European Health & Life Expectancy Information System, EH&LEiS (http://www.eurohex.eu/IS/web/app.php/Ehleis/LifeGeographic?Typ=Life&SubTyp=None) also provide information about the population in total. By decomposing life expectancy one obtains an estimate of Healthy Life Years (HLY) by sex and age (Sullivan, 1971). However, data which would enable the construction of complete life tables are not publicly available.

observed. Obviously, the disabled population is not closed, not only because of foreign migrations, which, in the case of disabled people, tend to be negligible. However, with advancing age, the population of disabled people increases due to a higher incidence of diseases, accidents and injuries. We do not have information that could be used to determine whether the overestimation of the population of disabled persons due to the adoption of higher survival rates is compensated for by an underestimation resulting from the growing number of persons who become disabled with age. This is definitely a strong and controversial assumption. Despite these reservations, this method was used for comparative purposes in order to explain as best as possible the existing discrepancy in estimates.

The results of the prediction of the disabled population identified in 2002 into the future for all categories are about 10% lower than those obtained in the 2011 census (Table 6, Figure 3). For example, in the age group 35–39, the share of disabled persons obtained after being aged into the future is higher by 26%, and in the next three age groups, by as much as 35%. The differences are somewhat bigger for women than for men. It is worth noting that estimates for ages 60–70, recalculated with a 10-year shift, are in fact higher.

Table 6. The share of disabled persons by sex and age in NSP 2011 and the predicted share based on NSP 2002

Age	NSP 2011			prediction based on NSP 2002		
	Total	Men	Women	Total	Men	Women
Total	12.2	11.6	12.7	11.0	10.6	11.4
0-14	2.9	3.4	2.4	2.16	2.5	1.8
15-19	3.3	3.7	2.9	2.91	3.3	2.5
20-24	3.0	3.4	2.6	3.05	3.4	2.6
25-29	3.2	3.6	2.7	3.21	3.6	2.8
30-34	3.8	4.2	3.3	3.25	3.8	2.7
35-39	4.7	5.0	4.4	3.48	4.0	2.9
40-44	6.6	6.9	6.3	4.47	5.1	3.8
45-49	9.9	10.1	9.7	6.46	7.2	5.7
50-54	15.6	15.7	15.4	10.47	11.3	9.7
55-59	22.1	23.7	20.6	17.56	18.1	17.0
60-64	25.0	28.5	22.0	27.81	28.4	27.3
65-69	29.0	29.2	28.9	35.38	40.3	31.5
70-74	34.4	33.6	34.9	27.72	31.3	25.3
75-79	39.7	38.6	40.4	36.97	37.6	36.6
80 +	44.1	42.8	44.6	46.78	46.5	46.9

Source: Estimates based on NSP 2002 and NSP 2011.
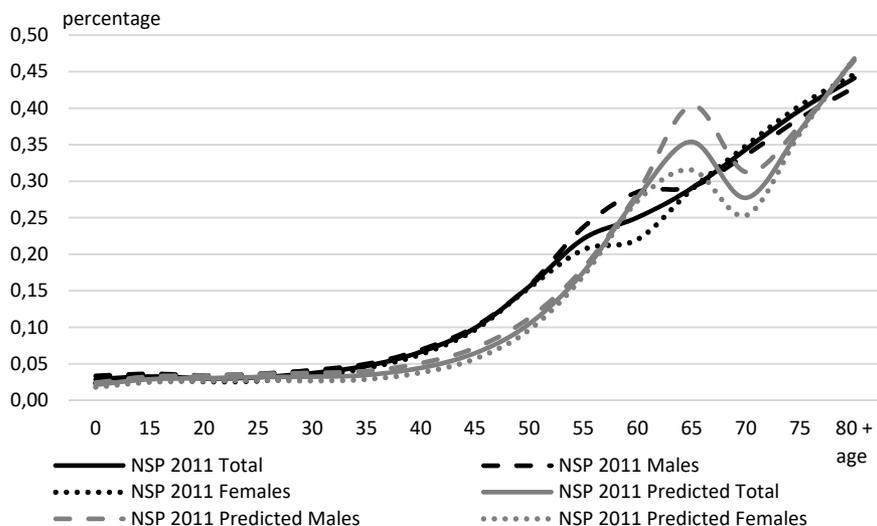
Figure 3. Disabled persons per 100 population by sex and age according to NSP 2011 and after aging the population of disabled people from NSP 2002

Source: Estimates based on NSP 2002 and NSP 2011.

The discrepancies described above inspired a further investigation seeking to determine how the estimates would change if one accounted for the distinction between disability in the legal and biological sense (Figures 4–6). The estimated shares of disabled persons in the 2011 census and those obtained after aging the disabled population from NSP 2002 into the future are roughly consistent with the pattern described above only in the case of the total population, without accounting for different categories of disability. The relationship is less evident in the case of disabled people in the biological and legal sense. One can clearly see the discrepancy resulting from the higher share of disabled persons in the retirement age in 2002, described above. This contrasts considerably with the relationship between estimates from NSP 2011 and the predicted values for the categories of exclusively legal and exclusively biological. It is this very difference that explains the lower estimates obtained in NSP 2011 compared to NSP 2002. This difference should be analysed separately for each of the two categories.
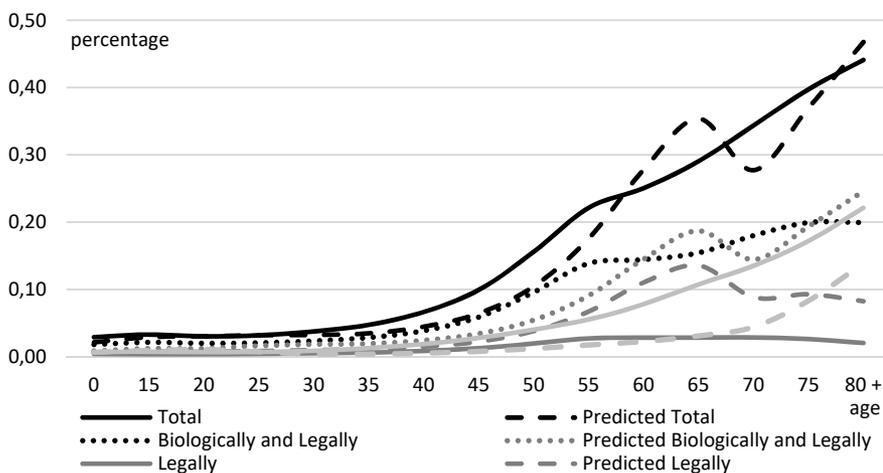
Figure 4. Disabled persons per 100 population by disability category and age according to NSP 2011 and after aging the population of disabled people from NSP 2002

Source: Estimates based on NSP 2002 and NSP 2011.
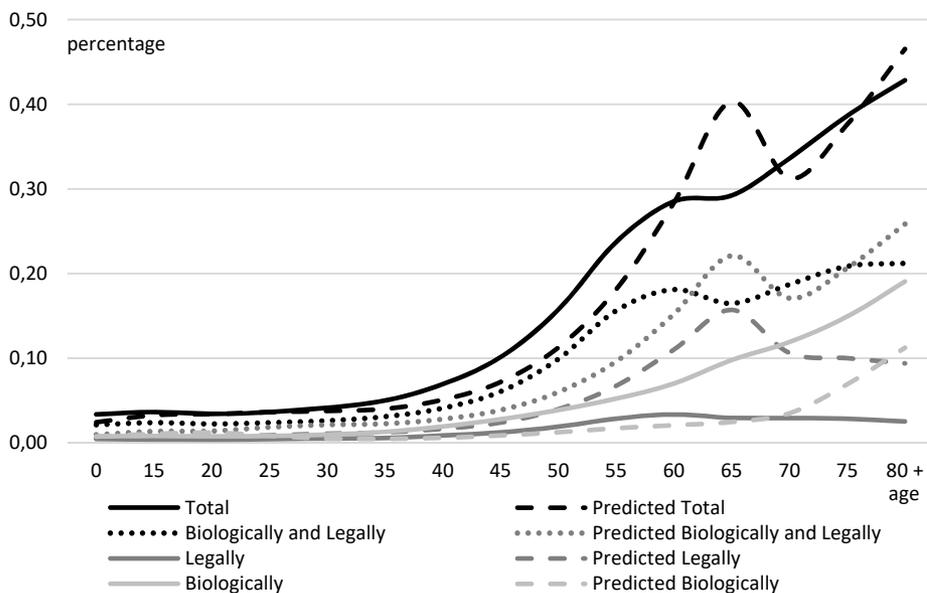
Figure 5. Disabled men per 100 population by disability category and age according to NSP 2011 and after aging the population of disabled people from NSP 2002

Source: Estimates based on NSP 2002 and NSP 2011.
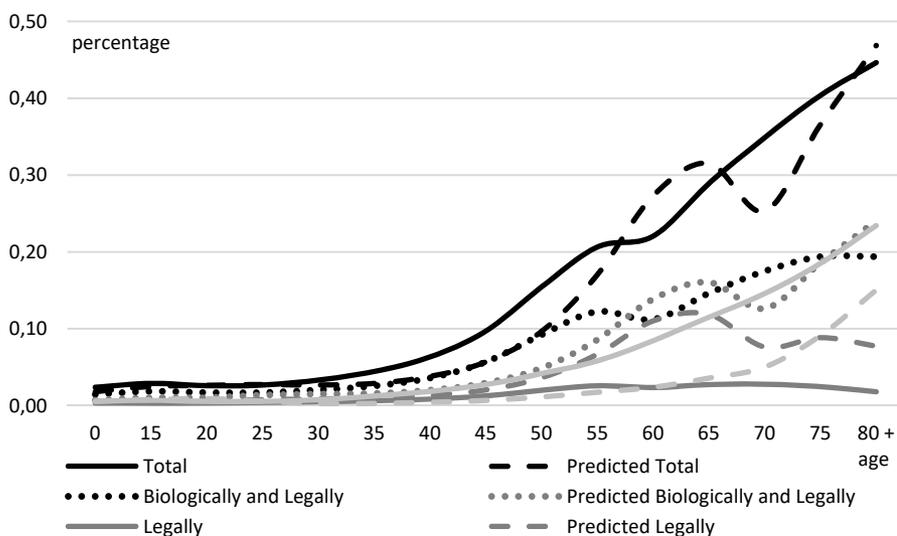
Figure 6. Disabled women per 100 population by disability category and age according to NSP 2011
and after aging the population of disabled people from NSP 2002

Source: Estimates based on NSP 2002 and NSP 2011.

The demographic aging was applied separately for different categories of disabled persons from NSP 2002 in order to track changes in these subpopulations after 9 years. The subpopulation of disabled persons exclusively in the legal sense was identified in the census based on subjective assessments of persons who had valid disability statements but did not experience any limitations in performing activities typical for their age. This category does raise certain doubts as to the grounds on which an authorized administrative body issued a disability statement about someone who did not have biological disabilities. The number of such persons, according to NSP 2011, is estimated to be around 480 thousand. In the previous census, this group amounted to over 1.6 million. In other words, the share of disabled persons exclusively in the legal sense declined from 4.25% in 2002 to 1.24% in 2011.

For men and women the estimates of disabled persons exclusively in the legal sense in NPS 2011 are lower than those obtained for the same year by prediction based on data from NSP 2002. There is a slight but noticeable increase in the share of the exclusively legal category of disability for people aged 50 and older, which rises to 2%. In the group of people aged 60, the share is equal to 3% and remains at a stable level for the following age groups. This situation can be explained by the desire to obtain a disability benefit associated with a disability statement as a supplement to the pension.

As regards the prediction for 2011 based on the subpopulation of disabled persons exclusively in the legal sense from NSP 2002, the estimated share is clearly higher and

equals 3.5% of the total population. It is twice as high as the share in NSP 2011. Starting from the age of 60, we can observe a threefold increase, which reaches its maximum (fivefold rise) for the age group 65–69. There is a similar pattern for women, although the differences are somewhat smaller. The biggest difference can be observed for the age group 60–64, where the share of disabled persons exclusively in the legal sense is over four times as high as that estimate.

The predicted values cannot be treated as precise estimates but they do show a trend reflecting inappropriate practices. The relationship between the share of this category of disability based on NSP 2011 and the predicted share for the same year based on data from NSP 2002 seems to be the result of efforts to counteract abuses concerning disability assessment decisions. This means that disability statements issued earlier may have become invalid, but also that in the following decade it was particularly difficult to obtain a positive decision and the number of disability statements actually issued was smaller. However, with respect to the exclusively biological category, the relationship between the share estimated in 2011 and the predicted share is exactly reverse. Given the clearly bigger scale of disability exclusively in the biological sense, a different conclusion can be drawn. From a social point of view, is it appropriate that such a high percentage of disabled people are classified as 'exclusively biological', which can, if fact, mean that these people are not able to successfully apply for a disability benefit.

Shares of disabled people exclusively in the biological sense in 2011 are clearly higher than the values obtained for 2011 by applying demographic prediction to data from 2002. There are two possible explanations for the discrepancy. The first one is the possibility that the degree of limitations in the performance of activities typical for a given age increases over the decade. The second possibility is that the difference is due to the declining share of disabled people classified as "exclusively legal" as a result of stricter disability assessment procedures.

The share of disabled people classified as "exclusively legal" in 2011 was 4.1% for the total population, 3.3% for men and 4.8% for women. The corresponding values obtained through prediction are 1.8%, 1.5% and 2.1%, respectively. Starting from the age of 50, there is a steady increase in the share of disabled people in this category from 4% to over 22% for people aged 80 and older. In contrast, with respect to the predicted shares (based on data from 2002), the onset of the intensive increase in the incidence of biological disability is delayed by 15–20 years. Consequently, at the age of 70 the share of people with biological disability is 3.5% for men and 5% for women. For people aged 80 the share of biological disability equals 14% (11% for men and 15% for men, Figures 3–5).

The above analysis contains a comparison of the relationship between the share of people with biological disabilities according to 2011 and 2002 censuses. The analysis took into account all people who reported the fact of experiencing activity limitations,

regardless of whether or not they had official statements of disability (Figure 6). It turns out that the share of biologically disabled estimated in 2011 was about 9% higher than in 2002. Thus, earlier suggestions about the possible underestimation of disability in 2011 compared to 2002 are not confirmed. Moreover, the variation across age groups shows an evident pattern: for people up to the age of 40, the share of disability estimated in 2011 is higher than that indicated in 2002. The difference amounts to as much as 30% for children up to the age of 15. For people aged 25–30, according to 2011, the share of disability for men is 16% higher than in the previous census, while for women it equals 22%. For people aged 40 the shares of disabled persons in both censuses are equal (ratio = 1, Figure 7). The biggest difference can be observed for people aged 50 – the share of biological disability among men in 2002 is 15% higher than in 2011 and 12% for women. For older age groups the discrepancy between the two censuses decreases and even disappears completely at one point.
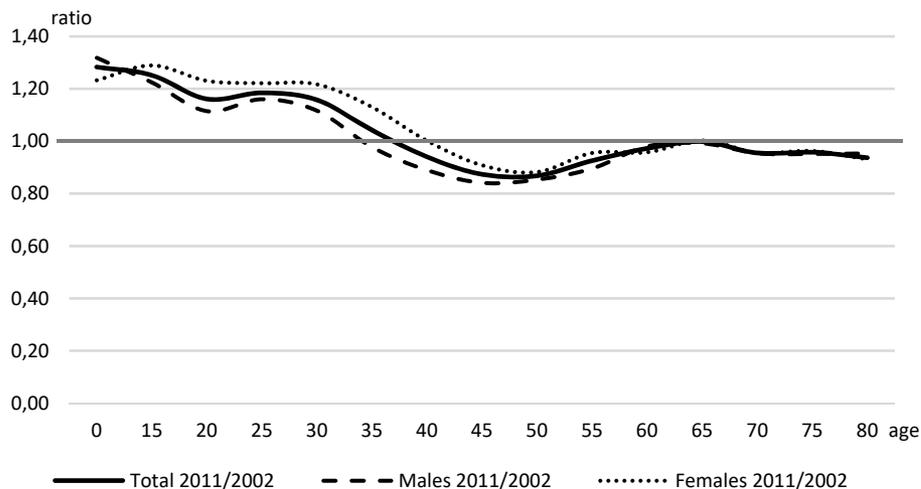
Figure 7. The relationship between the number of disabled persons in the biological sense by sex and age, NSP 2011/NSP 2002

Sources: Estimates based on NSP 2002 and NSP 2011.

The above considerations do not indicate that disability estimates in NSP 2011 are 'true'. However, the results of the comparison with shares predicted on the basis of data from NSP 2002 indicate that estimates obtained in both censuses are compatible. The analysis confirmed a similar variation in the share of disability across age groups. A higher share of disability was observed in younger age groups, especially among children. The situation calls for additional reflection on the method of measuring disability in this age group, which is also a concern pointed out by statisticians from the Washington Group (Crialesi et al., 2016; Meltzer, 2016). The analysis also helped to

identify one of the causes for the lower estimates of disability in NSP 2011, namely the change of rules in disability assessment procedures. The obvious consequence of the lower number of disability statements is the lower estimated share of disabled people classified as "exclusively legal".

6. Conclusions

The results of the analysis have highlighted a few aspects that should be taken into account when assessing the quality of estimates from the NSP 2011. For one thing, one should mention the consequences of the methodology adopted in the 2011 census. The voluntary nature of the question about such an essential topic as disability has two effects. First of all, it means that at the planning stage it was considered sufficient to estimate the number of disabled people in the legal sense on the basis of information from administrative registers. It is well known that regulations used for purposes of disability assessment vary across countries. For this reason, the comparability of results was supposed to be ensured by data from the survey. Following the example of other European countries, the recommendations of the Washington Group were not implemented in the census. However, leaving it up to respondents to decide whether or not to answer the disability question resulted in a high rate of non-response, which was the main cause of the bias in the results. Particularly, when one realises that the group of 1.3 million respondents who refused to answer the question most likely included disabled persons. To be fair, it was possible to link information from the sample survey in the 2011 census with data from the short questionnaire or from administrative registers, also for those who refused to answer the disability question. This additional information was then used to counteract the effect of non-response, inter alia, by means of calibration (Szymkowiak, 2012, 2014).

The lower estimates of disability in NSP 2011 were mainly due to decline in the subpopulation of disabled persons in the legal sense. The introduction of stricter criteria of disability assessment discouraged many potential applicants from applying for a disability benefit, which led to a fall in the actual size of this subpopulation and its estimates. This was accompanied by a rise in disability due to a higher incidence of diseases and injuries, which was confirmed by higher estimates of disability in the biological sense.

One can see an evident effect of new regulations used in the system of disability assessment and on decisions made by assessors. The natural consequence of this change is the decline in the number of disabled persons in the legal sense. Without passing judgement on how appropriate these administrative solutions actually were, there is no doubt that their effects were confirmed by statistical data. This fact should be viewed as evidence of the reliability of measuring the subpopulation of people with disability

statements. There is a separate question of using an appropriate definition of a disabled person, which focuses on people's limitations in the performance of basic activities for a given age and their participation in social life.

Acknowledgements

References

Altman, B. M., (2016). International *Measurement of Disability* (B. M. Altman, ed.). https://doi.org/10.1007/978-3-319-28498-9.

Altman, B. M., Meltzer, H., (2016). Developing Tools to Identify Environmental Factors as Context for Disability: A Theoretical Perspective. In *International Measurment of Disability*, pp. 183–206, https://doi.org/10.1007/978-3-319-28498-9_12.

Antczak, R., Grabowska, I., & Polańska, Z., (2018). Podstawy i źródła danych statystyki osób niepełnosprawnych. *Wiadomości Statystyczne*, (2), pp. 21–43, Retrieved from http://stat.gov.pl/czasopisma/wiadomosci-statystyczne/archiwum/wiadomosci-statystyczne-nr-22018-681,7,38.html.

Bogaert, P., Van Oyen, H., Beluche, I., Cambois, E., & Robine, J. M., (2018). The use of the global activity limitation Indicator and healthy life years by member states and the European Commission. *Archives of Public Health*, 76(1), pp. 1–7, https://doi.org/10.1186/s13690-018-0279-z.

Cambois, E., Jagger, C., Nusselder, W., Van Oyen, H., & Robine, J.-M., (2016). International Comparisons of Disability Prevalence Estimates: Impact of Accounting or Not Accounting for the Institutionalized Population. In *International Measurement of Disability*, pp. 207–229 https://doi.org/10.1007/978-3-319-28498-9_13.

Crialesi, R., De Palma, E., & Battisti, A., (2016). Building a "Module on Child Functioning and Disability." In *International Measurement of Disability,* pp. 151–165, https://doi.org/10.1007/978-3-319-28498-9_10.

Dehnel, G., Klimanek, T., (2016). Disa,bility in the national censuses of 2002 and 2011– a comparison of information scope. *Acta Universitatis Lodziensis Folia Oeconomica*, 5(325), pp. 127–141, https://doi.org/10.18778/0208-6018.325.10.

Dz.U. 2016 poz. 2046, (2016). Obwieszczenie Marszałka Sejmu RP z dnia 14 listopada 2016 r. w sprawie ogłoszenia jednolitego tekstu ustawy o rehabilitacji zawodowej i społecznej oraz zatrudnianiu osób niepełnosprawnych. *Dziennik Ustaw*, poz. 2046, Retrieved from http://prawo.sejm.gov.pl/isap.nsf/download.xsp/WDU20160002046/T/D20162046L.pdf.

Dz.U. 2017 poz. 1383, (2017). Obwieszczenie Marszałka Sejmu Rzeczypospolitej Polskiej z dnia 7 lipca 2017 r. w sprawie ogłoszenia jednolitego tekstu ustawy o emeryturach i rentach z Funduszu Ubezpieczeń Społecznych, *Dziennik Ustaw*.

Dz.U. Nr 47 poz. 277, (2010). Ustawa z dnia 4 marca 2010 r. o narodowym spisie powszechnym ludności i mieszkań w 2011 r., *Dziennik Ustaw*.

Ehlis, (2015). European Health and Life Expectancy Information System, Retrieved from https://www.bridge-health.eu/sites/default/files/BRIDGE Health EHLEIS_SHORT.pdf.

Fihel, A., (2011). Płeć a trwanie życia : analiza demograficzna. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego : Uniwersytet Warszawski. Wydział Nauk Ekonomicznych.

GUS, (2011a). Kwestionariusz NSP 2011. Badanie reprezentacyjne. Ludność, Retrieved from ttps://stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultstronaopisowa/5781/1/1/nsp_2011_badanie_reprez_ludnosc_wykaz_pytan.pdf.

GUS, (2011b). Stan zdrowia ludnosci Polski w 2009 r. (N. Koehne, B. Lednicki, M. Piekarzewska, R. Wieczorkowski, & A. Zajenkowska-Kozłowska, eds.), Retrieved from https://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie/stan-zdrowia-ludnosci-polski-w-2009-r,6,5.html.

GUS, (2012). Osoby niepełnosprawne na rynku pracy w 2011 r., Retrieved from https://stat.gov.pl/obszary-tematyczne/rynek-pracy/pracujacy-bezrobotni-bierni-zawodowo-wg-bael/osoby-niepelnosprawne-na-rynku-pracy-w-2011-r-,24,1.html.

GUS, (2013a). Ludność, stan i struktura demograficzno-społeczna, Narodowy Spis Powszechny Ludności i Mieszkań 2011, Warszawa.

GUS, (2013b). Ludność i gospodarstwa domowe. Stan i struktura społeczno-ekonomiczna. Część I. Ludność – NSP 2011, Retrieved from https://stat.gov.pl/spisy-powszechne/nsp-2011/nsp-2011-wyniki/ludnosc-i-gospodarstwa-domowe-stan-i-struktura-spoleczno-ekonomiczna-czesc-i-ludnosc-nsp-2011,11,1.html.

GUS, (2016). Stan zdrowia Ludnosci Polski w 2014 r. (M. Piekarzewska, R. Wieczorkowski, & A. Zajenkowska-Kozłowska, eds.), Retrieved from https://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie/stan-zdrowia-ludnosci-polski-w-2014-r-,6,6.html?pdf=1#%5B%7B%22num%22%3A1%2C%22gen%22%3A0%7D%2C%7B%22name%22%3A%22XYZ%22%7D%2C0%2C841.89%2Cnull%5D.

Loeb, M., (2016a). Development of Disability Measures for Surveys: The Washington Group Extended Set on Functioning. In *International Measurement of Disability*, pp. 97–122, https://doi.org/10.1007/978-3-319-28498-9_7.

Loeb, M., (2016b). *International Census/Survey Data and the Short Set of Disability Questions Developed by the Washington Group on Disability Statistics*, https://doi.org/10.1007/978-3-319-28498-9_15.

Madans, J., Mont, D., & Loeb, M., (2015). Comments on Sabariego et al. Measuring Disability: Comparing the Impact of Two Data Collection Approaches on Disability Rates. Int. J. Environ. Res. Public Health, 2015, 12, pp. 10329–10351. *International Journal of Environmental Research and Public Health*, 13(1), p. 65, https://doi.org/10.3390/ijerph13010065

Meltzer, H., (2016). The Challenges of Conducting National Surveys of Disability Among Children. In *International Measurement of Disability*, pp. 137–149, https://doi.org/10.1007/978-3-319-28498-9_9.

Molden, T. H., Tøssebro, J., (2010). Measuring disability in survey research: Comparing current measurements within one data set. *Alter*, 4(3), 174–189, https://doi.org/10.1016/j.alter.2010.05.003

Mont, D. (2007). Measuring Disability Prevalence. *Social Protection, World Bank*, (0706), pp. 1–54.

Sabariego, C., Oberhauser, C., Posarac, A., Bickenbach, J., Kostanjsek, N., Chatterji, S.,Cieza, A., (2015). Measuring disability: Comparing the impact of two data collection approaches on disability rates. *International Journal of Environmental Research and Public Health*, 12(9), pp. 10329–10351, https://doi.org/10.3390/ijerph120910329.

Slany, K., (2014). Osoby niepełnosprawne w świetle Narodowego Spisu Powszechnego Ludności i Mieszkań z 2011 r. – wybrane aspekty. *Niepełnosprawność*, (11), pp. 44–62, Retrieved from http://www.pfron.org.pl/download/5/503/04-KrystynaSlany.pdf.

Sullivan, D. F., (1971). A Single Index of Mortality and Morbidity. *HSMHA Health Reports*, 86(4), pp. 347–354. https://doi.org/10.2307/4594169.

Szukalski, P., (2018). Nadumieralność mężczyzn w Polsce w układzie regionalnym. *Demografia i Gerontologia Społeczna – Biuletyn Informacyjny*, (8).

UN, (2008a). *Principles and Recommendations for Population and Housing Censuses Revision 2*, New York: United Nations.

UN, (2008b). *The measurement of Disability Recommendations for the 2010 Round of Censuses.*

UN, (2015a). Principles and recommendations for population and housing censuses - Revision 3. In *Statistical Papers Series M.No.67/Rev.3*. New York: United Nations.

UN, (2015b). *Przekształcamy nasz świat: Agenda na rzecz zrównoważonego rozwoju 2030*, Retrieved from http://www.unic.un.org.pl/files/164/Agenda 2030_pl_2016_ostateczna.pdf.

UN WG, (2006). *The Washington Group Short Set of Questions on Disability*. Retrieved from http://www.washingtongroup-disability.com/wp-content/uploads/2016/01/The-Washington-Group-Short-Set-of-Questions-on-Disability.pdf.

US CENSUS BUREAU, (2017). *Measuring Disability in a Census.*

Van Oyen, H., Bogaert, P., Yokota, R. T. C., & Berger, N., (2018). Measuring disability: A systematic review of the validity and reliability of the Global Activity Limitations Indicator (GALI). *Archives of Public Health*, 76(1), pp. 1–11, https://doi.org/10.1186/s13690-018-0270-8.

WBG, (2018). World Bank Group Announces New Commitments on Disability Inclusion.

WBG, (2018).World Bank Group Commitments on Disability-Inclusive Development.

WHO, (2002). Towards a Common Language for Functioning, Disability and Health *ICF*, Retrieved from https://www.who.int/classifications/icf/icfbeginnersguide.pdf.

WHO, & The World Bank, (2011). World report on disability 2011, 91, Retrieved from http://www.who.int/disabilities/world_report/2011/report.pdf.

Interviewer allocation through interview–reinterview nested design for response error estimation in sample surveys

Fidan Mahmut Fahmi[1], **H. Öztaş Ayhan**[2], **İnci Batmaz**[3]

ABSTRACT

In surveys, non-sampling errors, due to their complex nature, are more challenging to quantify compared to sampling errors. Avoiding the release of these errors, however, results in biased survey estimates. In our previous paper, we devised the best interviewer allocation technique by using a nested experimental design to study response error estimation. In this study, in order to illustrate the effectiveness of this methodology in a different context, we apply it in interview-reinterview surveys relating to the time use and life satisfaction of academicians at Middle East Technical University, Turkey. An analysis of the pilot survey data showed that only half of the data was reliable, while the other half revealed interviewer effects. Prior to the main survey, interviewers underwent training in the course of which particular emphasis was put on the above-mentioned questions. In effect, the previously observed response variances which accounted for the total variance and data unreliability, were reduced considerably, increasing the quality of the main survey.

Key words: correlated response error, interviewer allocation assignments, quality of survey research, reinterview procedure, sample survey design.

1. Introduction

Research in the survey area mostly concerns the errors involved during the survey (Biemer *et al.* 2004); while some are dealing with the ways of eliminating errors, others try to measure the effect of them on the results by estimating the components of Total Survey Errors (TSE) involving both sampling and nonsampling errors (Kalton 1983, Salant and Dillman 1994). There are many different types of nonsampling errors. McNabb (2014) has defined nonsampling errors covering; frame error, measurement error, response error, interviewer error, and nonresponse error. McNabb (2014) also defines response error as basically respondent error. Measurement errors occur when

---

[1] Graduate student of METU, Ankara, Turkey. E-mail: fidantelaferli@yahoo.com.

[2] Professor Emeritus, Department of Statistics, METU, Ankara, Turkey. Corresponding author.
E-mail: oayhan@metu.edu.tr. ORCID: https://orcid.org/0000--0003-3818-483X.

[3] Professor, Department of Statistics, METU, Ankara, Turkey. E-mail: ibatmaz@metu.edu.tr.
ORCID: https://orcid.org/0000-0002-0821-3786.

the observed value differs from the true value according to the definition of the variable (Biemer and Lyberg, 2003). In several studies, response errors and measurement errors are used interchangeably (see; Hansen *et al.* 1951 and 1961).

The interviewer error, which may cause correlated errors in surveys, is one of the nonsampling errors. Since there exist only a few methods for compensating correlated errors caused by nonsampling errors, the current practice is to prevent the occurrence of them in data analysis (Biemer 2012). Due to the complex nature of the nonsampling error components, they are not usually examined in every survey report causing considerable bias in the survey estimates. Naturally, this study aims to highlight the importance of releasing such results, which covers a higher portion of the TSE.

Nonsampling error research was originally initiated by the U.S. Census Bureau Methodology Section around the middle of the last century. An early study of the methodology of response errors in surveys was published by Hansen, Hurwitz, Marks and Mauldin (1951). U.S. Bureau of the Census' survey model was described in Hansen, Hurwitz and Bershad (1961) and also in Hansen, Hurwitz and Pritzker (1964). Estimating the response variance components of the U.S. Bureau of the Census' Survey model was also evaluated by Bailar and Dalenius (1969). Later studies from the former Bureau researchers were done by Biemer and Stokes (1985, 2004) on the modelling of measurement error.

On the European side, response error related research was initiated by World Fertility Survey (WFS) Methodology Team, who worked on the single round high quality data collection. By this team, the methodology of the response errors was evaluated by O'Muircheartaigh and Marckwardt (1980) for the assessment of the reliability of WFS data. Sampling issues for national fertility surveys were also covered by Verma (1980). Computation methodology of response errors was proposed by O'Muircheartaigh (1984) for Peru Fertility Survey. Response errors for attitudinal surveys was also formulated by O' Muircheartaigh (1976, 1977). Later, simple response variance estimation methodology was overviewed by O' Muircheartaigh (2004). An excellent overview of the response error methodology is also covered by Moser and Kalton (1979). Estimation in the presence of measurement error is also evaluated by Fuller (1995).

Reinterview methodology was developed for the first time by the U.S. Census Bureau Methodology Division during 1950's (Hansen et al. (1951, 1961). Since then, several other researchers have improved on the methodology (Bailar and Dalenius 1969, Biemer and Stokes (1985, 1991). Identical approaches have also been taken for the WFS Methodology Division by O'Muircheartaigh (1977, 1984). Using design of experiments (DOEs) in interviewer allocation is relatively new in the survey research. O'Muircheartaigh (1984) has applied the methodology of interviewer-reinterviewer allocation to Peru Fertility Survey using a DOE, called Latin Square Design.

Estimators of nonsampling errors in interview–reinterview supervised surveys with interpenetrated assignments were proposed by Bassi and Fabbris (1997). Especially, O'Muircheartaigh (1977, 1984), Bassi and Fabbris (1997), Ayhan (2003) and Fahmi (2013) have also highlighted the use of supervised interview-reinterview design in their methodological research. It has been widely used since then in a variety of surveys (Biemer *et al.* 2004). It is an important tool to evaluate field work and to estimate and reduce the error components in a survey.

Interview–reinterview designs are necessary to estimate response variance. Hox et al. (2004) stated that "Several authors have criticized the existing studies on interviewer effects (Hagenaars and Heinen, 1982). A central criticism concerns the adequacy of the statistical models used. The structure of the data to be analyzed is hierarchical, since respondents are nested within interviewers." Lyberg and Kasprzyk (2004) also stated that "Interviewer errors and interviewer variability can be measured in various ways. Basically, different systems for reinterviews (replication; McCarthy, 1966) and interpenetration (Mahalanobis, 1946) are used." Based on these, is it possible to say that the researchers here use a reinterview design to estimate interviewer effects, because such a design eliminates the nested structure where one respondent is interviewed by one interviewer only. The rationale for the current study is the nested design of respondents within interviewers.

In this article, one kind of nonsampling error, called response error, is investigated in a personal interview in sample surveys, where multiple stages of sampling are employed. Response errors may occur because of the respondent error, interviewer error, or their interactions causing correlated response error. Under this assumption, in this article, a nested experimental design (ND) is utilized for developing response error models and obtaining efficient estimators for response such as simple and correlated response variance in interview-reinterview surveys as suggested by Ayhan (2003, 2012).

Ayhan (2003) used ND to make interviewer allocation for the interview-reinterview process. Ayhan (2012) also mentioned that experimental settings of the interviewer allocation can be based on the Nested and Factorial Design (NFD), or the Split Plot Design (SPD). Then, Fahmi (2013) investigated these designs for interviewer allocation in personal interview surveys. Next, Batmaz and Fahmi (2015) established theoretical backgrounds of the simple and correlated response error estimation procedure.

In general, to make the analysis simple, the expected value of the interviewer effect is assumed to be zero although it is not in reality. To measure it, the survey is designed to have different interviewers for the respondents within the main and reinterview survey. The advantage of the ND allocation is that it provides different respondents for the same interviewer in both the pilot and main survey, enabling to compute the response variance independently for each survey. It also provides flexibility in the field

allocation and application. Note here that there are some important issues to be dealt with in designing reinterviews such as selection of sample, reinterviewer, respondent, mode of interview as well as designing of reinterview questionnaire (Biemer and Stokes 2004).

Purposes of reinterview were extensively covered by Forsman and Schreiner (2004) where they have proposed "purposes of interview–reinterview designs", which are classified as:

(1) *To evaluate fieldwork*: (a) reinterview is used to identify interviewers who are falsifying data, and (b) reinterview is also used for misunderstood procedures and require remedial training.
(2) *To estimate error components in a survey model*: (a) reinterview is also used to estimate simple response variance, and (b) reinterview is also used to estimate response bias.

In the current study, reinterview is used to estimate simple response variance (Design 2a). In addition, this study also estimates correlated response variance and interviewer variance.

Forsman and Schreiner (2004) further cover other reinterview design issues, evaluating interviewer performance, model-based analysis of reinterview data, and the use of computer assisted interviewing. Correlates of reinterview response inconsistency are also examined by O'Muircheartaigh (1986) in the Current Population Survey.

A very recently edited book by Olson *et al.* (2020 a) covers interviewer effects from a total survey error perspective. An overview of research on interviewer effects is covered by Olson *et al.* (2020 b) within the book. They state that the errors introduced by interviewers can take the form of bias or variance. Early research also found that interviewers vary in how they administer survey questions and their effects were similar to sample clusters in both face-to-face (Hansen *et al.* 1961, Kish 1962) and telephone surveys (Groves and Magilavy 1986). In particular, similar to a design effect for cluster samples, interviewers increase the variance of an estimated mean as a function of their average workload and the intra-interviewer correlation.

Given the nesting of respondents within interviewers, following Kish's ANOVA-based model (Kish 1962), hierarchical or random effects models have long been used for the study of interviewer effects (Dijkstra 1983, Hox 1994, O'Muircheartaigh and Campanelli 1998).

Recently, Edwards *et al.* (2020) studied behaviour change techniques for reducing interviewer contributions to total survey error. Modelling interviewer effects in the National Health Interview Survey is also investigated by Dahlhamer *et al.* (2020). On the other hand, West (2020) designed studies for comparing interviewer variance in two groups of survey interviewers.

In this study, the methodology previously developed by Ayhan (2003, 2012), Fahmi (2013) and Batmaz and Fahmi (2015) is applied to an interview-reinterview survey for inquiring about the time use and life satisfaction of academicians working at the Middle East Technical University (METU), Turkey. This study contributes to the literature

in various aspects. Keeping in mind that nonsampling errors are equally important as sampling errors, it guides the survey researchers to measure the response errors, particularly interviewer effects on the responses as well as to measure the total, sampling, response and the correlated interviewer variances, efficiently. Thus, an important concern regarding the estimation of nonsampling errors is overcome. However, nonsampling errors are neglected in most surveys, due to "hardness in quantifying" as well as "additional data coasting," and consequently, error based information on most of the TSE components cannot be obtained completely. In the case the researchers want to compute and report these errors along with the survey results, it definitely will increase the validity and reliability, and hence, quality of surveys. This way, usefulness and limitations of surveys conducted will be appreciated better. Moreover, these tools will provide feedbacks regarding errors involved, particularly in surveys conducted periodically. By evaluating the experiences gained, quality of the survey can be continuously improved.

This article is organized as follows: the response reliability measures are defined in Section 2. In Section 3, the methodology for interviewer allocation by ND and its application are presented. In Section 4, findings of the applications are presented and the work is concluded in Section 5.

2. Measures of response reliability

We know that for each individual covered by the survey, there is an individual true value. The difference between an individual true value and the value recorded on the schedule is the individual response error. There is always a possibility that true values may change. To determine the optimum period between interview and reinterview, we followed the guide suggested by Biemer and Stokes (2004), and it is mentioned clearly in Section 3.2 Pilot Survey Application. We hope only few such changes in true values left, and they are represented by the residual error in the model.

In investigating the reliability of data, we can focus on two different but related aspects of the data: bias and variance. For each individual j, we have for each variable y, the results of two separate observations, y_{j1} and y_{j2}. In this case, they are assumed to be obtained from an interview-reinterview survey, respectively. The differences within the pairs of observations provide the raw material for the reliability investigation. Measures of reliability used depend on the types of data.

Measures of response reliability and response error estimation are proposed by Hansen et al. (1961). The methodology was also extended by O'Muircheartaigh (1977, 1984) and O'Muircheartaigh and Marckwardt (1980). They have covered the methodology for several data measurement scales, which can be categorized as in the following sections.

2.1. For categorical data

In comparing the responses obtained for a particular variable, the data may be represented by the square matrix $\{n_{ij}\}$, where n_{ij} is the number of elements classified in category *i* according to the first interview, and in category *j* according to the second interview, i.e. reinterview. The diagonal of this square matrix, with entries, n_{ii}, contains the cases of exact agreement. The simplest measure of reliability (bivariate agreement) is the *index of crude agreement* (or *crude index*), which can be written as

$$A = \frac{1}{n}\sum_i n_{ii}. \tag{1}$$

It represents the proportion of correctly classified units. Another simpler one is the *index of crude disagreement*

$$D = 1 - A. \tag{2}$$

It represents the proportion of incorrectly classified units. Here, values of *A* and *D* close to one (1) and zero (0), respectively, indicate good agreement.

However, the *crude index A* in formulae (1) has a fairly serious drawback; it does not take into account the fact that some agreements will occur by chance even if the measurement is completely unreliable (random). To overcome the problem, Cohen (1960) define an *index of consistency*, called *kappa*, of the following form:

$$\text{K} = 1 - \frac{1 - Po}{1 - Pe} = \frac{Po - Pe}{1 - Pe}, \tag{3}$$

where $P_o = \sum_{i=1}^{L}\left(\frac{n_{ii}}{n}\right)$ and $P_e = \sum_{i=1}^{L}\left(\frac{n_{i.}}{n}\right)\left(\frac{n_{.i}}{n}\right)$. Here, P_o is the sum of the observed proportions reflecting agreement, and P_e is the sum of the expected proportions reflecting agreement. Under the assumption of independence between the two observations, formulae (3) can be written as

$$\widehat{K} = 1 - \frac{\sum_{i \neq j} n_{ij}}{\frac{1}{n}\sum_{i \neq j} n_{i.}n_{.j}} = \frac{\sum_{i=1}^{L}\left(\frac{n_{ii}}{n}\right) - \left(\frac{n_{i.}}{n}\right)\left(\frac{n_{.i}}{n}\right)}{1 - \sum_{i=1}^{L}\left(\frac{n_{i.}}{n}\right)\left(\frac{n_{.i}}{n}\right)}, \tag{4}$$

where *L* represents the number of categories. For evaluating the magnitude of *kappa*, *K*, the standards proposed by Landis and Koch (1977) are utilized. Note also that *CI's* non containing 0 (zero) indicate significant *kappa* values.

2.2. For ordinal data

When the scales are ordinal, interval or ratio, any measure of agreement should take into account the degree of disagreement, which is a function of the difference between scale values. Accordingly, formula (1) is modified by redefining *agreement* to mean that the two interviews obtain values within some acceptable distance (*k* units) of each other. Then, agreement can be written as

$$A_k = \sum_{|i-j| \leq k} n_{ij} = 1 - D_k. \tag{5}$$

Cohen (1968) introduce a modified form of K, which allows for the scaled disagreement or partial credit in terms of weights $\{W_{ij}\}$, which reflect the contribution of each cell in the table to the degree of disagreement.

$$K_W = \frac{Po^* - Pe^*}{1 - Pe^*} = 1 - \frac{\sum_{i \neq j} w_{ij} n_{ij}}{\frac{1}{n} \sum_{i \neq j} w_{ij} n_{i.} n_{.j}}. \tag{6}$$

Here, any monotonically decreasing function of the differences between the values i and j can be used as weights. Cicchetti (1972) suggests the use of the following weights for the ordinal and metric data, respectively.

$$w_{ij} = 1 - |i - j|/(L - 1); \quad w_{ij} = 1 - (i - j)^2. \tag{7}$$

2.3. For interval and ratio scale data

For metric measurements, Hansen *et al.* (1961) proposed the basic mathematical or response error given below. The same methodology was reformulated by O'Muircheartaigh (1977) and O'Muircheartaigh and Marckwardt (1980). For simplicity, the discussion is restricted to the estimation of the population mean, where

$$\mu = \frac{1}{N} \sum_{j=1}^{N} \mu_j. \tag{8}$$

Assume that an observation for the jth element in the survey for trial t is denoted by y_{jt}. An estimator of μ obtained from a survey (one trial) is

$$\bar{y}_t = \frac{1}{n} \sum_{j=1}^{n} y_{jt}. \tag{9}$$

Here, the population consists of N individuals from which a sample of size n is sampled.

The total variance of the survey estimator is

$$\sigma_t^2 = E(\bar{y}_t - \bar{Y})^2, \tag{10}$$

where $\bar{Y} = E(\bar{y}_t)$, and the mean square error (MSE) of the estimator becomes

$$MSE = E(\bar{y}_t - \mu)^2 = \sigma_t^2 + \beta^2. \tag{11}$$

The expected value over all possible trials for the element j is

$$E(y_{jt}|j) = Y_j. \tag{12}$$

The difference between the observation on the jth unit of a particular survey (say trial t) and the expected value is

$$d_{jt} = y_{jt} - Y_j \tag{13}$$

2.3.1. Simple response variance

This is the *response deviation*, which is measured from the expected value. For the estimator obtained from the survey, the total variance can be partitioned as follows:

$$\sigma_t^2 = E(\bar{y}_t - \bar{Y})^2$$
$$= E(\bar{y}_t - \bar{y})^2 + 2E(\bar{y}_t - \bar{y})(\bar{y} - \bar{Y}) + E(\bar{y} - \bar{Y})^2, \tag{14}$$

where $\bar{y}_t = \frac{1}{n}\sum_{j=1}^{n} y_{jt}$ and $\bar{y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$.

The first term in equation (14) is the response variance, $\sigma_{\bar{d}_t}^2$; the second term involves the covariance between $\bar{d}_t$ and $\bar{y}$, and the third term is the sampling variance, $\sigma_{\bar{y}}^2$. The response variance can be restated as

$$\sigma_{\bar{d}_t}^2 = E(\bar{d}_t^2) = \frac{\sigma_d^2}{n}[1 + (n-1)\rho], \tag{15}$$

where σ_d^2 is simple response variance and ρ is the interclass correlation coefficient among the response deviations within a trial. Fellegi (1964), permits in principle the estimation of a number of components of the correlated response variance, $(\sigma_d^2(n-1)\rho)/n$. The sample variance can be written as

$$\sigma_{\bar{y}}^2 = E(\bar{y} - \bar{Y})^2 = \sigma_s^2[1 + \rho(n-1)], \tag{16}$$

where σ_s^2 is the population variance, and ρ is the intracluster correlation coefficient.

Then, index of inconsistency, IOI, is defined to be

$$IOI = \frac{\sigma_d^2}{\sigma_s^2 + \sigma_d^2}, \tag{17}$$

which measures the proportion of the total element variance due to the response variability. To measure how much the data are reliable, another statistic, called *reliability of data*,

$$r = 1 - IOI \tag{18}$$

has been proposed by Yu et al. (2000). Note that values of *IOI* and *r* close to zero (0) and one (1), respectively, indicate that data are consistent and reliable.

2.3.2. Correlated response variance

The analysis of response deviations presented above treats them as uncorrelated. The basic model of the response process for individual *j* is

$$y_{jt} = \mu_j + \beta_j + d_{jt}, \tag{19}$$

where y_{jt} is the response obtained from the individual *j* on the occasion *t*; μ_j is the true value for the individual; β_j is the individual response bias and d_{jt} is the response deviation. Since we cannot, unless we have external validating information for the individual, estimate β_j, we rewrite equation (13) as

$$y_{jt} = Y_j + d_{jt}, \tag{20}$$

where Y_j is the expected value of the observation for individual *j* over a large number of trials under the same essential survey conditions. However, if the interviewers cause a systematic distortion of the responses, we can write

$$y_{ijt} = Y_j + \alpha_i + \varepsilon_{ijt},$$

where the subscript *i* is added to denote the interviewer and split the response deviation d_{jt} into two additive components α_i and ε_{ijt}. Here, the α_i represents the net systematic

effect of interviewer i on the responses; it is the net bias introduced by the interviewer i. The ε_{ijt} is the residual response deviation, which is assumed to be unrelated to the interviewer. For making the relations simpler among the interviewers as a whole, it is assumed that the expected value of the interviewer error is taken as zero $[E(\alpha_i) = 0]$. This assumption naturally makes life easy, instead of computing interviewer error. Note also that that the interviewer effect can also be measured by the correlation between the responses for the first and the second interview as follows:

$$\alpha = \text{corr}(y_{j1}, y_{j2}). \tag{21}$$

Values of α close to 1 (one) indicate no interviewer effect at all.

Making the usual assumptions about the variances and covariances, we can write the variance of a single observation y_{ijt} as

$$\text{Var}\,(y_{ijt}) = \sigma_S^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2. \tag{22}$$

If $\bar{y}_t$ is the sample mean for the survey, then its variance is

$$\text{Var}\,(\bar{y}_t) = \frac{\sigma_S^2}{n} + \frac{\sigma_\alpha^2}{k} + \frac{\sigma_\varepsilon^2}{n}, \tag{23}$$

where k is the number of interviewers. And it is

$$= \frac{\sigma_S^2}{n} + \frac{\sigma_d^2}{n}(1 + \rho(m-1)), \tag{24}$$

where σ_S^2 is the population variance of the $\{y_j\}$; $\sigma_d^2 (= \sigma_\alpha^2 + \sigma_\varepsilon^2)$ is the simple response variance; m is the average workload per interviewer; and ρ is the intra-interviewer correlation coefficient $(= \sigma_\alpha^2/\sigma_d^2)$.

The usual estimate of the survey variance will include both σ_S^2 and σ_d^2, and if the sample is a simple random sample, it will be $(\sigma_S^2 + \sigma_d^2)/n$. Thus, the survey variance will be underestimated by an amount equal to

$$\frac{\sigma_d^2}{n}\rho(m-1). \tag{25}$$

2.3.3. Simple response variance, correlated response variance, and interviewer error

Computations of response variance and correlated interviewer variance are based on the following estimators. Let us denote μ and σ_t^2, the mean and the variance of y, respectively. Then, the total variance is the combination of sampling and response variances.

$$\mu = \frac{\sum_i^a \sum_j^b \mu_{ij}}{ab}, \tag{26}$$

$$\sigma_t^2 = E\{\frac{1}{n}\sum_i^a \sum_j^b (y_{ij} - \mu)^2 = \sigma_s^2 + \sigma_r^2, \tag{27}$$

where σ_s^2 is the sampling variance and σ_r^2 is the response variance. The sampling variance is defined (Fellegi 1964; Bassi and Fabbris 1997) as

$$\sigma_s^2 = E\left(\frac{1}{ab}\sum_i^a \sum_j^b c_{ij}^2\right), \tag{28}$$

where $c_{ij} = \mu_{ij} - \mu_i$ is the sample deviation of the same unit. Also, the response variance is defined as

$$\sigma_r^2 = E\left(\frac{1}{ab}\sum_i^a \sum_j^b d_{ij}^2\right), \tag{29}$$

where d_{ij} is the *response deviation* of unit j, enumerated by interviewer i. Here, the response deviation is determined as $d_{ij} = \tau_i + \beta_{j(i)}$. Values of d close to zero (0) and/or *CI for d* containing zero (0), indicate no significant deviation of the response between two interviews. Besides, δ_2 is the correlation coefficient between the response deviations within interviewer's assignments, and it is defined as

$$\delta_2 = \frac{1}{\sigma_r^2} E\left\{\frac{1}{n(b-1)} \sum_i^a \sum_{j=j'}^b d_{ij} d_{ij'}\right\}, \tag{30}$$

and $\delta_2 \sigma_r^2$ is the correlated interviewer variance.

The response error analysis is conducted and response reliability measures are calculated for each question listed in Table 2 by using the formulas shown in Section 2. Particularly, the response reliability statistics, A, D, A_k, IOI, r, K, K_w are calculated by using the formulae (1), (2), (5), (17), (18), (4), and (6), respectively. In addition, the interviewer effect, total, sampling, response and correlated interviewer variances are obtained by using the formulae (21), (27), (28), (29), and (29) and (30), respectively. Note here that to calculate the response reliability statistics for the Likert scale type questions, the agreement proportions between two interviews are obtain first (see Fahmi, 2013).

3. Application of the methodology

The main purpose of this study is to provide an insightful application whose results shed new light on the success of the methodology already developed by Ayhan (2003 & 2012), Fahmi (2013), and Batmaz and Fahmi (2015). The use of DOE technique for allocating interviewers provides a novel approach for estimating response error variance. In order to apply the methodology, an interview-reinterview survey is designed and conducted at Middle East Technical University (METU), Ankara, Turkey, inquiring about time use and life satisfaction of its academicians.

3.1. Interviewer allocation by a nested design

The experiments involved two or more random factors and the levels of at least one factor are similar but not identical for different levels of another factor is generally designed as nested experiments, and are commonly used to determine the sources of variation in the system (Box *et al.* 2005, Montgomery 2012). To illustrate this, suppose that the levels of a factor (e.g. B) are similar but not identical for the levels of another factor (e.g. A). Such an arrangement is called an ND with the levels of factor B nested

under the levels of factor A. A linear statistical model for analyzing such an experiment, a two-stage ND, is written as

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k} \begin{cases} i = 1,2, \dots, a \\ j = 1,2, \dots, b. \\ k = 1,2, \dots, n \end{cases} \quad (31)$$

There are a levels of factor A, b levels of factor B nested under each level of A. The subscript $j(i)$ indicates that the jth level of factor B is nested under the ith level of factor A. The replicates, if exist, are assumed to be nested within the combination of levels of A and B; so that the subscript $(ij)k$ is used for the white noise error term. In addition, this is a balanced ND because there are an equal number of levels of B within each level of A and an equal number of replicates. Because not every level of factor B appears with every level of factor A, there can be no interaction between A and B. In our particular case, this implies that respondents in different domains can only be visited by different interviewers, hence, data collected by ND can only be analyzed under the assumption that there is no interaction between interviewer and respondent factors. To measure this interaction, factorial designs can be used. However, in such allocations, the number of interviewers to be allocated for the field application may be combinatorically problematic.

Although ND does not allow measuring the interaction between the interviewer and respondent factors, it provides flexibility in allocating interviewers to respondents. Therefore, when compared with the factorial design, ND is much more time and cost efficient. Moreover, due to the fact that the factors involved are assumed to be random here, ND naturally provides estimates for the variance components, which are sample, interviewer and response variance in this case.

3.2. Pilot survey application

As the first step, a pilot survey is applied to a METU department. The main purpose of conducting pilot surveys is diverse; it includes pretesting the questionnaires, estimating the duration of interview, and planning the timing of reinterviews. In addition, data obtained from pilot studies (Fahmi 2013) are also analyzed to get feedback on the applicability of the methodology considered. Here, we present the pilot survey, in which interviewer allocation is done by using ND in a 10 question life satisfaction survey for the academicians in two rounds (i.e. pilot parent survey interview and reinterview).

Twenty-two academicians are involved as respondents in this survey; 10 of them are faculty members and 12 of them are research assistants. They are randomly clustered into four domains, where two contain five and other two contain six respondents. The interviewers selected randomly from the graduate students of the department are randomly assigned to one of these domains. As a result, an unbalanced

ND design is formed. The nested layout of the pilot fieldwork interview is shown in Figure 1. Also, the fieldwork allocation of the interviewers to respondent groups for the pilot survey is given in Table 1. Here, all domains are assigned to one supervisor, who controls the completion errors within the completed questionnaire, in the field, after its field data collection. Interviewers match with the respondents according to the preplanned survey design and interview allocation schemes. An interviewer is not allocated to the same respondent in the interview and the reinterview. Thus, we do not have replications in ND in our case (i.e., k=1). Note that the interviewers have training for sample respondent selection and questionnaire execution for few days. In case of nonresponse, a new respondent is determined by random substitution. The same approach is also used during the reinterview. Timing of interview, reinterview, and reconciliation survey was proposed by the World Fertility Survey Methodology Division for their 42 country surveys (WFS, 1977).
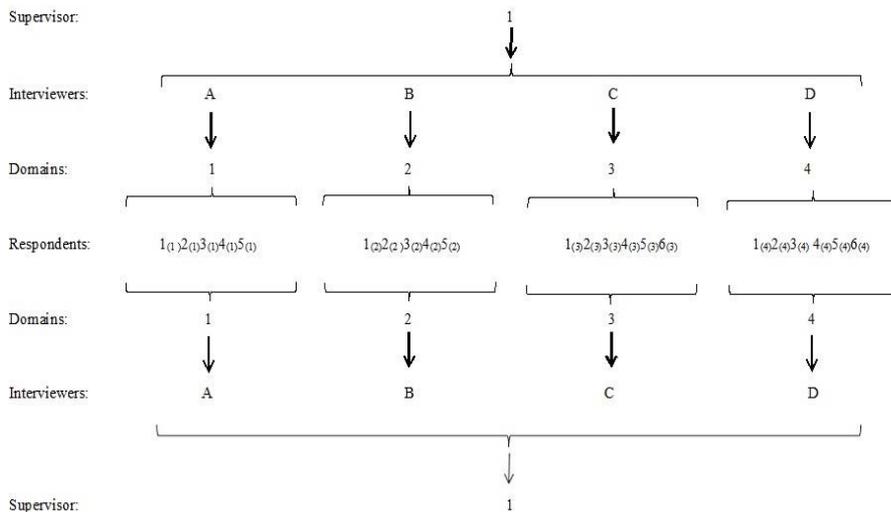
Figure 1. Pilot parent survey layout of the fieldwork

Table 1. Fieldwork allocation of the interviewers to respondent groups for the pilot survey

Domain Number	Respondent Number	Interview	Reinterview
1	$1_{(1)};2_{(1)};3_{(1)};4_{(1)};5_{(1)}$	Interviewer A	Interviewer B
2	$1_{(2)};2_{(2)};3_{(2)};4_{(2)};5_{(2)}$	Interviewer B	Interviewer A
3	$1_{(3)};2_{(3)};3_{(3)};4_{(3)};5_{(3)};6_{(3)}$	Interviewer C	Interviewer D
4	$1_{(4)};2_{(4)};3_{(4)};4_{(4)};5_{(4)};6_{(4)}$	Interviewer D	Interviewer C

By following the suggestions (Biemer and Stokes, 2004), after an interval of one month, the second round of the survey, the reinterview, is applied to the same respondents by exchanging the interviewers' domains using the same questionnaire paper (see Figure 2 and Table 1). Since the purpose of the reinterview is to estimate response variance and response bias, the original questions are repeated in their exact forms as suggested by Kish (1965). Note also that the execution of the questionnaires in both interview and reinterview is made on a voluntary basis to the respondent.
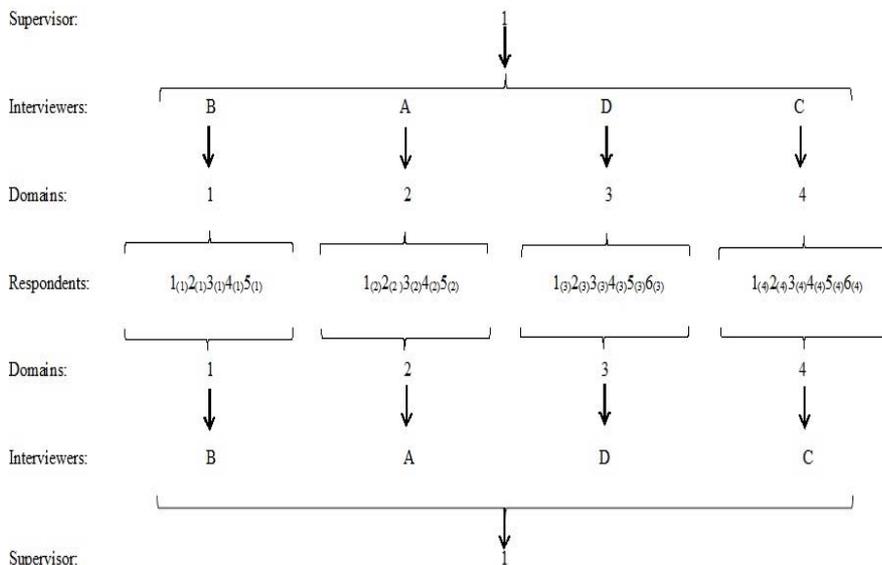
Figure 2. Pilot reinterview survey layout of the fieldwork

As a field operation, reinterviews are expensive, in face to face surveys. Because of its complex methodology, some survey designers would like to neglect this operation. On the other hand, nonsampling errors cover a larger amount of the total error, when compared with sampling errors. One should make a decision on the error versus cost of the survey operation. The reinterview is always conducted on a subsample of the original survey sample, the costs can be moderate. However, with the use of computer assisted interviewing, operational costs can be kept minimal while the usefulness of the reinterview is increased. The survey contained 10 basic questions, which are designed to cover a different range of data measurement levels such as dichotomy, polytomy, ordinal, interval (see Table 2).

The random effects model used for this survey is developed from model (31), and written as

$$y_{ij} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)} \begin{cases} i = 1,2, \dots ,4 \\ j = 1,2, \dots ,22 \end{cases}. \tag{32}$$

Here, μ represents the true value, τ_i the ith interviewer error, $\beta_{j(i)}$ is the jth respondent error nested under the ith interviewer, $\varepsilon_{(ij)}$ is the NID $(0,\sigma^2)$ random error term. Thus, in this design, we assume that there are four domains and from each domain a sample of size five, five, six and six respondents are drawn, respectively, without replacement.

This is an unbalanced design because the sizes of each interviewer's assignment are not the same. The response error analysis is conducted and response reliability measures are calculated for each question listed in Table 2 by using the formulas shown in Section 2, and the results obtained are presented in Table 3 and Table 4. Note here that to calculate the response reliability statistics for the Likert scale type questions, the agreement proportions between two interviews are obtain first (see Fahmi 2013).

Table 2. Pilot study questions and related information

Pilot Survey Question Number	Measurement scale	Variable name	Main Survey Question Number
1	Dichotomy	Gender of respondent	1
2	Interval	Age of respondent	2
3	Interval	Height of respondent	3
4	Likert scale	Last degree owned	5
5	Dichotomy	Title of respondent	6
6	Interval	Working duration in years in the university	8
7	Interval	Payment on clothing in TL* per month**	15
8	Interval	Payment on cultural activities in TL* per month**	16
9	Likert scale	Job satisfaction	17
10	Likert scale	Salary satisfaction	18

* Note that TL refers to Turkish Lira as currency; ** "per month" refers to any average month within the year.

Table 3. Response reliability statistics for the pilot survey

Ques. No.	A	D	$A_{k=1}$	$D_{k=1}$	IOI	IOI Eval.	r	K	K_w	K Eval.	CI for K
1	1.000	0.000	1.000	0.000	NA	NA	NA	1.000	1.000	AP	(1.00, 1.00)*
2	1.000	0.000	1.000	0.000	0.000	L	1.000	1.000	1.000	AP	(1.00, 1.00)*
3	1.000	0.000	1.000	0.000	0.000	L	1.000	1.000	1.000	AP	(1.00, 1.00)*
4	1.000	0.000	1.000	0.000	0.000	L	1.000	1.000	1.000	AP	(1.00, 1.00)*
5	1.000	0.000	1.000	0.000	NA	NA	NA	1.000	1.000	AP	(1.00, 1.00)*
6	1.000	0.000	1.000	0.000	0.001	L	0.999	1.000	1.000	AP	(1.00, 1.00)*
7	0.636	0.363	0.818	0.182	0.399	M	0.601	0.480	0.528	M	(0.20, 0.76)*
8	0.454	0.545	0.864	0.136	0.685	H	0.315	0.248	0.313	F	(-0.03, 0.53)
9	0.864	0.136	0.908	0.092	0.469	M	0.531	0.749	0.891	S	(0.53, 0.97)*
10	0.682	0.318	0.955	0.045	0.326	M	0.674	0.566	0.596	M	(0.32, 0.82)*

Notes: 1. NA indicates that this statistic is irrelevant for this type of variable; M: Moderate; L: Low; H: High; F: Fair; AP: Almost Perfect; S: Substantial 2. * denotes statistically significant kappa value for that particular type of question at α=0.05 level of significance by paired-t test. 3. Indices of *A*: crude agreement; *D*: crude disagreement; *IOI*: inconsistency; *r*: reliability of data; *K*: consistency (kappa) 4. Values of *A*, *r* and *K* close to one (1) indicate consistent and reliable data.

3.3. Main survey application

In the main survey, the questionnaires of the pilot survey are extended with some additional questions, and executed to the randomly selected faculty members of METU, Turkey, under a preplanned schema according to an ND, again in two rounds, namely, main and reinterview surveys. Details of both response error applications are presented below. Note that main survey data can be found in Fahmi (2013).

In this part of the study, following the methodology proposed by Ayhan (2003, 2012), Fahmi (2013), and Batmaz and Fahmi (2015), an ND is applied to allocate the interviewers to respondents in a life satisfaction and time use survey for METU academicians. The survey contains 20 questions, and it is applied to 168 academicians. They are randomly selected from METU's five faculties which have 839 academicians. The sample corresponds to 20% of the total number of the academicians working at METU. The number of faculty members and the corresponding sizes of the selected samples are given in Table 5. Note here that since this is an academic research, the sample size is limited to 168 academicians.

Table 4. Other response error statistics for the pilot survey

Ques. No.	$\bar{y}$	$\bar{y}_1$	$\bar{y}_2$	$d = \bar{y}_1 - \bar{y}_2$	CI for d	α	s_t^2	s_r^2	s_s^2	$\hat{\delta}_2 s_r^2$
1	NA	NA	NA	NA	NA	1.000	NA	NA	NA	NA
2	32.86	32.86	32.86	0	(0.00, 0.00)	1.000	0.00	0.00	0.00	0.00
3	170.56	170.59	170.55	0.04	(-0.24, 0.33)	0.998	0.00	0.00	0.00	0.00
4	2.18	2.18	2.18	0	(0.00, 0.00)	1.000	0.00	0.00	0.00	0.00
5	NA	NA	NA	NA	NA	1.000	NA	NA	NA	NA
6	7.4	7.30	7.60	-0.30	(-0.48, -0.10)*	0.998	53.13	0.09	53.04	0.11
7	142.70	133.86	151.59	-17.73	(-61.91, 26.46)	0.624	13.459	5.380	8.079	979
8	90.45	77.50	103.41	-25.91	(-47.99, 3.83)*	0.802	1.863	1.277	587	789
9	4.25	4.27	4.23	0.04	(-0.24, 0.34)	0.599	0.49	0.23	0.27	0.094
10	3.00	3.04	2.95	0.09	(-0.21, 0.39)	0.820	1.38	0.45	0.93	0.052

Notes: 1. NA indicates that this statistic is irrelevant for this type of variable, 2. * denotes statistically significant difference between the parent and reinterview values for that particular type of question at α=0.05 level of significance by paired-t test. 3. d: response deviance (error); α: interviewer effect; s_t^2: total variance; s_r^2: response variance; s_s^2: sampling variance $\hat{\delta}_2 s_r^2$: correlated interviewer variance 4. Values of D, IOI, and d close to zero (0) and also CI for K and for d containing zero indicate consistent and reliable data.

Table 5. Number of academicians at each METU faculty and the selected sample sizes

Faculties	Total number of academicians	Selected number of academicians
Architecture	52	10
Economic and Administrative Sciences	91	18
Education	80	16
Engineering	384	77
Arts and Sciences	232	47
Total	**839**	**168**

The respondents are divided into eight domains, and one interviewer is sent to each domain, randomly. As in the case of the pilot study, the numbers of each interviewer's assignments are not the same. Hence, an unbalanced ND design is formed again.

One month later, the second round of the main survey is applied to six respondents from each domain (See Table 5), which are again selected randomly by exchanging the interviewers' domains, and using the same original questionnaire (See Figure 4). The nested layout of the first round main survey fieldwork interview is shown in Figure 3.

In the main survey, the questionnaire is expanded to 20 questions by including 10 more questions related to academicians' time use in addition to the life satisfaction questions covered in the pilot study. Main survey questions and the related information is presented in Table 6. The number of respondents at each sample department in the main survey first round (interview) and the second round (reinterview) is given in Fahmi (2013). Note here that the same random effects model given in model (32), which is developed from model (31), is used.

4. Findings and discussion

In order to exemplify the methodology considered, a pilot and also a main sample survey are executed both in two rounds (interview–reinterview). The response error analysis is conducted and response reliability measures are calculated using formulas given in Section 2 for each question listed in Table 6, and the results obtained are presented in Table 7 and Table 8.

Response reliability measures for questions based on different levels of measurement scales are obtained for data collected from the sample surveys. Simple and correlated response errors are also estimated for different measurement scaled data. In this Section, the data obtained from these applications are evaluated with respect to the reliability measures and other statistics for all variables.
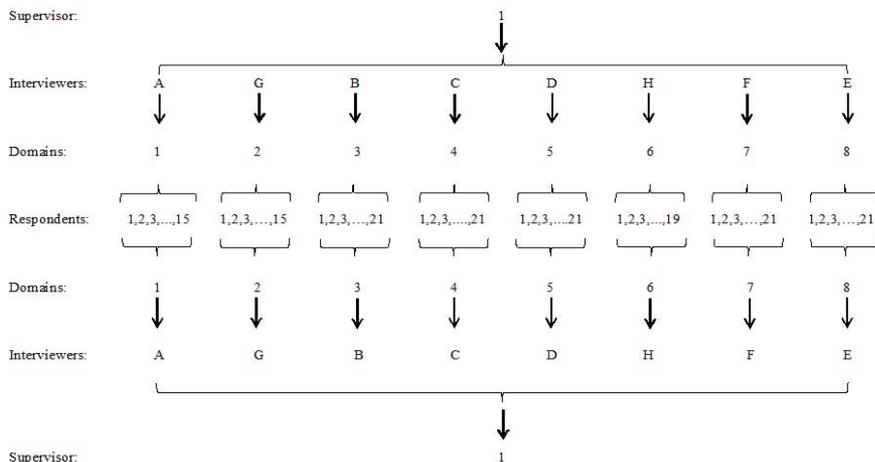
Figure 3. Parent main survey layout of the fieldwork
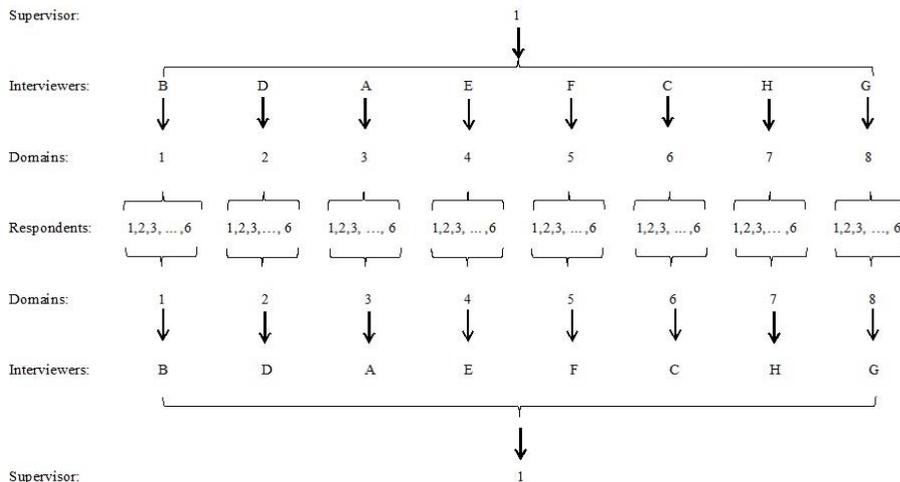
Figure 4. Main survey reinterview layout of the fieldwork

4.1. Findings of the pilot survey

When the results are examined, the following findings of the pilot study may be given as follows:

- For questions 1-5, we have completely reliable data with respect to all relevant indices considered; there exists ignorable response variance for question 6 with respect to *index of consistency* (*IOI*) and of *data reliability* ($r = 1 - IOI$) in Table 3, and also with respect to s_r^2 in Table 4. For the rest of questions (7-10), there exist response variances with respect to *IOI*, r and also s_r^2. Among them *interviewer effects* (α) are observed on questions 9, 7, 8 and 10, in decreasing order (Table 4).

Table 6.   Main survey questions and related information

Main Survey Question Number	Measurement scale	Variable name	Pilot Survey Question Number
1	Dichotomy	Gender of respondent	1
2	Interval	Age of respondent	2
3	Interval	Height of respondent	3
4	Dichotomy	Marital status of respondent	
5	Likert scale	Last degree owned	4
6	Likert scale	Title of the respondent	5
7	Likert scale	Number of languages known	
8	Interval	Working duration in years in the university	6
9	Interval	Fixed working duration in a day in hours[*]	
10	Interval	Sleeping duration a day in hours	
11	Interval	Time spent for leisure and sports per week	
12	Interval	Time spent for eating and drinking per day	
13	Interval	Time spent with their family per week	
14	Dichotomy	Interested in cooking	
15	Interval	Payment on clothing in TL[**] per month[***]	7
16	Interval	Payment on cultural activities in TL[**] per month[***]	8
17	Likert scale	Job satisfaction	9
18	Likert scale	Salary satisfaction	10
19	Likert scale	Working duration satisfaction	
20	Likert scale	Current use for time satisfaction	

[*] Note that the time frame for the working duration may not be well defined for this variable, and may create limitation and potential reason of variability from one round to another. [**] TL refers to Turkish Lira as a currency. [***] "per month" refers to any average month within the year.

- There is no response error at all in questions 2, 3 and 4 of type interval, interval, and ordinal, respectively, according to response deviation (d), and *CI for d*, which includes zero (Table 4). Also, the associated almost perfect *kappa* values (K=1.0) are found to be statistically significant based on *CI for K* (Table 3). Besides, according to α, responses in the two interviews are perfectly correlated (Table 4).

- For questions 1 and 5 of the dichotomy type, *crude agreement (A), disagreement (D)* and *consistency* (K) index values indicate statistically significant with respect to *CI for K* and perfect agreement with respect to evaluation of K between two interviews (Table 3).

- There is very small but statistically significant response error for question 6 according to *response deviation* (d), and *CI for d* (Table 4). However, response variance is accounted for only 0.1% of the total variance with respect to *index of inconsistency* (*IOI*) (Table 3). For the question inquiring about the working period duration of respondents, the estimators of the *uncorrelated response variance* (s_r^2) and *correlated interviewer variance* ($\hat{\delta}_2 s_r^2$) are found to be low (Table 4), indicating smaller interviewer effect on the respondents.

- For questions 7 and 8 of the interval type, which ask the respondents about the amount of money spend on clothing and on cultural activities, respectively, *response errors* (*d*) and their associated *response variances* (s_r^2) are the largest among the other questions (Table 4). However, there is a statistically significant difference in responses between two interviews only for question 8 with respect to *CI for d* (Table 4). The response variances are attributed to 40% and 69% of the total variance with respect to *IOI* for questions 7 and 8, respectively (Table 3). In addition, interviewer variances are attributed to 7% (=979/13.459) and 42% (=789/1.863) of the total variance for questions 7 and 8, respectively. According to the *kappa*, there is a moderate and fair agreement between responses of two interviews (Table 3), although correlation statistics do not indicate (α values are 0.775 and 0.947 for question 7 and 8, respectively (Table 4)).

Table 7. Response reliability statistics for the main survey

Ques. No.	A	D	A_k	D_k	IOI	IOI Eval.	r	K	K_w	K Eval.	CI for K
1	1.000	0.000	1.000	0.000	NA	NA		1.000	1.000	AP	(1.00, 1.00)*
2	0.937	0.063	1.000	0.000	0.011	L	0.989	0.908	0.952	AP	(0.81, 1.01)*
3	0.915	0.085	1.000	0.000	0.019	L	0.981	0.877	0.897	AP	(0.76, 0.99)*
4	1.000	0.000	1.000	0.000	NA	NA	NA	1.000	1.000	AP	(1.00, 1.00)*
5	1.000	0.000	1.000	0.000	0.000	L	1.000	1.000	1.000	AP	(1.00, 1.00)*
6	0.917	0.083	1.000	0.000	0.296	M	0.704	0.875	0.904	AP	(0.76, 0.99)*
7	0.833	0.167	1.000	0.000	0.143	L	0.857	0.583	0.621	M	(0.34, 0.83)*
8	0.896	0.104	1.000	0.000	0.048	L	0.952	0.859	0.883	AP	(0.73, 0.99)*
9	0.708	0.295	0.979	0.021	0.296	M	0.704	0.494	0.576	M	(0.27, 0.77)*
10	0.625	0.375	0.958	0.042	0.382	M	0.618	0.434	0.500	M	(0.23, 0.64)*
11	0.688	0.313	0.917	0.083	0.562	H	0.438	0.407	0.455	M	(0.19, 0.62)*
12	0.688	0.313	0.938	0.062	0.800	H	0.200	0.444	0.558	M	(0.23, 0.66)*
13	0.646	0.354	0.896	0.104	0.608	H	0.392	0.496	0.575	M	(0.32, 0.68)*
14	0.813	0.188	1.000	0.000	NA	NA	NA	0.631	0.698	S	(0.42, 0.84)*
15	0.542	0.458	0.813	0.187	0.340	M	0.660	0.317	0.414	F	(0.15, 0.49)*
16	0.667	0.333	0.917	0.083	0.171	L	0.829	0.301	0.385	F	(0.09, 0.51)*
17	0.604	0.396	1.000	0.000	0.442	M	0.558	0.360	0.427	F	(0.16, 0.56)*
18	0.562	0.438	0.938	0.062	0.302	M	0.302	0.399	0.522	F	(0.21, 0.59)*
19	0.542	0.458	0.938	0.062	0.547	H	0.547	0.263	0.356	F	(0.05, 0.48)*
20	0.646	0.354	0.938	0.062	0.265	M	0.265	0.521	0.602	M	(0.34, 0.70)*

Notes: 1. NA indicates that this statistic is irrelevant for this type of variable; M: Moderate; L: Low; H: High; F: Fair: AP: Almost Perfect; S: Substantial 2. * denotes statistically significant kappa value for that particular type of question at α=0.05 level of significance by paired-t test. 3. Indices of *A*: crude agreement; *D*: crude disagreement; *IOI*: inconsistency; *r*: reliability of data; *K:* consistency (kappa) 4. Values of *A*, *r* and *K* close to one (1) indicate consistent and reliable data.

- There exist very small and not statistically significant response errors associated with questions 9 and 10 of the ordinal type with respect to *response deviation* (*d*) and associated *CI for d*, asking about overall job and salary satisfaction of the respondents, respectively (Table 4). However, there exist response variances which are attributed to 47% and 33% of the total variance for questions 9 and 10 with respect to *index of*

consistency, *IOI*, respectively (Table 3). Also, an *interviewer effect* is detected with respect to α on the responses for these questions; however, they only account for 19% (=0.094/0.49) and 3.8% (=0.052/1.38) of the total variance for questions 9 and 10, respectively (Table 4). Reliability statistics indicate that there is substantial and moderate agreement between the responses of two interviews for questions 9 and 10 with respect to evaluation of kappa, *K* (Table 3), respectively; nevertheless, correlation statistics do not approve this (α values are 0.599 and 0.820 for question 9 and 10, respectively (Table 4)).

4.2. Findings of the main survey

Findings of the main survey may be given as follows:

- For some questions such as 1 and 4 with respect to *index of crude agreement*, *A*, in Table 6, and 5, 19 with respect to *response error, d,* in Table 7, there are no changes in the given responses. Inquiring about the satisfaction of respondents with daily working hours (Question 19), *indices of crude agreement (A) and of consistency (K)* have a fair agreement between responses of two interviews.

Table 8. Other response error statistics for the main survey

Ques. No.	$\bar{y}$	$\bar{y}_1$	$\bar{y}_2$	$d=$ $\bar{y}_1 - \bar{y}_2$	CI for d	α	s_t^2	s_r^2	s_s^2	$\hat{\delta}_2 s_r^2$
1	NA	NA	NA	NA	NA	1.000	NA	NA	NA	NA
2	47.83	46.48	46.98	-0.5	(-3.75, 1.67)	0.989	122.72	1.40	121.32	0.000
3	172.24	171.38	171.33	0.05	(-0.49, 0.58)	0.979	80.79	1.54	79.25	0.130
4	NA	NA	NA	NA	NA	1.000	NA	NA	NA	NA
5	3.00	3.00	3.00	0	(0.00, 0.00)*	1.000	0.00	0.00	0.00	0.000
6	1.98	2.00	1.95	0.05	(-0.04, 0.13)	0.937	0.86	0.04	0.81	0.040
7	1.31	1.27	1.35	-0.08	(-0.20, 0.03)	0.775	0.56	0.08	0.48	0.110
8	18.52	16.33	17.45	-1.12	(-2.08, 0.23)	0.947	146.25	7.03	139.22	0.004
9	8.51	8.27	7.97	0.3	(-0.16, 0.76)	0.674	4.26	1.26	3.00	0.003
10	7.01	7.14	7.02	0.12	(-0.10, 0.33)	0.651	0.76	0.29	0.48	0.002
11	1.16	1.26	1.49	-0.23	(-0.66, 0.19)	0.635	1.85	1.04	0.81	0.047
12	1.58	1.65	1.70	-0.05	(-0.28,0.18)	0.403	0.45	0.36	0.10	0.046
13	3.45	3.56	3.31	0.25	(-0.44, 0.94)	0.641	4.72	2.87	1.86	0.068
14	NA	NA	NA	NA	NA	1.000	NA	NA	NA	NA
15	153.73	164.40	189.27	-24.87	(-66.69, 16.94)	0.822	30,538	10,369	20,169	2.395
16	144.20	116.98	106.56	10.42	(-19.37, 40.20)	0.659	26,475	4,525	21,950	822.44
17	4.281	4.31	4.25	0.06	(-0.12, 0.25)	0.665	0.43	0.19	0.24	0.229
18	2.615	2.60	2.63	-0.03	(-0.27, 0.23)	0.720	1.29	0.39	0.91	0.197
19	4.271	4.27	4.27	0	(-0.25, 0.25)	0.374	0.75	0.41	0.34	0.144
20	3.39	3.44	3.33	0.11	(-0.13, 0.34)	0.703	0.98	0.26	0.72	0.646

Notes: 1. NA indicates that this statistic is irrelevant for this type of variable, 2. * denotes statistically significant difference between the parent and reinterview values for that particular type of question at α=0.05 level of significance by paired-t test. 3. *d*: response deviance (error); *α*: interviewer effect; s_t^2: total variance; s_r^2: response variance; s_s^2: sampling variance $\hat{\delta}_2 s_r^2$: correlated interviewer variance 4. Values of *D*, *IOI*, and *d* close to zero (0) and also *CI for K* and for *d* containing zero indicate consistent and reliable data.

- Reliable data belong to questions 1-8 and 14 with respect to A and r in Table 6. Asking about the height of respondents (Question 3), the *correlated interviewer variance, $\hat{\delta}_2 s_r^2$,* is found to be very low (Table 7). Asking about the academic title of the respondents (Question 6), indices of *crude agreement, A, and of consistency, K,* have almost a perfect agreement between responses of the two interviews (Table 6). For the number of languages known asked in question 7, *index of consistency, K,* have a moderate agreement between responses of the two interviews (Table 6). Inquiring about if s/he is interested in cooking (Question 14), *index of consistency, K,* have a substantial agreement between responses (Table 6).

- Data belonging to questions 9, 10, 15, 16 have a moderate to fair agreement and reliability (Table 6).

- The least reliable data belong to questions 11, 12, 13, 19; response variance accounts for 56%, 80%, 61%, 55% with respect to *IOI* (Table 6), and interviewer variance accounts for 7% (=0.047/1.85), 10% (=0.046/0.45), 1.4% (=0.068/4.72), 19% (=0.144/0.75) of the total variance for questions 11, 12, 13, 19, respectively.

- Asking about the level of job satisfaction, Question 17 has a fair agreement between responses of the two interviews with respect to the index of consistency (K) (Table 6).

- Inquiring about the current salary satisfaction, Question 18 has a fair agreement between responses of the two interviews with respect to *indices of crude agreement, A, and of consistency, K* (Table 6).

- Asking about the satisfaction with the time use of respondents, Question 20 has a moderate agreement between responses of the two interviews with respect to indices of *crude agreement, A, and of consistency, K* (Table 6).

5. Conclusions

The main aim of this work is to investigate response errors which may stem from the respondent, interviewer or from their interaction, under interview-reinterview settings in sample surveys. We suggest using NDs in interview-reinterview surveys due to several reasons. First, an ND naturally provides estimation of interviewer effects due to its nested structure in which one respondent is interviewed by many interviewers. Next, it provides computing response errors independently in each survey. And also, it provides flexibilities in the field allocation and applications. In order to apply the suggested approach, an interview-reinterview survey is conducted at METU, Ankara, Turkey, to investigate the satisfaction of academicians' regarding life and time use.

Analysis of the pilot survey reveals that we have completely reliable data sometimes with an ignorable response variance on the questions inquiring about factual information about the participants such as gender, title, and so on. Nevertheless, there exist response variances in the questions involving elements hard to quantify, such as "amount of payment" or "duration". Besides, questions asking about respondents'

feelings, such as their "satisfaction level", seem to open higher interviewer effects. The last two questions are usually formulated with either ordinal or interval type of variables. Analysis of the pilot study reveals that the response reliability seems to be irrelevant to these two data types.

Analysis of the main survey results show that almost three fourth of the data are reliable and almost reliable. The rest of the questions are exposed to interviewer effects, and need more attention. Note that the response error that may be mostly attributed to the interviewer effect belongs to the one regarding respondents' satisfaction. As a result of the training provided to the interviewers' immediately after the pilot survey analysis, in the main survey, the associated response variances accounting for the total variance are considerably reduced from 69% to 29%. It is the outcome of having done two consecutive interview-reinterview designs.

As a future study, alternative interviewer allocation can also be examined on the basis of both the nested and factorial experimental design techniques. However, under such allocations, the number of interviewers which will be allocated for the field application may be combinatorically problematic.

The following limitations should be kept in mind while evaluating the response reliability measures. The time lag between two interviews should be reasonably large, enabling to recall their first response during the second interview. This issue is clarified by the following related literature. The World Fertility Survey's document on "Re-interview Survey Design" resulted in the fieldwork applications as the median lag between the first interview and the reinterview was 2 to 4 months for the planned five national studies (O'Muircheartaigh and Marckwardt, 1980). Also, O'Muircheartaigh (1982) suggests a regression type analysis to test independence between the two interviews. The time lag between the original interview and the reinterview varies between a few days to several months. Also, research on optimal time lags in different reinterview situations is rare in the literature (Forsman and Schreiner, 2004). Memory recall errors are affected by the *time duration between the two reference points* (interview and reinterview) as well as the *importance of the event*, *frequency of occurence of the event*, *measurement scale of the event*, and *bounded* or *aided recalling* (Ayhan and Işıksal, 2004). Consequently, the time difference between the two cannot be the only criteria for evaluation. When the time interval between the interview and the reinterview is very short, then reinterviewed respondents can recall their earlier responses, and they may be losing interest, also it is possible to agree with their previous reply to the interview as a form of satisficing. For very lengthy questionnaires, respondent fatigue is also possible, but may not be the case for short questionnaire surveys.

Acknowledgements

References

Ayhan, H. Ö., (2003). Models of response error components in supervised interview–reinterview surveys. *Journal of Applied Statistics*, 30(9), pp. 1047–1054.

Ayhan, H. Ö., (2012). Response reliability and response error estimation in personal interview surveys. AISC: International Conference on Advances in Interdisciplinary Statistics and Combinatorics, Greensboro, NC, October 5–7, 2012.

Ayhan, H. Ö., Işiksal, S., (2004). Memory recall errors in retrospective surveys: A reverse record check stıdy. *Quality and Quantity*, 38, pp. 475–493.

Bailar, B. A., Dalenius, T., (1969). Estimating the response variance components of the U.S. Bureau of the Census' survey model. *Sankhya: The Indian Journal of Statistics*, Series B, Vol. 31(3/4), pp. 341–360.

Bassi, F., Fabbris, L., (1997). Estimators of nonsampling errors in interview – reinterview supervised surveys with interpenetrated assignments. In *Survey Measurement and Process Quality*, L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz, and D. Trewin, eds., John Wiley and Sons, New York, pp. 733–751.

Batmaz, I., Fahmi, F. M., (2015). Response variance estimation in personal interview surveys with several allocation schemes, *Proceedings of the 60th World Statistics Congress of the International Statistical Institute*, Rio de Janeiro, Brazil, pp. 1–6.

Biemer, P. P., (2012). *Latent Class Analysis of Survey Error*. John Wiley and Sons, New Jersey.

Biemer, P. P., Lyberg, L. E., (2003). *Introduction to Survey Quality.* Hoboken: Wiley Series in Survey Methodology

Biemer, P. P., Stokes, S. L., (1985). Optimal design of interviewer variance experiments in complex surveys. *Journal of the American Statistical Association*, 80, pp. 158–166.

Biemer, P. P., Stokes, S. L., (2004). Approaches to the modeling of measurement error. In *Measurement Errors in Surveys* Edited by P. Biemer, R.M. Groves, L. Lyberg, N.A. Mathiowetz, and S. Sudman. Wiley Series in Survey Methodology, New Jersey, pp. 487–516.

Biemer, P. P., Groves, R.M., Lyberg, L. E., Mathiowetz, N. A., Sudman, S., (2004). *Measurement Errors in Surveys*. Wiley Series in Survey Methodology, New Jersey.

Box, G. E. P., Hunter, J. S., Hunter, W. G., (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*. John Wiley and Sons, New Jersey.

Cicchetti, D. V., (1972). A new measure of agreement between rank-ordered variables. *Proceedings of the 80th Annual Conference of the American Psychological Association* 1972, pp. 17–18.

Cohen, J., (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, pp. 37–46.

Cohen, J., (1968). Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, pp. 213–220.

Dahlhamer, J., Zablotsky, B., Zelaya, C., Maitland, A., (2020). Modeling interviewer effects in the National Health Interview Survey. Chapter 21 in Olson, K., Smyth, J.D., Dykema, J., Holbrook, A.L., Kreuter, F., West, B.T., Eds. (2020 a). *Interviewer Effects from a Total Survey Error Perspective.* New York: Chapman and Hall/CRC., pp. 295–310.

Dijkstra, W., (1983). How interviewer variance can bias the results of research on interviewer effects. *Quality and Quantity*, 17, 179–187.

Edwards, B., Sun, H., Hubbard, R., (2020). Behavior change techniques for reducing interviewer contributions to total survey error. Chapter 6 in Olson, K., Smyth, J.D., Dykema, J., Holbrook, A.L., Kreuter, F., West, B.T., Eds. (2020 a). *Interviewer Effects from a Total Survey Error Perspective.* New York: Chapman and Hall/CRC., pp. 77–90.

Fahmi, F. M., (2013). *Estimation of Response Errors in Complex Sample Surveys.* Ph.D. thesis, Graduate School of Natural and Applied Sciences, Middle East Technical University, Ankara, Turkey.

Fellegi, I. P., (1964). Response variance and its estimation. *Journal of the American Statistical Association*, 69, pp. 19–54.

Forsman, G., Schreiner, I., (2004). The design and analysis of reinterview: An overview. Chapter 15 in *Measurement Errors in Surveys* Edited by P. Biemer, R.M. Groves, L. Lyberg, N.A. Mathiowetz, and S. Sudman. Wiley Series in Survey Methodology, New Jersey, pp. 279–301.

Fuller, W. A., (1995). Estimation in the presence of measurement error. International Statistical Review, 63(2), pp, 121–147.

Groves, R. M., Magilavy, L. J., (1986). Measuring and explaining interviewer effects in centralized telephone facilities. *Public Opinion Quarterly*, 50, pp. 251–266.

Moser, C. A., Kalton, G., (1979). *Survey Methods in Social Investigation.* Heinemann Educational Books, London.

Hansen, M. H., Hurwitz, W. N., Bershad, M. A., (1961). Measurement errors in censuses and surveys. *Bulletin of the International Statistical Institute*, 38(2), pp. 359–374.

Hansen, M. H., Hurwitz, W. N., Marks, E. S., Mauldin, W. P., (1951). Response errors in surveys. *Journal of the American Statistical Association*, 46(254), pp. 147–190.

Hansen, M. H., Hurwitz, W. N., Pritzker, L., (1964). The estimation and interpretation of gross differences and the simple response variance. In C.R. RAO (Editor): Contributions to Statistics, Presented to Professor P.C. Mahalanobis on the Occasion of his 70th Birthday. Calcutta, India: Pergamon Press, Oxford and Statistical Publishing Society.

Hagenaars, J. A., Heinen, T. G., (1982). Effects of role-independent interviewer characteristics on responses. In W.Dijkstra and J.van der Zouwen (eds.), *Response Behaviour in the Survey Interview.* Academic Press, London, pp. 91–130.

Hox, J. J., (1994). Hierarchical regression models for interviewer and respondent effects. *Sociological Methods and Research*, 22, pp. 300–318.

Hox, J. J., De Leeuw, E. D., Kreft, I. G. G., (2004). The effect of interviewer and respondent characteristics on the quality of survey data: A multilevel model. Chapter 22 in Biemer *et al.* (2004) *Measurement Errors in Surveys*. Wiley Series in Survey Methodology, New Jersey, pp.439–461.

Kalton, G., (1983). Introduction to survey sampling. *Quantitative Applications in the Social Sciences, A SAGE University Papers Series,* No. 35, Newbury Park.

Kish, L., (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association*, 57, pp. 91–115.

Kish, L., (1965). Survey Sampling. John Wiley and Sons, New York.

Landis, J. R., Koch, G., (1977). A review of statistical methods in the analysis of data rising from observer reliability studies. *Statistica Neerlandica*, 29, pp. 101–123 and pp. 151–161.

Lyberg, L., Kasprzyk, D., (2004). Data collection methods and measurement error: An overview. Chapter 13 in Biemer *et al.* (2004) *Measurement Errors in Surveys*. Wiley Series in Survey Methodology, New Jersey, pp. 237–257.

Mahalanobis, P. C., (1946). Recent Experiments in Statistical Sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, Vol. 109, pp. 325–378.

McCarthy, P. J., (1966). Replication: An approach to the analysis of data from complex surveys. National Center for Health Statistics, *Vital and Health Statistics Series*, No. 2(14).

McNabb, D. E., (2014). *Nonsampling Error in Social Surveys.* Los Angeles: SAGE Publications, Inc.

Montgomery, D. C., (2012). *Design and Analysis of Experiments.* New York: John Wiley and Sons.

Olson, K., Smyth, J. D., Dykema, J., Holbrook, A. L., Kreuter, F., West, B. T., Eds., (2020a). *Interviewer Effects from a Total Survey Error Perspective.* New York: Chapman and Hall/CRC.

Olson, K., Smyth, J. D., Dykema, J., Holbrook, A. L., Kreuter, F., West, B. T., (2020 b). The past, present, and future of research on interviewer effects. Chapter 1 in Olson, K., Smyth, J.D., Dykema, J., Holbrook, A.L., Kreuter, F., West, B.T., Eds. (2020 a). *Interviewer Effects from a Total Survey Error Perspective.* New York: Chapman and Hall/CRC., pp. 3–16.

O'Muircheartaigh, C. A., (1976). Response errors in attitudinal sample surveys. *Quality and Quantity*, 10, pp. 97–115.

O'Muircheartaigh, C. A., (1977). Response errors. In *Analysis of Survey Data*, Vol. 2, John Wiley and Sons, London, pp. 193–239.

O'Muircheartaigh, C. A., (1982). *Methodology of the Response Errors Project.* Fertility Survey, Scientific Reports, No. 28, London, UK. 32 pp.

O'Muircheartaigh, C. A., (1984). The pattern of response variance in the Peru fertility survey. *WFS Scientific Report*, No. 45.

O'Muircheartaigh, C. A., (1986). Correlates of reinterview response inconsistency in the Current Population Survey. *Proceedings of the Second Annual Research Conference*, U.S. Bureau of the Census, pp. 208–234.

O'Muircheartaigh, C. A., (1991). Simple response variance: estimation and determinants. In *Measurement Errors in Surveys* Edited by P. Biemer, R.M. Groves, L. Lyberg, N.A. Mathiowetz, and S. Sudman. Wiley Series in Survey Methodology, New Jersey, pp. 551–574.

O'Muircheartaigh, C. A., Campanelli, P., (1998). The relative impact of interviewer effects and sample design effects on survey precision. Journal of the Royal Statistical Society, A 161, pp. 63–77.

O'Muircheartaigh, C. A., Marckwardt, A. M., (1980). An assessment of the reliability of WFS data. World Fertility Survey Conference 1980 London, Record of Proceedings Volume 3, pp. 313–379.

Salant, P., Dillman, D. A., (1994). *How to Conduct your own Survey?* John Wiley and Sons, New York.

Verma, V., (1980). Sampling for national fertility surveys. World Fertility Survey Conference 1980 London, Record of Proceedings, Vol. 3, pp. 389–448.

West, B. T., (2020). Designing studies for comparing interviewer variance in two groups of survey interviewers. Chapter 23 in Olson, K., Smyth, J.D., Dykema, J., Holbrook, A.L., Kreuter, F., West, B.T., Eds. (2020 a). *Interviewer Effects from a Total Survey Error Perspective.* New York: Chapman and Hall/CRC., pp. 323–334.

WFS, (1977). *Guidelines for Country Report No.1.* WFS Basic Documentation No.8, London, U.K.: World Fertility Survey.

Yu, C. H., Ohlund, B., Digangi, S., Jannasch-Pennell, A., (2000). Estimating the reliability of self-reporting data for web-based instruction. *Instruction and Research Support, Arizona State University*, US.

A Bayes algorithm for model compatibility and comparison of ARMA(p,q) models

Praveen Kumar Tripathi,[1] Rijji Sen,[2] S. K. Upadhyay[3]

ABSTRACT

The paper presents a Bayes analysis of an autoregressive-moving average model and its components based on exact likelihood and weak priors for the parameters where the priors are defined so that they incorporate stationarity and invertibility restrictions naturally. A Gibbs-Metropolis hybrid scheme is used to draw posterior-based inferences for the models under consideration. The compatibility of the models with the data is examined using the Ljung-Box-Pierce chi-square-based statistic. The paper also compares different compatible models through the posterior predictive loss criterion in order to recommend the most appropriate one. For a numerical illustration of the above, data on the Indian gross domestic product growth rate at constant prices are considered. Differencing the data once prior to conducting the analysis ensured their stationarity. Retrospective short-term predictions of the data are provided based on the final recommended model. The considered methodology is expected to offer an easy and precise method for economic data analysis.

Key words: ARMA model, exact likelihood, Gibbs sampler, Metropolis algorithm, posterior predictive loss, model compatibility, Ljung-Box-Pierce statistic, GDP growth rate.

1. Introduction

The general form of an autoregressive-moving average (ARMA) model, of order p and q, is defined as

$$y_t = \theta_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} \tag{1}$$

where y_t's are the time series observations, θ_0 is the intercept, ϕ_i's and ψ_j's are autoregressive (AR) and moving average (MA) coefficients, respectively, and ε's are the independent and identically distributed (iid) components of the Gaussian white noise distributed with mean zero and variance σ^2. We shall denote the model (1) by ARMA(p,q). There may be, of course, several choices of p and q but the large choices usually complicate the models and often lead to intractable solutions. This paper, therefore, considers a few arbitrary choices of p and q such that $p+q \leq 2$ and then finally recommends a model that is

---

[1]Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India.
E-mail: pkt2088@gmail.com. ORCID: https://orcid.org/0000-0002-5652-8301.
[2]Department of Statistics, Behala College, Calcutta University, India. E-mail: rijjisen@gmail.com.
ORCID: https://orcid.org/0000-0002-7805-037X.
[3]Department of Statistics, Banaras Hindu University, Varanasi, India. E-mail: skupadhyay@gmail.com.
ORCID: https://orcid.org/0000-0001-6051-5821.

most appropriate for our intended objective. It is to be noted that the choice of $p + q \leq 2$ simplifies the model considerably.

The literature is abundant with some significant references on classical analyses of ARMA models and its variants. Box and Jenkins (1976) popularized the ARMA model by simplifying the analysis in the classical framework, especially with reference to order identification of the model via autocorrelation function (ACF) and partial autocorrelation function (PACF). Later, Tsay and Tiao (1984, 1985) developed a unified approach for order identification for both stationary and non-stationary ARMA models. They proposed some consistent estimates of the auto-regressive parameters, which in turn were utilized to define an extended sample ACF for determination of the order of an AR model. The method proposed was mostly appropriate for non-seasonal data. However, in the case of seasonal data, a number of studies have used filtering approach. Reilly (1980) used an automatic methodology, similar to that used by Box and Jenkins (1976), to model the macroeconomic variables, like gross domestic product (GDP). A similar method was developed by Reynolds et al. (1995), which is more automatic and well illustrated by utilizing the time-series data for a single variable. Recently, Tripathi et al. (2018) have used Box-Jenkins methodology on the Integrated form of ARMA model.

The Bayesian analysis of ARMA models has a vast literature and the references to this context include Zellner (1971), Monahan (1983), Marriot and Smith (1992), Chib and Greenberg (1994), Marriot et al. (1996), Kleinbergen and Hoek (2000), Fan and Yao (2008) and Tripathi et al. (2017), etc. Among these, a sophisticated numerical integration technique was used by Monahan (1983) and that was later extended by Marriot and Smith (1992). Considered as a breakthrough in the analysis of ARMA models, the study done by Chib and Greenberg (1994) was the first to use the Markov chain Monte Carlo (MCMC) technique. In this paper, the authors relied on state space version of the model. Although based on more realistic assumptions than those used in the preceding works, Chib and Greenberg (1994) do have the disadvantage of carrying out the analysis on only a subset of the parameter set. Marriot et al. (1996) is another significant reference where the authors developed sampling-based approach for the estimation of parameters of ARMA model and its components. They suggested sampling from the conditional densities of AR and MA coefficients subject to the restriction of stationarity and invertibility. The present paper provides the full Bayesian analysis of ARMA models with special focus on model compatibility and comparison. It is based on a more logical and more computation friendly formulation of the ARMA likelihood in comparison with that of Tripathi et al. (2017). More elaborately speaking, the previous work uses an approximate conditional likelihood. The present work brings an improvement in the analysis by considering the joint density of the previous observations in addition to the conditional likelihood (as suggested by Box et al. (2004)). Thus, this is a closer approximation to the exact likelihood, which is given by Newbold et al. (1974). Naturally, the analysis proposed in the present paper appears to be more accurate.

To the best of our belief, none of the papers on ARMA models addresses the problems of verifying model compatibility as well as model comparison using the tools of the Bayesian paradigm. Since the idea of prediction is an integral part to any time series analysis, we use this to verify the compatibility of the considered models. The basic idea is to judge whether the predicted data are in compliance with the observed data. The discrepancy, if any, is

quantified by the calculation of the Bayesian p-value, which essentially uses a discrepancy measure, the Ljung-Box-Pierce statistic in our case. Realising the indispensability of the idea of prediction, we have further used this notion in comparing the different ARMA sub-models considered in the paper. It may be noted that the earlier papers mostly considered the Box-Jenkins methodology to select the models (see, for example, Reilly (1980), Tsay and Tiao (1984, 1985), Reynolds et al. (1995), Tripathi et al. (2018), etc.). This does not appear logical in some sense and, in no way, complies with the Bayesian paradigm. On the other hand, we have used the predictive loss criterion that successfully merges the ideas of prediction and the loss incurred thereof, making it true to Bayesian sensibilities.

Recently, Tripathi et al. (2017) performed an approximate Bayes analysis of ARMA model (1) and used the GDP growth rate data of India to illustrate their procedure. In this paper, the authors resorted to using an approximate form of the likelihood by considering the values of observations and the subsequent error terms, prior to the very first observation, say y_1, to be zero. Although the approximations were made keeping in mind the computational ease, it does have a logical lacuna to some extent. If one considers time series processes such as the GDP growth rate data, one must remember that the dataset considered is actually a part of a large data series and, in reality, there are non-zero observations before y_1. The present treatment of the problem, therefore, adds a reasonable amount of soundness by considering a more logical approximation of the likelihood function for the ARMA model and its components, based on a line of suggestion by Box et al. (2004). As a further extension of the work done by Tripathi et al. (2017), we used a Gibbs-Metropolis hybrid scheme to perform the complete Bayesian analysis of ARMA models instead of pure Gibbs sampler using adaptive-rejection sampling (see Tripathi et al. (2017)).

Undoubtedly, the ARMA model has the capability to model a variety of observations on time series data probably because of its generality and flexibility. The present paper is no more an exception and considers a time series data on GDP growth rate of India at constant prices (considering base year to be 2004-05) collected over a period of 1951-52 to 2013-14 and uses the ARMA model to explain and analyze the data (see also Tripathi et al. (2017)). The use of ARMA models to explain GDP growth rate data is quite prevalent in the literature (see, for example, Morley et al. (2003) and Ludlow and Enders (2000)). We can motivate the model based on two lines of thought. First, GDP growth in one quarter is obviously affected by the same in other quarters. It is similar to the situation when the current value in a time series can be considered to depend upon the lagged observations. Second, the GDP observations at a particular point of time are not only affected by the random shocks at that time, but also by the shocks (such as a natural disaster) that have taken place earlier. Hence, an ARMA model that takes care of these two aspects of modelling is quite plausible. It may be noted that this motivation is general and can be applied to other time series as well.

It is our understanding that the choice of a particular model cannot be completely pre-specified by theoretical consideration alone, rather it should be selected from many competing models using some model selection criteria. In most of the studies using the ARMA model, the particular model is selected using the Box-Jenkins methodology, that is, with the help of autocorrelation and partial autocorrelation functions (see, for example, Tsay and Tiao (1984, 1985) and Pankratz (1983)). Forecasting or predicting is an indispensable element of time series. Gelfand and Ghosh (1998) suggested a criterion based on minimization

of posterior predictive loss (PPL) arising due to a model. Going by this criterion, models are rewarded not only for their predictive capabilities, but also for their fidelity to the observed data. Since in time series analysis prediction is the ultimate objective, the proposed model selection criterion seems to be quite appealing while keeping in mind the criterion of prediction. We finally look into retrospective prediction because of the general belief that a model which predicts well retrospectively is expected to do well at least for the short term prospective prediction (see, for example, Tripathi et al. (2017, 2018)) in majority of cases, although not always.

The outline of the paper is as follows. The next section provides the stationarity and invertibility conditions for the various components of the considered ARMA model. Section 3 provides the Bayesian formulation of the ARMA model for its possible posterior analysis. The analysis is extended for AR and MA components as well, although the developments are routine once the general formulation for the ARMA model is obtained. Vague priors are used for the model parameters and posterior analysis is performed using Gibbs sampler algorithm after imposing the necessary conditions for stationarity and invertibility. The algorithm is actually a hybrid scheme with Metropolis algorithm within the Gibbs sampler as some of the full conditionals are not available for routine sample generation. Finally, an explanation is given for getting predictive samples once the posterior samples are obtained. Section 4 discusses the model compatibility issues based on Ljung-Box-Pierce statistic. The section also comments briefly on the PPL criterion of Gelfand and Ghosh (1998) as a tool for comparison of the compatible models. Section 5 considers the Indian GDP growth rate data and analyses the same using the formulation given in the previous sections. The analysis is done for a number of combinations of p and q such that $p+q \leq 2$ and finally the entertained models are compared using the PPL criterion. Some numeric evidences for the adequacy of the models are given and, also, the short term retrospective prediction based on the final selected model is considered. The last section is a brief conclusion that summarizes our general findings. The paper also has an appendix with some supplementary developments required for the completeness.

2. Stationarity and Invertibility conditions

In order that a time series model becomes reasonable, two vital conditions, namely the stationarity and the invertibility, need to be verified. The stationarity and invertibility conditions in the ARMA(p,q) $(p+q \leq 2)$ model can be defined separately in terms of its AR and MA components, respectively. The exact forms have been extensively worked out in the literature (see, for example, Pankratz (1983) and Box et al. (2004)). Without going into the various aspects of deriving these conditions, we can directly state them as follows. For AR(1) or ARMA(1,0), the condition of stationarity is simply, $|\phi_1| < 1$ whereas for AR(2) or ARMA(2,0), these conditions are

$$|\phi_2| < 1, \tag{2}$$

$$\phi_1 + \phi_2 < 1, \tag{3}$$

and
$$\phi_2 - \phi_1 < 1. \tag{4}$$

Similarly for MA(1) or ARMA(0, 1), the condition of invertibility is simply $|\psi_1| < 1$ whereas for MA(2) or ARMA(0, 2), these conditions are

$$|\psi_2| < 1, \tag{5}$$

$$-\psi_1 - \psi_2 < 1, \tag{6}$$

and
$$\psi_1 - \psi_2 < 1. \tag{7}$$

Also, for the ARMA(1, 1) model the conditions of stationarity as well as invertibility are applied simultaneously and these are simply,

$$|\phi_1| < 1 \qquad \text{and} \qquad |\psi_1| < 1. \tag{8}$$

We shall denote the stationarity and invertibility regions for AR(p) and MA(q) models by C_p and C_q, respectively. Non-stationarity in a time series can occur in several ways and should be handled accordingly. Say, for instance, non-stationarity in the variance can be checked by considering some appropriate transformations on the variates whereas non-stationarity in the location can be checked by using differenced data instead (see, for example, Shumway and Stoffer (2011)). It has been observed that the stationarity and invertibility conditions are very complicated for higher order ($p > 2$ and $q > 2$) ARMA models (see Marriot et al. (1996)) and this is perhaps the reason that higher order is often avoided in a true sense. Exception includes Okereke et al. (2015), where the authors derived and illustrated the consequences of the invertibility conditions on the parameters of MA process of order three. We shall not go into the details of such situations although the interested readers may refer to Fan and Yao (2008), Box et al. (2004), etc.

Several tests have been proposed to check the stationarity of a time series. Among them, two are commonly used in the time series literature. These are augmented Dickey-Fuller (ADF) test (see, for example, Dickey and Fuller (1979)) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (see, for example, Kwiatkowski et al. (1992)). ADF test is a unit root test and it is based on t-test of the coefficients of a generalized AR process. KPSS test, on the other hand, tests the stationarity of the process and assumes that the time series can be represented as the sum of a deterministic trend, a random walk and a stationarity error. The final conclusion in both the tests is normally drawn on the basis of p-values. For more details on the two tests, one may refer to Dickey and Fuller (1979) and Kwiatkowski et al. (1992).

3. Bayesian Model Formulation and Posterior Simulation

Let $\underline{y} : y_1, y_2, ..., y_T$ be the time series observations from the assumed ARMA(p, q) model. The exact likelihood function corresponding to ARMA(p, q) model is always difficult to write since the observation at any stage depends on its p lagged observations and we may not have lagged observations corresponding to the first p observed time series data sets. Newbold (1974) gave the exact form of likelihood for the general ARMA(p, q) model that

is certainly pragmatic but quite difficult at least computationally except in the case of very small sample sizes. Tools such as those based on Monte Carlo or other sample-based approaches also do not support much if employed directly on the likelihood suggested by Newbold (1974). As an alternative, Marriot et al. (1996) suggested a computational friendly form of the likelihood function for ARMA model by introducing the latent variables, $\mathbf{y_0} = (y_0, y_{-1}, ..., y_{1-p})$ and $\varepsilon_0 = (\varepsilon_0, \varepsilon_{-1}, ..., \varepsilon_{1-q})$ into the existing set of unknowns. This form was certainly easy to implement but resulted in the increase of dimensionality of (unknowns) parameter space.

Tripathi et al. (2017) recently used an approximate form of the likelihood function of a general ARMA (p,q) model. This approximation, although easy to implement, has a limitation in the sense that it assumes all the components of $\mathbf{y_0}$ and ε_0 as zero, providing no contribution of these components in the likelihood function. An alternative strategy was suggested by Box et al. (2004) although the strategy provided an approximation to the exact likelihood. Since the latent variables $\mathbf{y_0}$ and ε_0 are unknowns, Box et al. (2004) suggested to consider likelihood as the product of two terms where the first term is the joint density of first p observations $(y_1, y_2, ..., y_p)$ and the second term may be considered as the product of conditional density of remaining observations with each observation conditioned on its p lagged observations. This approach is certainly an approximation but makes sense when T is large but p is comparatively small. Moreover, the components of ε_0 can be easily taken to be zero. Obviously, the resulting likelihood function can be written as

$$L(y_1, y_2, ..., y_T | \theta_0, \Phi, \Psi) = f(y_1, y_2, ..., y_p | \theta_0, \Phi, \Psi) \times \prod_{t=p+1}^{T} f(y_t | y_{t-1}, y_{t-2}, ..., y_{t-p}; \theta_0, \Phi, \Psi),$$

(9)

where $\Phi = (\phi_1, ..., \phi_p)$ and $\Psi = (\psi_1, ..., \psi_q)$. The likelihood in (9) can be considered as an extension of the likelihood proposed by Tripathi et al. (2017) and it can be obtained, up to proportionality, as (see Appendix)

$$L(y_1, y_2, ..., y_T | \theta_0, \Phi, \Psi) \propto \left(\frac{1}{\sigma^2} \right)^{T/2} \times |V_{\phi, \psi}|^{-1/2} \times$$

$$\exp \left(-\frac{1}{2\sigma^2} \{ (Y_p - \mu_p)' V_{\phi, \psi}^{-1} (Y_p - \mu_p) + \sum_{t=p+1}^{T} (y_t - \theta_0 - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{q} \psi_j \varepsilon_{t-j})^2 \} \right).$$

(10)

It may be noted that for Bayesian implementation, we do not require the exact likelihood, but rather the likelihood defined up to proportionality is sufficient. Moreover, the two-component AR(p) and MA(q) models can be obtained from the general ARMA(p,q) model just by ignoring the MA and AR terms, respectively. The likelihood for AR(p) model

can be obtained, up to proportionality, as

$$L_{AR}(y_1, y_2, ..., y_T | \theta_0, \Phi) \propto \left(\frac{1}{\sigma^2}\right)^{T/2} \times |V_\phi|^{-1/2} \times$$

$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p - \mu_p)' V_\phi^{-1}(Y_p - \mu_p) + \sum_{t=p+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}\phi_i y_{t-i})^2\}\right),$$

$$(11)$$

where V_ϕ denotes a $p \times p$ matrix involving the terms of ϕ's that may be determined by the relationship $\Sigma_p = \sigma^2 V_\phi$. Similarly, the likelihood function for MA(q) model can be obtained as the joint distribution of the observations and it can be obtained, up to proportionality, as

$$L_{MA}(y_1, y_2, ..., y_T | \theta_0, \Psi) \propto (\sigma^2)^{\frac{-T}{2}} \times |V_\psi|^{-1/2} \times \exp\left(-\frac{1}{2\sigma^2}(Y_T - \mu_T)' V_\psi^{-1}(Y_T - \mu_T)\right),$$

$$(12)$$

where V_ψ is obtained from the relation $\Sigma_T = \sigma^2 V_\psi$ (see also Appendix for other relevant details).

3.1. Priors

Besides the likelihood function, another important component in any Bayesian analysis is the specification of prior distribution. If one has enough *a priori* information, it is advisable to go with the informative prior. The prior distribution in that case will have a dominant role in posterior-based inferences. If enough details on *a priori* evidence is not available to go for informative prior, it is often suggested that one should instead use a vague prior or a weakly informative prior. It is to be noted that a weakly informative prior can be very well described by a proper prior density with large to very large scatteredness. Obviously, if the considered prior is weak, the posterior will be completely dominated by the likelihood function and the inferences can be said to depend exclusively on the data-based information. Such a consideration will of course remove the possibility of any negative impact that inappropriately chosen strong *a priori* belief could have had on the analysis. Keeping this in mind, we have considered the following non-informative priors similar to those proposed by Tripathi et al. (2017) and defined under stationarity and/or invertibility restrictions as detailed in Section 2. The considered priors are

$$\pi_1(\sigma^2) \propto \frac{1}{\sigma^2}; \qquad \sigma^2 \geq 0, \tag{13}$$

$$\pi_2(\theta_0) \propto U[-M, M]; \qquad M > 0, \tag{14}$$

$$\pi_3(\phi_i) \propto U[-N_1, N_1]; \qquad N_1 > 0, \qquad i = 1, 2, ..., p, \tag{15}$$

and

$$\pi_4(\psi_j) \propto U[-N_2, N_2]; \qquad N_2 > 0, \qquad j = 1, 2, ..., q, \tag{16}$$

where the constants M and N_i, $i = 1, 2$, used in the priors are the hyperparameters and $U[a, b]$ is the uniform distribution over the interval $[a, b]$. The prior distribution (13) is obviously a Jeffreys' type of prior for the scale parameter and it has often been suggested in the literature by a number of authors (see, for example, Marriot et al. (1996) and Kleinbergen and Hoek (2000)). The ranges of uniform distributions can, in general, be recommended large enough in order that the priors remain vague over the corresponding intervals. Moreover, since stationarity and invertibility conditions are essential requirements, it is necessary that we restrict our parameters ϕ_i and ψ_j in (15) and (16) to lie in the regions of stationarity and invertibility, that is, C_p and C_q, respectively, as defined in Section 2. We thus consider the priors for ϕ_i and ψ_j to be uniform distributions defined over the regions C_p and C_q. That is the value of the hyperparameters N_1 and N_2 are so chosen that the stationarity and invertibility conditions are satisfied. We must also keep in mind that the restrictions were calculated only for models satisfying the condition $p + q \leq 2$.

3.2. Posterior Distributions

Updating the prior distributions (13) to (16) with the likelihood (10) via Bayes theorem yields the joint posterior distribution for the parameters of an ARMA(p, q) model which, up to proportionality, can be written as

$$p(\theta_0, \Phi, \Psi, \sigma^2 | \underline{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1} \times |V_{\phi,\psi}|^{-1/2} \times$$

$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p - \mu_p)' V_{\phi,\psi}^{-1}(Y_p - \mu_p) + \sum_{t=p+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}\phi_i y_{t-i} - \sum_{j=1}^{q}\psi_j \varepsilon_{t-j})^2\}\right)$$

$$\times I_{[-M,M]}(\theta_0) \times \prod_{i=1}^{p} I_{[-N_1,N_1]}(\phi_i) \times \prod_{j=1}^{q} I_{[-N_2,N_2]}(\psi_j), \tag{17}$$

where $I_{[v_1,v_2]}(.)$ is the indicator function that takes value unity if (.) belongs to $[v_1, v_2]$ and zero otherwise.

Next, combining the prior distributions (13) to (15) with the likelihood (11) via Bayes theorem yields the joint posterior distribution for the parameters of an AR(p) model, which

can be written, up to proportionality, as

$$
p_{AR}(\theta_0, \Phi, \sigma^2 | \underline{y}) \propto \left(\frac{1}{\sigma^2} \right)^{\frac{T}{2}+1} \times |V_\phi|^{-1/2} \times
$$

$$
\exp \left(-\frac{1}{2\sigma^2} \{ (Y_p - \mu_p)' V_\phi^{-1} (Y_p - \mu_p) + \sum_{t=p+1}^{T} (y_t - \theta_0 - \sum_{i=1}^{p} \phi_i y_{t-i})^2 \} \right)
$$

$$
\times I_{[-M,M]}(\theta_0) \times \prod_{i=1}^{p} I_{[-N_1,N_1]}(\phi_i).
$$

$$(18)$$

Similarly, the posterior distribution for the parameters of a MA(q) model can be obtained up to proportionality as

$$
p_{MA}(\theta_0, \Psi, \sigma^2 | \underline{y}) \propto \left(\frac{1}{\sigma^2} \right)^{\frac{T}{2}+1} \times |V_\psi|^{-1/2} \times \exp \left(-\frac{1}{2\sigma^2} \{ (Y_T - \mu_T)' V_\psi^{-1} (Y_T - \mu_T) \} \right)
$$

$$
\times I_{[-M,M]}(\theta_0) \times \prod_{i=1}^{q} I_{[-N_2,N_2]}(\psi_i).
$$

$$(19)$$

Obviously, this posterior is the result of updating of prior distributions (13), (14) and (16) through the likelihood (12) via the Bayes theorem.

The posteriors (17), (18) and (19) are analytically intractable and cannot be obtained in nice closed forms. We, therefore, do not have many options except going for sample-based approaches and then drawing the corresponding inferences based on the simulated samples from the corresponding posteriors. In the next subsection, we shall discuss briefly the implementation of our proposed MCMC scheme and model estimation.

3.3. MCMC Implementation and Model Estimation

Markov chain Monte Carlo procedures basically construct a Markov chain such that simulating from its stationary distribution renders samples from the target posterior. The Gibbs sampler requires that the target posterior may be reduced into full conditionals corresponding to every variate. The algorithm progresses by simulating from, often unidimensional, full conditionals in a cyclic fashion. Since the implementation of Gibbs sampler requires simulation from various full conditionals, it is indispensable that all the full conditionals are routinely available for sample generation. We may check that this latter requirement may not be easily ensured for all the full conditionals through any standard sample generating schemes. We, therefore, make use of the Metropolis algorithm for such full conditionals although procedures are there where indirect strategies can be always developed. The Metropolis algorithm implemented separately on some of the full conditionals works by generating a probable variate value from a proposal density. The resulting variate value is

accepted if it has the large posterior probability. This is decided with the help of an acceptance probability. Choosing the mean and variance of the proposal density is a vital decision. The mean of the proposal is usually taken as the maximum likelihood (ML) estimate and its variance can be taken as the inverse of observed Fisher's information. Moreover, as far as the variance is concerned, it must be noted that if the variance is too large, some of the generated variate values will be quite far away from the current value leading to rejection. On the other hand, if the variance is too small, the chain will take more time to cover the entire support of the density with slightly low probability regions being under-sampled. Hence, a properly centered and dispersed kernel is highly essential. We often use a tuning constant 'c' to make this adjustment. We skip further discussions on the Gibbs sampler and the Metropolis algorithms for want of space and refer Gelfand and Smith (1991) and Upadhyay and Smith (1994) for details.

For the posterior corresponding to the ARMA(p,q) model given in (17), the full conditionals up to proportionality for different parameters can be written as

$$
p_1(\theta_0|\sigma^2,\Phi,\Psi,\underline{y}) \propto
$$
$$
\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_{\phi,\psi}^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\}\right),
$$
$$
(20)
$$

$$
p_2(\phi_i|\sigma^2,\theta_0,\Psi,\underline{y}) \propto |V_{\phi,\psi}|^{-1/2}\times
$$
$$
\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_{\phi,\psi}^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\}\right);
$$
$$
i=1,2,...,p, \quad (21)
$$

$$
p_3(\psi_j|\sigma^2,\theta_0,\Phi,\underline{y}) \propto |V_{\phi,\psi}|^{-1/2}\times
$$
$$
\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_{\phi,\psi}^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\}\right);
$$
$$
j=1,2,...,q, \quad (22)
$$

and

$$
p_4(\sigma^2|\theta_0,\Phi,\Psi,\underline{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1}\times
$$
$$
\exp\left(-\frac{1}{2\sigma^2}\{(Y_p-\mu_p)'V_{\phi,\psi}^{-1}(Y_p-\mu_p) + \sum_{t=p+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\}\right).
$$
$$
(23)
$$

On the other hand, if we consider the posterior (18) corresponding to a particular case of AR(p) model, the corresponding full conditionals are

$$p_5(\theta_0|\sigma^2,\Phi,\underline{y}) \propto$$

$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p - \mu_p)'V_\phi^{-1}(Y_p - \mu_p) + \sum_{t=p+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}\phi_i y_{t-i})^2\}\right), \quad (24)$$

$$p_6(\phi_i|\sigma^2,\theta_0,\underline{y}) \propto |V_\phi|^{-1/2} \times$$

$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p - \mu_p)'V_\phi^{-1}(Y_p - \mu_p) + \sum_{t=p+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}\phi_i y_{t-i})^2\}\right);$$

$$i = 1,2,...,p \quad (25)$$

and

$$p_7(\sigma^2|\theta_0,\Phi,\underline{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1} \times$$

$$\exp\left(-\frac{1}{2\sigma^2}\{(Y_p - \mu_p)'V_\phi^{-1}(Y_p - \mu_p) + \sum_{t=p+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}\phi_i y_{t-i})^2\}\right). \quad (26)$$

Similarly, MA(q) model results in the full conditionals up to proportionality obtained from the posterior (19) as

$$p_8\left(\theta_0|\sigma^2,\Psi,\underline{y}\right) \propto \exp\left(-\frac{1}{2\sigma^2}\{(Y_T - \mu_T)'V_\psi^{-1}(Y_T - \mu_T)\}\right), \quad (27)$$

$$p_9(\psi_j|\sigma^2,\theta_0,\underline{y}) \propto |V_\psi|^{-1/2} \times \exp\left(-\frac{1}{2\sigma^2}\{(Y_T - \mu_T)'V_\psi^{-1}(Y_T - \mu_T)\}\right);$$

$$j = 1,2,...,q \quad (28)$$

and

$$p_{10}(\sigma^2|\theta_0,\Psi,\underline{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1} \times \exp\left(-\frac{1}{2\sigma^2}\{(Y_T - \mu_T)'V_\psi^{-1}(Y_T - \mu_T)\}\right). \quad (29)$$

Obviously, we have a total of $(p+q+2)$ full conditionals corresponding to a general ARMA(p,q) model given by (1). Similarly, a total of $(p+2)$ full conditionals corresponding to a general AR(p) model and a total of $(q+2)$ full conditionals corresponding to a general MA(q) model are obtained. Among the various full conditionals, samples from (23), (26) and (29) corresponding to the posterior distributions of σ^2 can be obtained using a gamma generating routine after making the transformation $\tau = 1/\sigma^2$. It can be noted that the gamma

distribution, with shape and scale parameters $\alpha > 0$ and $\beta > 0$, respectively, for a random variable X is defined as;

$$g(x) = \frac{\beta^\alpha}{\Gamma\alpha} x^{\alpha-1} e^{-\beta x}; \ \ 0 < x < \infty, \tag{30}$$

where

$$\Gamma\alpha = \int_0^\infty e^{-x} x^{\alpha-1} dx.$$

Thus, it can be seen that the transformed variate τ follows a gamma density with shape parameter $\alpha = \frac{T}{2}$ and scale parameter $\beta = \frac{1}{2}\{(Y_p - \mu_p)'V_{\phi,\psi}^{-1}(Y_p - \mu_p) + \sum_{t=p+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}\phi_i y_{t-i} - \sum_{j=1}^{q}\psi_j\varepsilon_{t-j})^2\}$ in the case of ARMA(p,q) model. In the case of AR(p) model, τ follows a gamma density with shape parameter $\alpha = \frac{T}{2}$ and scale parameter $\beta = \frac{1}{2}\{(Y_p - \mu_p)'V_\phi^{-1}(Y_p - \mu_p) + \sum_{t=p+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}\phi_i y_{t-i})^2\}$. Similarly, in the case of MA(q) model, τ follows a gamma density with shape parameter $\alpha = \frac{T}{2}$ and the scale parameter $\beta = \frac{1}{2}\{(Y_T - \mu_T)'V_\psi^{-1}(Y_T - \mu_T)\}$.

The full conditionals (20) to (22), for each i and j, are not available from the point of view of sample generation although they may be managed easily by using the Metropolis algorithm. Our final scheme can, therefore, be a kind of hybrid (Metropolis within Gibbs) because of these full conditionals and the same can be referred to as the Gibbs-Metropolis hybrid sampler. Similar hybrid schemes can be thought of for generating samples from (24) to (25), for each i, corresponding to AR(p) model and from (27) to (28), for each j, corresponding to MA(q) model. For Metropolis implementation separately on each of these full conditionals, one can use univariate normal candidate generating density with mean taken as the current realization and standard deviation to be approximated on the basis of the particular element of the Hessian matrix obtained at ML estimates. It is important to mention that although the candidate generating density is univariate and the corresponding full conditionals are each univariate, the ML estimates and the corresponding Hessian-based approximations are obtained for a multi-parameter likelihood function. Moreover, as mentioned earlier, the value of the standard deviation so obtained is adjusted by multiplying with a scaling constant c, taken in the range 0.5 and 1.0 (see, for example, Upadhyay and Smith (1994)) for reducing the number of rejections in the Metropolis step. The other important thing in the implementation of the Metropolis algorithm is the choice of initial values for running the chain. One can, of course, use any standard classical estimates to begin running the chain. We have used the ML estimates in particular. For relevant details on the Metropolis algorithm and issues on its convergence diagnostics, one can refer to Smith and Roberts (1993) and Upadhyay et al. (2001), among others.

Thus, the Gibbs-Metropolis hybrid sampler strategy can be easily implemented on the posteriors given in (17), (18) and (19) corresponding to ARMA, AR and MA models, respectively. The Gibbs-Metropolis hybrid strategy is being referred to simply because some of the full conditionals do not ascribe to any standard form of distributions and, as such, the generations are difficult. We, therefore, implement intermediate Metropolis steps for

generating variate values from the same and base our decision on a single long run of the chain. Obviously, the implementation of the algorithm results in a long sequence of iterating chains. After sufficiently large number of iterations, the generating sequence converges in distribution to random samples from the target joint posterior and the generated components to the corresponding marginal posteriors (see, for example, Smith and Roberts (1993)). In order to get the final samples, one has to discard the initial burn-in samples and then pick up the variate values at equidistant intervals to minimize serial correlation among the generating variates (see, for example, Upadhyay et al. (2012)). These final samples can be used in a variety of ways to draw the desired posterior-based inferences. Say, for instance, one can use sample-based posterior estimates for the model parameters or one can estimate the entire posterior densities by means of some nonparametric density estimates. Other desired features of the posteriors can also be likewise obtained once the posterior samples are made available (see also Upadhyay et al. (2012)).

3.4. Prediction

As mentioned, an important part of our analysis includes predicting the unobserved future data y_{T+1} given the informative data $\underline{y} = y_1, y_2, ..., y_T$. Now, one can easily confirm from (1) that for the given set of observations $\underline{y}$, the future observation y_{T+1} follows normal distribution with mean

$$\mu_{T+1} = \theta_0 + \sum_{i=1}^{p} \phi_i y_{t+1-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t+1-j} \tag{31}$$

and variance σ^2. Thus, the future observation y_{T+1} can be easily simulated from this normal distribution after replacing the corresponding parameters by their appropriately chosen posterior estimates (say, for example, estimated posterior modes) obtained on the basis of final posterior simulated samples. The ε_t's in (1) can be simulated from the normal density with mean zero and variance equal to the corresponding posterior estimate of σ^2. This strategy can be easily applied to get the predictive samples of y_{T+1} from which any desired sample-based predictive characteristic can be assessed. It is essential to mention here that the posterior mode is the highest probable value and, therefore, it makes sense if the resulting posterior distribution is non-symmetric. On the other hand, if the resulting posterior distribution is symmetric, it is immaterial whether one uses mean or median or mode. Moreover, in the case of AR and MA models, one can simply ignore the MA and AR components, respectively, in the general form of the ARMA model (1) and proceed to obtain the corresponding predictive samples. It may, however, be noted that one requires the posterior samples corresponding to AR and MA models.

4. Model Compatibility

A model compatibility study can be performed using a variety of tools. An important one among these may utilize the idea of predictive simulation where the predictive observations are obtained from the model under consideration and then compared with the observed

data $\underline{y}$. Obviously, a model may be considered compatible with the observed data if its predictive output compliances with the former. In the case of no or poor resemblance of the two data sets, the considered model creates a suspicion and cannot be recommended for the data in hand. Among other things, this resemblance can often be measured by means of a quantitative summary in the form of p-value that can be obtained using an appropriate model discriminating statistic. Informally, graphical tools are also suggested in the literature where some characteristics based on the two data sets may be shown on the same graphical scale (see, for example, Gelman et al. (1996), Bayarri and Berger (1998), Upadhyay and Peshwani (2003), etc.).

The present study, however, begins with an informal approach where simple time series plots for the observed and the predicted data from the model(s) are shown graphically in a way that the plots corresponding to the latter are superimposed over that corresponding to the former. For p-value based study of model compatibility, we advocate for a Bayesian version of the same based on an appropriately chosen statistic. Obviously, if the calculated p-value happens to be large, the considered model may be regarded compatible with the data.

The Ljung-Box-Pierce statistic that follows a chi-square distribution if the null hypothesis is true, is commonly used in autoregressive integrated moving average (ARIMA) modelling to test the overall randomness behaviour based on a number of lags. It may be noted that the test is applied on the residuals of a fitted ARIMA model and not on the original series. Thus, the hypothesis actually being tested in this case is that the residuals from the ARIMA model have no autocorrelation and that the model is adequate. When testing the residuals of an estimated ARIMA model, the degrees of freedom need to be adjusted to reflect the parameter estimation. For example, for an ARIMA$(p, 0, q)$ or ARMA(p, q) model, the degrees of freedom should be set to $(l - p - q)$, where l is the number of lags, that is, the order of autocorrelation being tested and the correction due to p and q is because of the fact that the degrees of freedom must account for the estimated model parameters. The statistic, used to test for uncorrelated residuals, is then calculated by

$$Q(l) = T(T+2) \sum_{j=1}^{l} \frac{r_j^2(\hat{\varepsilon}_t)}{T - j}, \tag{32}$$

where

$$r_j^2(\hat{\varepsilon}_t) = \frac{\sum_{t=1}^{T-j} \hat{\varepsilon}_t \hat{\varepsilon}_{t+j}}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}, \tag{33}$$

$\hat{\varepsilon}_t = y_t - \hat{y}_t$ and the statistic $Q(l)$ follows chi-square distribution with $(l - p - q)$ degrees of freedom. Here $\hat{y}_t$ is the predicted value of y_t that can be obtained from (1) in a way described in subsection 3.4, T is the number of residuals computed for the model and $\hat{\varepsilon}_t$ is the residual at time t. This statistic was advocated by Ljung and Box (1978) and is often referred to as the Ljung-Box statistic or the Ljung-Box-Pierce statistic. The Bayesian p-value can then be calculated by the probability P^{post} (based on the posterior samples) of acceptance of the

null hypothesis and it is given by

$$pval = P^{post}[Q(l) < \chi^2_{(l-p-q),(1-\alpha)}], \tag{34}$$

where $\chi^2_{(l-p-q),(1-\alpha)}$ is the tabulated value of chi-square at $(l-p-q)$ degrees of freedom and the level of significance α. The model is adequate if this probability is large, that is, at least larger than the assumed significance level (usually 0.05). There is, however, very little practical advice on how to choose the number of lags for the test. A recommendation based on power considerations is to consider $l = 10$ for non-seasonal data and $l = 2k$ for seasonal data where k is the period of seasonality (see, for example, Hyndman and Athanasopoulos (2018)). Needless to mention that we want to ensure that l is large enough to capture any meaningful and troublesome correlations. So for our data set, we prefer to choose $l = 10$ to increase the power of the test.

4.1. Model Comparison: A PPL Approach

After the compatibility of a model is ascertained, we go for its comparison with other compatible models. Model comparison is intuitively based on two major criteria - its fitting to the observed data and its inherent complexity. Most of the model comparison tools, Bayesian or otherwise, are based on a weighted trade-off between the two criteria, the weights being decided according to some specific needs. One such criterion, known as PPL criterion, was initially given by Gelfand and Ghosh (1998). Based on predictive simulation, this criterion parallels to standard utility ideas and partitions the total loss into loss due to fit and loss due to complexity (see also Upadhyay and Mukherjee (2008)). A simplified version of this criterion was given by Sahu and Dey (2000) (see also Upadhyay et al. (2012)). This criterion recommends a model m that minimizes the joint effect of two measures, namely, the closeness of observed and predictive data sets and variability of the prediction. The criterion can be defined as

$$D(m) = G(m) + P(m) \tag{35}$$

where $G(m) = \sum_{t=1}^{T} (\mu_t^{(m)} - y_t)^2$, $P(m) = \sum_{t=1}^{T} \sigma_t^{2(m)}$, $\mu_t^{(m)} = E(z_t|y_t, m)$, $\sigma_t^{2(m)} = \text{Var}(z_t|y_t, m)$ and z_t denotes the tth component of predictive data, $t = 1, 2, 3, ..., T$. Obviously, the term $\mu_t^{(m)}$ represents the predictive mean and the term $\sigma_t^{2(m)}$ represents the predictive variance of tth component under the model m.

In (35), $G(m)$ represents the goodness of fit term and it increases when the entertained model provides poor fitting at the observed data points. Similarly, the second term $P(m)$ represents the penalty term and it increases with the increasing complexity in the model. A model m that provides least value of $D(m)$ when compared with all other models, is finally recommended. We are not going into details of its formulation rather refer to Sahu and Dey (2000) (see also Upadhyay et al. (2012) and Tripathi et al. (2017)).

5. Real Data Illustration

For numerical illustration, we consider a real data set on GDP growth rate of India at constant prices. The data set given in Table 5 (see Appendix) is collected over a period of sixty three years, 1951-52 to 2013-14, and is taken from the publication of Central Statistical Organization (CSO) (2014) (see `http://planningcommission.nic.in/data/datatable/0814/comp_databook.pdf`). This data set has been used by a number of authors, a recent reference being Tripathi et al. (2017).
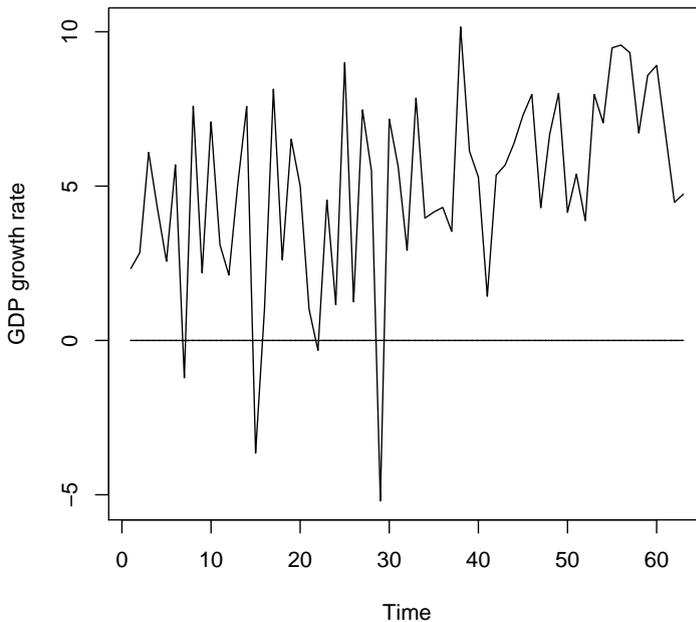
Figure 1: Time series plot showing the GDP growth rate of India (straight line shows the mean level).

Before we begin the intended numerical illustration with GDP growth rate data, let us draw the simple time series plot corresponding to the given GDP observations to see if there is stationarity behaviour in the series. The corresponding time series plot is shown in Figure 1. It can be seen that the series exhibits non-stationarity behaviour indicated by its growth (see Figure 1) and, therefore, it is essential to perform an appropriate transformation (see also Clement (2014)) on the data to remove its non-stationarity behaviour. To resolve this issue, we took first difference of the data and the corresponding time series plot for the transformed data, as shown in Figure 2. Obviously, the figure shows stationarity behaviour. Some outliers at intermediate stages can also be seen, which cause some abrupt hikes in

the first half of the series. Since the study of outliers is beyond the scope of our work, we continue our study up to this level of stationarity only. We further strengthen our claim using some numerical evidences. We assess stationarity of data (or its absence) using the ADF and KPSS tests, first on non-differenced data and then on differenced data. The two tests for stationarity are found to be significant at 5% level and, therefore, provide enough evidence against stationarity of actual data. The p-value in the ADF test is found to be 0.1 and that in the KPSS test is 0.02, which evidently refuse the presence of stationarity in the data. Moreover, after differencing the data, the p-values are found to be 0.01 for the ADF test and 0.1 for the KPSS test, which now ensure the stationarity in the data. Hence, this transformed data can be used for the final analysis.
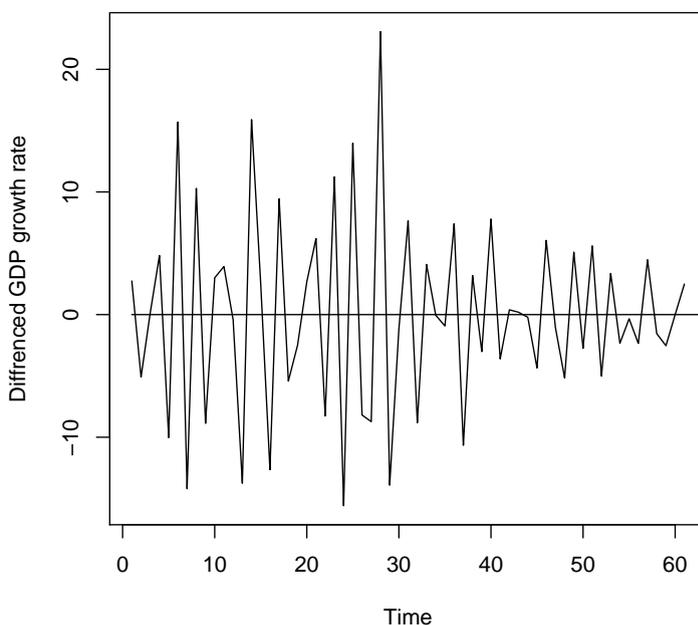
Figure 2: Time series plot corresponding to the first difference of GDP growth rate data (straight line shows the mean level).

Our analysis of ARMA(p,q) was restricted to $p+q \leq 2$ and, therefore, we considered five different ARMA modelling combinations by taking (p,q) as (0, 1), (0, 2), (1, 0), (1, 1) and (2, 0), respectively. These restrictions on the values of p and q were imposed for the ease of performing the analysis and also for the ease of implementation of invertibility conditions as mentioned in Section 2. It is to be noted that for higher values of p and q, as mentioned in Section 2, the invertibility conditions are not available in analytically closed forms and the situation might be difficult to ensure these numerically too.

To perform the complete posterior analysis, we first obtained the ML estimates of the corresponding parameters of the considered models by maximizing the log likelihood function in each case. These ML estimates were utilized as the initial values for the necessary MCMC implementation. The complete posterior analysis was done for each considered model as discussed in Section 3. In order to nullify the prior effect and hence to draw exclusively data dependent inferences, the values of prior hyperparameters M and N_i, $i = 1, 2$, were chosen to be 100 in each case. The priors for AR and/or MA coefficients were chosen under the restrictions of stationarity and invertibility as discussed in Section 2 and subsection 3.1.

Under the prior assumptions as mentioned above, we analyzed the posteriors (17), (18) and (19) using the Gibbs-Metropolis hybrid scheme as discussed in subsection 3.3. It is important to mention that for the full conditional distributions corresponding to the intercept, the AR coefficients and/or the MA coefficients, the corresponding generating algorithm was Metropolis with properly centred and dispersed normal kernel as the proposal density. The mean value of the kernel was approximated by the ML estimate of the corresponding parameter (see Table 1). As far as the standard deviation is concerned, the exact value was difficult to obtain either analytically or numerically. We, therefore, considered its numerical approximation by evaluating the second derivative numerically at the corresponding ML estimate. The need for some adjustments occurred probably because of the approximations that we considered. This adjustment was carried out with the help of a scaling constant, $c = 0.6$, that provided a good acceptance probability in each case.

To get the posterior sample by implementing the Gibbs-Metropolis hybrid sampler, we considered a single long run of the chain. After an initial transient behaviour, convergence of the chain, based on ergodic averages, was observed at about 40K iterations. This cannot be considered as a deterrent issue considering the development in high speed computing that took approximately 3.1423 minutes in 40K iterations. Thus, it can be said that the algorithm works reasonably well. After successfully monitoring the convergence, we picked up equally spaced generated values to form the samples from the corresponding posteriors (see also Upadhyay et al. (2001)). These gaps were taken to be 20 to make the serial correlation negligibly small.

Some of the important estimated posterior characteristics based on the posterior samples of size 1K from each model are shown in Table 1. These estimates are shown in the form of estimated posterior means, medians, modes and 0.95 highest posterior density (HPD) intervals (HPD interval with coverage probability 0.95). Besides, we have also shown the ML estimates of different parameters for each of the considered models, which were actually obtained for defining the initial values for running the proposed Gibbs-Metropolis hybrid algorithm but can also be used for the purpose of comparison with the corresponding Bayes estimates. It can be seen that the ML estimates are, in general, close enough to the corresponding Bayes estimates, a conclusion that is expected for the considered modelling combination. From the posterior estimates shown in Table 1, one can draw several conclusions, the one among these may be to get an overall idea of various estimated posterior densities. It can be seen that the estimated marginal posterior of σ^2 reveals slight positive skewness. The corresponding estimates for other parameters, however, exhibit almost symmetrical behaviour for their posterior densities except for the parameter θ_0 for the first two

models, which reveal negatively skewed posterior densities. It may also be noted that the degree of skewness, in general, is less for ARMA$(0,1)$ model as compared to all other models. As a final remark, it can be said that these estimates are obtained under the restrictions of stationarity and invertibility and, as mentioned earlier, the results are reasonable from that point of view as well. One can also see the significance of preceding observations in our reported results. It can be seen that the estimates of ϕ's and ψ's are, in general, appreciably larger than the corresponding estimated values for the intercepts. This of course advocates the usability of the considered model in some sense. Moreover, the estimates of σ^2's also convey an important message that increasing complexity in the model, in general, decreases the variability inherent in the model.

Table 1. Posterior summaries for different variates of considered ARMA models based on first difference data

Model	Parameter	ML Estimates	Posterior Means	Posterior Medians	Posterior Modes	0.95 HPD Interval	
ARMA$(1,0)$	θ_0	0.054	0.039	0.057	0.099	-0.829	1.084
	ϕ_1	-0.558	-0.557	-0.555	-0.550	-0.770	-0.346
	σ^2	13.079	14.076	13.777	13.667	9.297	19.491
ARMA$(2,0)$	θ_0	0.065	0.039	0.059	0.124	-0.701	0.924
	ϕ_1	-0.808	-0.791	-0.796	-0.794	-0.981	-0.603
	ϕ_2	-0.435	-0.421	-0.424	-0.444	-0.654	-0.218
	σ^2	10.501	11.374	11.005	10.563	7.178	15.689
ARMA$(0,1)$	θ_0	0.071	0.064	0.067	0.068	-0.043	0.201
	ψ_1	-0.900	-0.884	-0.896	-0.913	-0.998	-0.744
	σ^2	8.015	9.197	9.023	8.779	6.011	12.838
ARMA$(0,2)$	θ_0	0.073	0.053	0.059	0.065	-0.093	0.228
	ψ_1	-1.230	-0.903	-0.925	-0.969	-0.999	-0.743
	ψ_2	0.230	0.074	0.072	0.066	-0.106	0.250
	σ^2	7.595	9.482	9.253	8.745	6.051	13.175
ARMA$(1,1)$	θ_0	0.072	0.083	0.086	0.087	-0.044	0.183
	ϕ_1	-0.195	-0.226	-0.228	-0.251	-0.476	0.019
	ψ_1	-0.990	-0.901	-0.916	-0.970	-0.999	-0.738
	σ^2	7.665	8.738	8.533	8.068	5.876	12.291

We next consider the issue of examining compatibility of the assumed models for the given time series data. Since the stationarity behaviour is established on the basis of first difference data, we shall consider the same set of observations for examining compatibility. Our study is based on some graphical as well as quantitative summaries. Graphically, we have studied it by plotting the time series of observed first difference data along with the corresponding predictive differenced data. For this purpose, 10 predictive samples were generated from each of the considered models using the final posterior samples of size 10 generated using the Gibbs-Metropolis hybrid sampler algorithm (see subsection 3.3). Thus, each posterior sample resulted in one predictive sample of size equal to that of the observed data. We next considered obtaining the first difference from each of the predictive samples. Now, the differenced form of 10 predictive samples are plotted along with the corresponding observed (differenced) data in the form of time series. One such plot corresponding to ARMA$(0,1)$ model is shown in Figure 3, where the bold line corresponds to first difference of observed time series data. The time series plots corresponding to the first differenced predictive samples are shown by means of dotted lines. It can be seen that the predictive sample plots and the observed data plot exhibit more or less similar overlapping behaviour and, therefore, the ARMA$(0,1)$ model can be considered compatible for the observed first

differenced data. We had a similar message from the plots corresponding to all other models although the plots are not shown due to space restriction. Thus, all the models can be regarded compatible with the observed time series data.
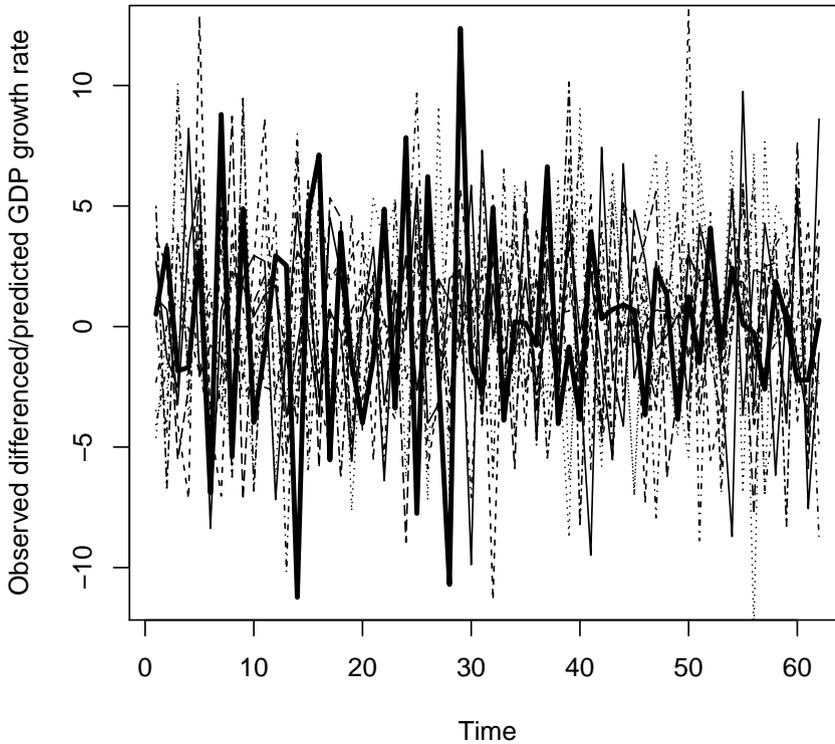
Figure 3: Time series plots for the first difference of observed and predictive data sets from ARMA(0, 1) model (the bold line corresponds to observed data).

For the study of model compatibility based on quantitative evidence, we considered evaluating the Bayesian p-value based on Ljung-Box-Pierce chi-square statistic (see Section 4). It is to be noted that the values of residuals can be obtained once the predictive observations corresponding to each of the original observations are made available. To clarify the computational stages, we first simulated 1K posterior samples as discussed in Section 3.3 and then obtained 1K predictive samples, each predictive sample of size equal to that of the observed time series data. Based on these simulated samples, we can have 1K samples of residuals where each sample of residuals is of size equal to that of the observed data. Finally, each set of residuals is used to get the predicted value of $Q(l)$ by substituting the values of residuals in (32). Hence, we calculate a total of 1K values of $Q(l)$. These values can then be used to

obtain the estimated p-values (see (34)) by counting the number of values of $Q(l)$ which are less than the tabulated value of $\chi^2_{(l-p-q),(1-\alpha)}$ and calculating the corresponding fraction. The p-values have been calculated for the considered ARMA models and are presented in Table 2.

Table 2. p-values for the considered ARMA models

Model	*pval*
ARMA$(1,0)$	0.758
ARMA$(2,0)$	0.723
ARMA$(0,1)$	0.739
ARMA$(0,2)$	0.743
ARMA$(1,1)$	0.716

Obviously, the values are large enough to support the adequacy of the models. Thus, our model compatibility study conveys that each of the considered models may be a good candidate to describe the data in hand. One can, of course, use parsimony principle and recommend a model that happens to be the simplest among these but we shall compare these models using the PPL criterion described in subsection 4.1 before recommending a model.

The values of $P(m)$, $G(m)$ and hence $D(m)$ for all the considered models are given in Table 3. These values are based on 1K posterior and correspondingly 1K predictive samples from each of the considered models. It can be seen that the value of $D(m)$ is least for ARMA$(0,1)$ model mainly because it has the smallest value of loss due to complexity although it provides poor fitting when compared with ARMA$(1,0)$, ARMA$(2,0)$, ARMA$(0,2)$ and ARMA$(1,1)$ models (see Table 3). Thus, our final recommended model is ARMA$(0,1)$, a model that has no autoregressive component.

Table 3. Results based on PPL criterion for the considered ARMA models

Model	$P(m)$	$G(m)$	$D(m)$
ARMA$(1,0)$	1345.309	1206.245	2551.554
ARMA$(2,0)$	1280.723	1168.402	2449.125
ARMA$(0,1)$	1022.513	1194.609	**2217.123**
ARMA$(0,2)$	1045.388	1182.324	2227.711
ARMA$(1,1)$	1188.329	1185.996	2374.325

Before we end the section, let us consider the problem of predicting the future observations based on the finally selected ARMA$(0,1)$ model. We, however, confine ourselves to short term retrospective prediction so that the scope of predicting the GDP values through the considered model can be verified based on the comparison of predicted values with the observed data points. To proceed with the task of retrospective prediction, let us consider the first 55 observations as the informative data out of a total of 63 entertained observations (see Table 5) and obtain the predictive estimates for the next 56th observation. We may then include this predicted observation in the informative data and proceed with 56 informative observations to develop the prediction for 57th observation. The process may be continued until all the 63 observations are predicted.

To explain the implementation, let us consider the informative data size as r. Thus, we begin with the complete posterior analysis based on these r observations before going for the actual prediction. The details of running the hybrid strategy on the posterior corresponding to ARMA$(0, 1)$ model and hence obtaining the posterior and corresponding predictive samples are provided in subsections 3.3 and 3.4. The results are shown in terms of point prediction as well as the corresponding predictive interval in Table 4. These results are based on a posterior sample of size 1K and corresponding predictive observation for the next unobserved future data $((r + 1)^{th})$ obtained for each value of the simulated posterior sample. The predictive estimates are then given as the corresponding predictive modes based on such 1K predictive samples. Similarly, the predictive intervals correspond to highest predictive density intervals each with coverage probability 0.95 obtained on the basis of such 1K predictive samples.

Table 4. The retrospective predictions of GDP growth rate data for the period 2006-07 to 2013-14

Year	y_t	True value	Predictive point estimate	0.95 predictive interval	
2006-07	y_{56}	9.57	10.694	1.133	17.665
2007-08	y_{57}	9.32	9.844	0.967	16.924
2008-09	y_{58}	6.72	9.051	1.105	17.099
2009-10	y_{59}	8.59	8.915	1.427	17.770
2010-11	y_{60}	8.91	8.092	0.551	16.684
2011-12	y_{61}	6.69	9.194	1.015	16.721
2012-13	y_{62}	4.47	10.358	0.993	16.773
2013-14	y_{63}	4.74	9.633	2.714	18.467

It can be seen from the results that the predictive point estimates in the form of modal values are, in general, not too far away from the actual observed data points except for the situations where there is a high fluctuation in the values from those in the previous years. Say, for instance, the values corresponding to the years 2011-12 (y_{61}) to 2013-14 (y_{63}) where the predictive point estimates are too far away from the actual values. It might be possible that there are structural breaks in the GDP data for these years perhaps because of global economic recession and, as such, our model fails to reflect the same. The situation is, however, not too susceptible when we see the estimated predictive intervals with coverage probability 0.95. All these estimated intervals not only cover the true values but also indicate that the true values fall in the high probability central regions of the corresponding predictive density estimates. Although such predictive density estimates are not shown, an idea can be derived based on the values of estimated predictive intervals. As a word of final remark- in no way we claim that our model is most appropriate for the situation rather it appears as if there is always a scope for its improvement. The simplicity of our model and its analysis are certainly the important features in its favour.

6. Conclusions

The paper emphasizes the analysis of a general ARMA model along with its components in a fully Bayesian framework, although a few classical tools, such as evaluation of the ML estimates and the observed information, are employed to support our analysis. The analysis finally proceeds for five particular cases of the considered general form under stationarity and invertibility restrictions. A sample-based approach based on the Gibbs-Metropolis hybrid sampler appears to provide routine posterior implementation on the considered ARMA model and its particular forms. The paper then considers model compatibility and comparison, the former using predictive simulation ideas and the latter using predictive loss criterion. A real data illustration of GDP growth rate data of India at constant prices conveys that ARMA$(0, 1)$ model appears to be the most appropriate, although other components also provide good compatibility with the data in hand. A short-term retrospective prediction based on the final chosen model conveys that the proposed model can be used, in general, except when there is abrupt fluctuation in the data from those of previous years.

Acknowledgements

References

Bayarri, M. J., Berger, J. O., (1998). Quantifying Surprise in the Data and Model Verification, In *Bayesian Statistics*, 6, ed. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, pp. 53–82. London: Oxford University Press.

Box, G. E. P., Jenkins, G. M., (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., (2004). *Time Series Analysis: Forecasting and Control*. Second Indian Reprint, Pearson Education.

Chib, S., Greenberg, E., (1994). Bayes Inference in Regression Models with ARMA(p, q) Errors. *Journal of Econometrics*, 64, pp. 183–206.

Clement, E. P., (2014). Using Normalized Bayesian Information Criterion (BIC) to Improve Box-Jenkins Model Building. *American Journal of Mathematics and Statistics*, 4(5), pp. 214–221.

Dickey, D. A., Fuller, W. A., (1979). Distribution of the Estimators for the Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, pp. 427–431.

Fan, C., Yao. S., (2008). Bayesian Approach for ARMA Process and Its Application. *International Business Research*, 1(4), pp. 49–55.

Gelfand, A. E., Smith, A. F. M., (1991). Gibbs Sampling for Marginal Posterior Expectations. *Communications in Statistics, Theory and Methods*, 20(5–6), pp. 1747–1766.

Gelfand, A. E., Ghosh, S. K., (1998). Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika*, 85, pp. 1–11.

Gelman, A., Meng, X. L., Stern, H. S., (1996). Posterior Predictive Assessment of Model Fitness via Realized Discrepancies. *Statistica Sinica*, 6, pp. 733–807.

Hyndman, R. J., Athanasopoulos, G., (2018). *Forecasting: principles and practice.* OTexts.

Kleibergen, F., Hoek, H., (2000). Bayesian Analysis of ARMA Models. *Tinbergen Institute Discussion Paper* Amsterdam, the Netherlands.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, P., (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. *Journal of Econometrics*, 54, pp. 159–178.

Ljung, G. M., Box, G. E. P., (1978). On A Measure of Lack of Fit in Time Series Models. *Biometrika*, 65, 297–303.

Ludlow, J., Enders, W., (2000). Estimating non-linear ARMA models using Fourier coefficients. *International Journal of Forecasting*, 16(3), pp. 333–347.

Morley, J. C., Nelson, C. R., Zivot, E., (2003). Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different? *Review of Economics and Statistics*, 85(2), pp. 235–243.

Marriot, J. M., Smith A. F. M., (1992). Reparameterization Aspects of Numerical Bayesian Methodology for Autoregressive Moving Average Models. *Journal of Time Series Analysis*, 13(4), pp. 327–343.

Marriot, J. M., Ravishanker, N., Gelfand, A., PAI, J., (1996). Bayesian Analysis of ARMA Processes: Complete Sampling Based Inference Under Exact Likelihoods. In: Berry, D. A., K. M. Chaloner, and J. Geweke (eds.): *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner* John Wiley, New York.

Monahan, J. F., (1983). Fully Bayesian Analysis of ARMA Time Series Models. *Journal of Econometrics*, 21, pp. 307–331.

Newbold, P., (1974). The Exact Likelihood Function for a Mixed Autoregressive-Moving average Process. *Biometrika*, 61(3), pp. 423–426.

Okereke, O. E., Iwueze, I. S., Johnson, O., (2015). Necessary Conditions for the Application of Moving Average Process of Order Three. *Applied Mathematics*, 6(01), pp. 173–181.

Pankratz, A., (1983). *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases.* John Wiley & Sons, New York.

Reilly, D. P., (1980). Experiences with An Automatic Box-Jenkins Modeling Algorithm. *Time Series Analysis, Proceedings of Houston Meeting on Time Series Analysis*, North-Holland publishing, Amsterdam.

Reynolds, S. B., Mellichamp, J. M., Smith, R. E., (1995). Box-Jenkins Forecast Model Identification. *AI Expert* June, pp. 15–28.

Sahu, S. K., Dey D. K., (2000). A Comparison of Frailty and Other Models for Bivariate Survival Data. *Lifetime Data Analysis*, 6, pp. 207–228.

Shumway, R. H., Stoffer, D. S., (2011). *Time Series Analysis and Its Applications*, Springer-Verlag, New York.

Smith, A. F. M., Roberts, G. O., (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, Series B*, 55, pp. 3–25.

Tripathi, P. K., Ranjan, R., Pant, R., Upadhyay, S. K., (2017). An approximate Bayes analysis of ARMA model for Indian GDP growth rate data. *Journal of Statistics and Management Systems*, 20(3), pp. 399–419.

Tripathi, P. K., Mishra, R. K., Upadhyay, S. K., (2018). Bayes and Classical Prediction of Total Fertility Rate of India Using Autoregressive Integrated Moving Average Model. *Journal of Statistics Applications & Probability*, 7(2), pp. 233–244.

Tsay, R. S., Tiao, G. C., (1984). Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Non-Stationary ARMA Models. *Journal of the American Statistical Association*, 79, pp. 84–96.

Tsay, R. S., Tiao, G. C., (1985). Use of Canonical Analysis in Time Series Model Identification. *Biometrika*, 72(2), pp. 299–315.

Upadhyay, S. K., Gupta, A., Dey, D. K., (2012). Bayesian Modeling of Bathtub Shaped Hazard Rate Using Various Weibull Extensions and Related Issues of Model Selection. *Sankhya, Series B*, 74, pp. 15–43.

Upadhyay, S. K., Mukherjee, B., (2008). Assessing the Value of the Threshold Parameter in the Weibull Distribution Using Bayes Paradigm. *IEEE Transactions on Reliability*, 57(3), pp. 489–497.

Upadhyay, S. K., Peshwani, M., (2003). Choice Between Weibull and Lognormal Models: A Simulation Based Bayesian Study. *Communications in Statistics, Theory and Methods*, 32, pp. 381–405.

Upadhyay, S. K., Smith, A. F. M., (1994). Modelling Complexities in Reliability and the Role of Simulation in Bayesian Computation. *International Journal of Continuing Engineering Education*, 4, pp. 93–104.

Upadhyay, S. K., Vasishta, N., Smith, A .F. M., (2001). Bayes Inference in Life Testing and Reliability via Markov Chain Monte Carlo Simulation. *Sankhya, Series A*, 63, pp. 15–40.

Zellner, A., (1971). *An Introduction to Bayesian Inference in Econometrics.* John Wiley & Sons, New York.

Appendix

Table 5. GDP growth rate of India at constant prices for the period 1951-52 to 2013-14 (from left to right)

2.33	2.84	6.09	4.25	2.56	5.69	-1.21	7.59	2.19
7.08	3.10	2.12	5.06	7.58	-3.65	1.02	8.14	2.61
6.52	5.01	1.01	-0.32	4.55	1.16	9.00	1.25	7.47
5.5	-5.2	7.17	5.63	2.92	7.85	3.96	4.16	4.31
3.53	10.16	6.13	5.29	1.43	5.36	5.68	6.39	7.29
7.97	4.30	6.68	8.00	4.15	5.39	3.88	7.97	7.05
9.48	9.57	9.32	6.72	8.59	8.91	6.69	4.47	4.74

Likelihood for ARMA(p, q) model:

Let us consider the likelihood of ARMA(p, q) model as given in (9) and let $Y_p = (y_1, y_2, ..., y_p)'_{p \times 1}$ be the vector of first p observations in the sample of size T. Then Y_p follows a p-variate normal distribution with mean vector $\mu_p = (\mu, \mu, ..., \mu)'_{p \times 1}$ of dimension $p \times 1$ and variance-covariance matrix Σ_p of dimension $p \times p$. It may be noted that $\mu = \theta_0 / (1 - \phi_1 - \phi_2 - ... - \phi_p)$ and the elements of Σ_p can be obtained by solving the Yule-Walker equations of autocorrelation in terms of AR parameters (see, for example, Box et al. (2004)). Thus, the joint distribution of Y_p can be written as

$$f(y_1, y_2, ..., y_p | \theta_0, \Phi, \Psi) \propto (\sigma^2)^{\frac{-p}{2}} \times |V_{\phi, \psi}|^{-1/2} \times$$
$$\exp\left(-\frac{1}{2\sigma^2}(Y_p - \mu_p)' V_{\phi, \psi}^{-1}(Y_p - \mu_p)\right), \quad (36)$$

where $V_{\phi, \psi}$ is a $p \times p$ matrix involving terms of ϕ's and ψ's and it may be determined by using the relationship $\Sigma_p = \sigma^2 V_{\phi, \psi}$.

The second term on the right-hand side of (9) is the product of conditional densities of y_t with conditioning variates $(y_{t-1}, y_{t-2}, ..., y_{t-p})$, $t = (p+1), ..., T$. This conditional density can be shown to follow $N((\theta_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j}), \sigma^2)$ and the same can be written as,

$$f(y_t | y_{t-1}, y_{t-2}, ..., y_{t-p}; \theta_0, \Phi, \Psi) \propto \left(\frac{1}{\sigma^2}\right) \times$$
$$\exp\left(-\frac{1}{2\sigma^2}(y_t - \theta_0 - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{q} \psi_j \varepsilon_{t-j})^2\right). \quad (37)$$

Obviously, the conditional likelihood function (the second term in the right-hand side of (9)) reduces to,

$$f_c \propto \left(\frac{1}{\sigma^2} \right)^{(T-p)/2} \times \exp \left(-\frac{1}{2\sigma^2} \sum_{t=p+1}^{T} (y_t - \theta_0 - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{q} \psi_j \varepsilon_{t-j})^2 \right). \qquad (38)$$

Now, combining (9) and (38), the likelihood for ARMA(p,q) model in (9) can be written as in (10).

The ARMA(p,q) model is generally difficult to implement in practice due to the fact that stationarity and invertibility conditions are not easily encountered for higher values of p and q beyond, say, 2 in each case. This is perhaps the reason that we restrict to $p+q \leq 2$ in our present study. Moreover, for ARMA$(1,1)$ model, the things are comparatively easier. In this case, we have $Y_1 = y_1 \sim N(\mu_1, \Sigma_1)$, where $\mu_1 = \frac{\theta_0}{1-\phi_1}$ and $\Sigma_1 = \frac{\sigma^2(1+\psi_1^2+2\psi_1\phi_1)}{1-\phi_1^2}$.

Exact likelihood for AR(p) model:

The AR(p) model corresponds to ARMA$(p,0)$ and, therefore, the corresponding likelihood function can be obtained by putting zero in the place of MA component q in the likelihood (10) and the same can be written as in (11).

Moreover, as mentioned earlier for ARMA process, the AR(p) process is also not easy to implement in practice for values of p beyond 2. For AR(1) model, we have $Y_1 = y_1 \sim N(\mu_1, \Sigma_1)$ where $\mu_1 = \frac{\theta_0}{1-\phi_1}$ and $\Sigma_1 = \frac{\sigma^2}{1-\phi_1^2}$ and for AR(2) model, we have $Y_2 = (y_1, y_2)' \sim N(\mu_2, \Sigma_2)$ with $\mu = \frac{\theta_0}{1-\phi_1-\phi_2}$ and

$$\Sigma_2 = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where $\Sigma_{11} = \Sigma_{22} = \frac{\sigma^2(\phi_2-1)}{(\phi_2-1)(1-\phi_1^2-\phi_2^2)+2\phi_1^2\phi_2}$ and $\Sigma_{12} = \Sigma_{21} = \frac{-\sigma^2\phi_1}{(\phi_2-1)(1-\phi_1^2-\phi_2^2)+2\phi_1^2\phi_2}$.

Exact likelihood for MA(q) model:

The MA(q) model corresponds to ARMA$(0,q)$ and, therefore, MA(q) process can be easily written from (1) as

$$y_t = \theta_0 + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} + \varepsilon_t. \qquad (39)$$

The likelihood function corresponding to MA(q) model cannot be obtained directly by putting zero in the place of AR component p in the likelihood (10). This may be because of the fact that the process (1) will not have any component of lagged time series observation in its right-hand side and the components of ε_0 are already taken to be zero while writing the likelihood corresponding to ARMA(p,q) model. One can, however, write a computationally friendly form of the exact likelihood for MA(q) model by an alternative argument.

Let the vector of complete sample observations be given by $Y_T = (y_1, y_2, ..., y_T)'_{T \times 1}$ and correspondingly the mean vector be given by $\mu_T = (\mu, \mu, ..., \mu)'_{T \times 1}$ where $\mu = \theta_0$. The variance-covariance matrix Σ_T is a $T \times T$ matrix and the same can be obtained as

$$
\Sigma_T = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} & ... & \Sigma_{1T} \\
\Sigma_{21} & \Sigma_{22} & ... & \Sigma_{2T} \\
. & . & & . \\
. & . & & . \\
. & . & & . \\
\Sigma_{T1} & \Sigma_{T2} & ... & \Sigma_{TT}
\end{bmatrix},
$$

where

$$
\Sigma_{11} = \Sigma_{22} = ... = \Sigma_{TT} = \sigma^2(1 + \psi_1^2 + \psi_2^2 + ... + \psi_q^2); \qquad \text{for} \quad t = 1, 2, ..., T \quad (40)
$$

and

$$
\Sigma_{t,t-k} = \Sigma_{t-k,t} = \begin{cases} \sigma^2(\psi_k + \psi_{k+1}\psi_1 + ... + \psi_q\psi_{q-k}), & \text{for} \quad k = 1, 2, ..., q \\ 0, & \text{for} \quad k > q. \end{cases} \qquad (41)
$$

The exact likelihood for the MA(q) model is, therefore, given by the T-variate normal distribution with mean vector μ_T and variance-covariance matrix Σ_T and it can be written up to proportionality as in (12).

Developing calibration estimators for population mean using robust measures of dispersion under stratified random sampling

Ahmed Audu[1], Rajesh Singh[2], Supriya Khare[3]

ABSTRACT

In this paper, two modified, design-based calibration ratio-type estimators are presented. The suggested estimators were developed under stratified random sampling using information on an auxiliary variable in the form of robust statistical measures, including Gini's mean difference, Downton's method and probability weighted moments. The properties (biases and MSEs) of the proposed estimators are studied up to the terms of first-order approximation by means of Taylor's Series approximation. The theoretical results were supported by a simulation study conducted on four bivariate populations and generated using normal, chi-square, exponential and gamma populations. The results of the study indicate that the proposed calibration scheme is more precise than any of the others considered in this paper.

Key words: calibration, outliers, percentage relative efficiency (PRE), stratified sampling.

1. Introduction

In sampling survey, calibration is a commonly used technique to produce estimation weights. These calibrations weights in turn satisfy calibration equation that incorporates auxiliary information. The calibration approach consists of (a) computation of new weights that incorporate specified auxiliary information and are restrained by calibration equations (b) the use of these weights to compute linearly weighted estimate of mean, totals and other finite population parameters satisfying an objective of obtaining nearly unbiased estimate. This technique has been used to develop cosmetic estimators (estimators interpretable both as design-based and as prediction-based estimators) (see Sarndal and Wright (1984), Brewer (1995, 1999), etc.). The calibration technique has also been utilized to develop design-based estimator

---

[1] Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria.
E-mail: ahmed.audu@udusok.edu.ng. ORCID: https://orcid.org/0000-0002-6915-7589.
[2] Department of Statistics, Banaras Hindu University, Varanasi, 221005, India. E-mail: rsinghstat@gmail.com. ORCID: https://orcid.org/0000-0002-9274-8141.
[3] Department of Statistics, Banaras Hindu University, Varanasi, 221005, India. Corresponding author.
E-mail: supriya.khare@bhu.ac.in. ORCID: https://orcid.org/0000-0002-7683-6715.

under different sampling schemes like stratified random sampling, stratified random double sampling, two-stage sampling, etc. In this direction many authors like Deville and Sarndal (1992), Singh and Mohl (1996), Estevao and Sarndal (2000), Estevao and Sarndal (2002), Singh (2003),Tracy et al. (2003), Kim et al. (2007), Barktus and Pumputis (2010), Sud et al. (2014), Clement and Enang (2016), Rao et al. (2016) and Subzar et al. (2018) have proposed estimators and studied their properties for estimating population mean under different calibration constraints in stratified random sampling. Tracy et al. (2003) obtained calibration weights for population mean by using first and second order moments of auxiliary variable in stratified random sampling. Nidhi et al. (2017) considered estimation of population mean using calibration approach in stratified and stratified double sampling schemes. Kim et al. (2007) utilized calibration approach in defining estimators for population variance in stratified random sampling. Other authors like Horvitz and Thompson (1952), Estevao and Sárndal (2006), Aditya et al. (2016), Salinas et al. (2019) considered estimation of population mean under two stage sampling scheme using the calibration approach.

In this paper, we have suggested two calibrated schemes in stratified random sampling by utilizing auxiliary information on certain robust statistical measures like Gini's mean difference, Downton's method and Probability weighted moments, all of which are insensitive against the presence of outliers in the population and are less susceptible to fluctuations in sampling whenever extreme observations are present as alternatives to Rao et al. (2016) calibration estimators.

2. Some existing estimators in literature

Let $\Theta_N = \left\{ \Theta_{N_h}, h = 1, 2, ..., K \right\}$ be a stratified non-overlapping heterogeneous population with K strata of size $N = \sum_{h=1}^{K} N_h$ with units $y_{hi}, i = 1, 2, ..., N_{hi}$ and $x_{hi}, i = 1, 2, ..., N_{hi}$ for study variable y and auxiliary variable x respectively. $\overline{Y}_h = N_h^{-1} \sum_{i=1}^{N_h} y_{hi}$ and $\overline{X}_h = N_h^{-1} \sum_{i=1}^{N_h} x_{hi}$ are means of study and auxiliary variables respectively. A random sample of size $n = \sum_{h=1}^{K} n_h$ is selected from the population using SRSWOR. The conventional unbiased estimator of the population mean and its variance is given in Eq. (2.1) and Eq. (2.2), respectively.

$$\overline{y}_{st} = \sum_{h=1}^{K} \Delta_h \overline{y}_h \tag{2.1}$$

$$Var\left(\overline{y}_{st}\right) = \sum_{h=1}^{K} \Delta_h^2 \left(n_h^{-1} - N_h^{-1}\right) S_{yh}^2 \tag{2.2}$$

Where,

$$\Delta_h = N_h / N, \ \overline{y}_h = n_h^{-1} \sum_{i=1}^{n_h} y_{hi}, S_{yh}^2 = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (y_{hi} - \overline{Y}_h)^2, \overline{Y} = \sum_{h=1}^{K} \Delta_h \overline{Y}_h$$

Singh (2003) suggested a design-based calibration estimator with two constraints for estimating population mean in stratified sampling. The suggested calibration estimator is given in Eq. (2.3).

$$\overline{y}_S = \sum_{h=1}^{K} \Delta_h^S \overline{y}_h \tag{2.3}$$

where Δ_h^S is the new calibration weight of stratum K^{th} to be obtained by solving (2.4).

$$\left. \begin{array}{ll} \min & Z_S = \sum_{h=1}^{K} \left(\Delta_h^S - \Delta_h \right)^2 / \Delta_h \phi_h \\ s.t & \sum_{h=1}^{K} \Delta_h^S \overline{x}_h = \sum_{h=1}^{K} \Delta_h \overline{X}_h, \quad \sum_{h=1}^{K} \Delta_h^S = \sum_{h=1}^{K} \Delta_h \end{array} \right\} \tag{2.4}$$

where ϕ_h are suitably chosen positive scale factors, which decide the form of the estimator.

Eq.(2.4) yields a calibration weight in Eq. (2.5) and the estimator $\overline{y}_S$ was obtained as in Eq. (2.6).

$$\Delta_h^S = \Delta_h + \frac{\phi_h \Delta_h \overline{x}_h \sum_{h=1}^{K} \Delta_h \phi_h - \Delta_h \phi_h \sum_{h=}^{K} \Delta_h \phi_h \overline{x}_h}{\sum_{h=}^{K} \Delta_h \phi_h \sum_{h=}^{K} \Delta_h \phi_h \overline{x}_h^2 - \left(\sum_{h=}^{K} \Delta_h \phi_h \overline{x}_h \right)^2} \left(\overline{X} - \sum_{h=}^{K} \Delta_h \overline{x}_h \right)$$

$$\tag{2.5}$$

$$\overline{y}_S = \sum_{h=1}^{K} \Delta_h \overline{y}_h + \frac{\sum_{h=1}^{K} \phi_h \Delta_h \sum_{h=1}^{K} \phi_h \Delta_h \overline{x}_h \overline{y}_h - \sum_{h=1}^{K} \phi_h \Delta_h \overline{x}_h \sum_{h=1}^{K} \phi_h \Delta_h \overline{y}_h}{\sum_{h=1}^{K} \phi_h \Delta_h \sum_{h=1}^{K} \phi_h \Delta_h \overline{x}_h^2 - \left(\sum_{h=1}^{K} \phi_h \Delta_h \overline{x}_h \right)^2} \left(\overline{X} - \sum_{h=1}^{K} \Delta_h \overline{x}_h \right)$$

$$\tag{2.6}$$

Clement and Enang (2016) suggested a design-based calibration estimator for the combined ratio estimator in stratified random sampling. The suggested estimators with the associated calibration constraint are given in Eq. (2.7) and Eq. (2.8).

$$\overline{y}_{CE} = \sum_{h=1}^{K} \Delta_h^{CE} \hat{R} \overline{X} \tag{2.7}$$

$$\left. \begin{array}{ll} \min & Z_{CE} = \sum_{h=1}^{K} \left(\Delta_h^{CE} - \Delta_h \right)^2 / \Delta_h \phi_h \\ s.t & \sum_{h=1}^{K} \Delta_h^{CE} \overline{x}_h = \overline{X} \end{array} \right\} \tag{2.8}$$

where $\hat{R}_h = \overline{y}_h / \overline{x}_h$ $\Delta_h^{\textit{Œ}}$ is the proposed calibration weight of Kth stratum.

The calibration weight Δ_h^*, estimator $\overline{y}_{CE}$ and $\mathrm{var}\left(\overline{y}_{CE}\right)$ were obtained as given in Eq. (2.9), Eq. (2.10) and Eq. (2.11) respectively.

$$\Delta_h^{CE} = \Delta_h + \frac{\Delta_h \phi_h \overline{x}_h}{\sum_{h=1}^K \Delta_h \phi_h \overline{x}_h^2}\left(\overline{X} - \sum_{h=1}^K \Delta_h \overline{x}_h\right) \tag{2.9}$$

$$\overline{y}_{CE} = \overline{X}\sum_{h=1}^K \Delta_h \overline{y}_h \Big/ \sum_{h=1}^K \Delta_h \overline{x}_h \tag{2.10}$$

$$Bias\left(\overline{y}_{CE}\right) = \overline{X}^{-1}\sum_{h=1}^K \Delta_h \left(n_h^{-1} - N_h^{-1}\right)\left(RS_{xh}^2 - S_{yxh}\right) \tag{2.11}$$

$$Var\left(\overline{y}_{CE}\right) = \left(\overline{X}^2 Var\left(\overline{x}_{st}\right)\sum_{h=1}^K \Delta_h^2 \left(n_h^{-1} - N_h^{-1}\right)S_h^{*2}\right) \Big/ \hat{\overline{X}}^2 Var\left(\overline{x}_{st}^*\right) \tag{2.12}$$

where

$$\hat{\overline{X}} = \sum_{h=1}^K \Delta_h \overline{x}_h, Var\left(\overline{x}_{st}^*\right) = \sum_{h=1}^K \Delta_h S_{xh}^2, S_h^{*2} = S_{yh}^2 + R^2 S_{xh}^2 - 2RS_{yxh}, R = \overline{Y} / \overline{X},$$
$$S_{yxh} = \left(N_h - 1\right)^{-1}\sum_{i=1}^{N_h}\left(y_{hi} - \overline{Y}_h\right)\left(x_{hi} - \overline{X}_h\right)$$

Rao et al. (2016) proposed two new design-based calibration schemes by incorporating coefficient of variation in the constraint to the chi-square distance function for the new calibration weight defined to improve the precision of the sample mean estimator in stratified random sampling. The first scheme proposed is given in Eq. (2.13).

$$\overline{y}_{RTK} = \sum_{h=1}^K \Delta_h^* \overline{y}_h \tag{2.13}$$

where Δ_h^* is the new calibration weight such that the chi-square function Z_* is defined as

$$\left.\begin{aligned} \min \quad & Z_* = \sum_{h=1}^K \frac{\left(\Delta_h^* - \Delta_h\right)^2}{\Delta_h \phi_h} \\ s.t \quad & \sum_{h=1}^K \Delta_h^*\left(\overline{x}_h + c_{xh}\right) = \sum_{h=1}^K \Delta_h\left(\overline{X}_h + C_{Xh}\right) \end{aligned}\right\} \tag{2.14}$$

where

$$c_{xh} = s_{xh} / \bar{x}_h, C_{Xh} = S_{Xh} / \bar{X}_h, s_{xh}^2 = (n_h - 1)^{-1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2, \bar{x}_h = n_h^{-1} \sum_{i=1}^{n_h} x_{hi}$$

$$S_{Xh}^2 = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$$

Solving Eq. (2.14) and let $\phi_h = (\bar{x}_h + c_{xh})^{-1}$, the calibration weight Δ_{h1}^* and the estimator $\bar{y}_{RTK}$ are given by Eq. (2.15) and Eq. (2.16) respectively.

$$\Delta_{h1}^* = \Delta_h + \Delta_h \left(\sum_{h=1}^{K} \Delta_h (\bar{X}_h + C_{Xh}) - \sum_{h=1}^{K} \Delta_h (\bar{x}_h + c_{xh}) \right) \left(\sum_{h=1}^{K} \Delta_h (\bar{x}_h + c_{xh}) \right)^{-1} \qquad (2.15)$$

$$\bar{y}_{RTK1} = \sum_{h=1}^{K} \Delta_h \bar{y}_h \sum_{h=1}^{K} \Delta_h (\bar{X}_h + C_{Xh}) \left(\sum_{h=1}^{K} \Delta_h (\bar{x}_h + c_{xh}) \right)^{-1} \qquad (2.16)$$

Similarly, function Z_* is also subjected to another constraint defined in Eq. (2.17),

$$\sum_{h=1}^{K} \Delta_h^* (1 + \bar{x}_h + c_{xh}) = \sum_{h=1}^{K} \Delta_h (1 + \bar{X}_h + C_{Xh}) \qquad (2.17)$$

which lead to another estimator given as

$$\Delta_{h2}^* = \Delta_h + \Delta_h \left(\sum_{h=1}^{K} \Delta_h (1 + \bar{X}_h + C_{Xh}) - \sum_{h=1}^{K} \Delta_h (1 + \bar{x}_h + c_{xh}) \right) \left(\sum_{h=1}^{K} \Delta_h (1 + \bar{x}_h + c_{xh}) \right)^{-1} \qquad (2.18)$$

$$\bar{y}_{RTK2} = \sum_{h=1}^{K} \Delta_h \bar{y}_h \sum_{h=1}^{K} \Delta_h (1 + \bar{X}_h + C_{Xh}) \left(\sum_{h=1}^{K} \Delta_h (1 + \bar{x}_h + c_{xh}) \right)^{-1} \qquad (2.19)$$

However, estimators $\bar{y}_{RTK1}$ and $\bar{y}_{RTK2}$ are functions of coefficients of variation which are easily affected outliers or extreme values.

3. Suggested calibration estimators

Motivated by Clement and Enang (2016) and Rao et al. (2016), we proposed two classes of design-based calibration estimators in stratified random sampling using robust measures such as Gini's mean difference G_{MD}, Downton's method D_M and probability weighted moments P_{WM} of the auxiliary information, which are insensitive to the presence of outliers or extreme values in the data.

Let $z \in \Re^+$ with units $z_i, i = 1, 2, ..., N$, then:

$$G_{MD}(z) = 2N^{-1}(N-1)^{-1}\sum_{i=1}^{N}(2i-N-1)z_i \tag{3.1}$$

$$D_M(z) = 2\sqrt{\pi}N^{-1}(N-1)^{-1}\sum_{i=1}^{N}(i-(N+1)/2)z_i \tag{3.2}$$

$$P_{WM}(z) = \sqrt{\pi}N^{-2}\sum_{i=1}^{N}(2i-(N+1))z_i \tag{3.3}$$

3.1. First calibration scheme proposed

Consider an estimator defined in Eq. (3.4) under stratified sampling having distance function as given in Eq. (3.5),

$$\overline{y}_{AR} = \sum_{h=1}^{K}\Omega_{hi}^{*}\overline{y}_h, \; i=1,2,3. \tag{3.4}$$

where Ω_{hi}^{*} is the new calibration weights such that the chi-square function Z^{*} is defined as

$$\left.\begin{array}{l} \min \quad Z^{*} = \sum_{h=1}^{K}\dfrac{\left(\Omega_{hi}^{*}-\Delta_h\right)^2}{\Delta_h\phi_h} \\[3mm] s.t \; \sum_{h=1}^{K}\Omega_{hi}^{*}\left(\overline{x}_h+\lambda_{hi}(x)\right) = \sum_{h=1}^{K}\Delta_h\left(\overline{X}_h+\lambda_{hi}(x)\right), i=1,2,3 \end{array}\right\} \tag{3.5}$$

where $\lambda_{h1}(x) = G_{MDh}(x), \lambda_{h2}(x) = D_{Mh}(x), \lambda_{h3}(x) = P_{WMh}(x)$

To compute the new calibrated weights Ω_{hi}^{*}, we use the Lagrange multipliers function of the form given by Eq. (3.6),

$$\Phi = \sum_{h=1}^{K}\dfrac{\left(\Omega_{hi}^{*}-\Delta_h\right)^2}{\Delta_h\phi_h} - 2\eta\left(\sum_{h=1}^{K}\Omega_{hi}^{*}\left(\overline{x}_h+\lambda_{hi}(x)\right)-\sum_{h=1}^{K}\Delta_h\left(\overline{X}_h+\lambda_{hi}(x)\right)\right) \tag{3.6}$$

Partially differentiating Eq. (3.6) with respect to Ω_{h}^{*} and η and equating to zero, we have

$$\Omega_{hi}^{*} = \Delta_h + \eta\Delta_h\phi_h\left(\overline{x}_h+\lambda_{hi}(x)\right) \tag{3.7}$$

$$\sum_{h=1}^{K}\Omega_{hi}^{*}\left(\overline{x}_h+\lambda_{hi}(x)\right) - \sum_{h=1}^{K}\Delta_h\left(\overline{X}_h+\lambda_{hi}(x)\right) = 0 \tag{3.8}$$

Substituting Eq. (3.7) in Eq. (3.8) to get η and then substituting the expression obtained into Eq. (3.7). By putting $\phi_h = \left(\overline{x}_h+\lambda_{hi}(x)\right)^{-1}$, the new calibration weight Ω_{hi}^{*} is obtained as

$$\Omega_{hi}^* = \Delta_h + \Delta_h \left(\sum_{h=1}^{K} \Delta_h \left(\overline{X}_h + \lambda_{hi}(x) \right) - \sum_{h=1}^{K} \Delta_h \left(\overline{x}_h + \lambda_{hi}(x) \right) \right) \left(\sum_{h=1}^{K} \Delta_h \left(\overline{x}_h + \lambda_{hi}(x) \right) \right)^{-1} \quad (3.9)$$

Now, substituting Eq. (3.9) in Eq. (3.4) and letting $\lambda_{hi}(x)$ be either $G_{MDh}(x)$ or $D_{Mh}(x)$ or $P_{WMh}(x)$, the new estimators are obtained as,

$$\left.\begin{array}{l} \overline{y}_{AR1} = \sum_{h=1}^{K} \Delta_h \overline{y}_h \sum_{h=1}^{K} \Delta_h \left(\overline{X}_h + G_{MDh}(x) \right) \left(\sum_{h=1}^{K} \Delta_h \left(\overline{x}_h + G_{MDh}(x) \right) \right)^{-1} \\[20pt] \overline{y}_{AR2} = \sum_{h=1}^{K} \Delta_h \overline{y}_h \sum_{h=1}^{K} \Delta_h \left(\overline{X}_h + D_{Mh}(x) \right) \left(\sum_{h=1}^{K} \Delta_h \left(\overline{x}_h + D_{Mh}(x) \right) \right)^{-1} \\[20pt] \overline{y}_{AR3} = \sum_{h=1}^{K} \Delta_h \overline{y}_h \sum_{h=1}^{K} \Delta_h \left(\overline{X}_h + P_{WMh}(x) \right) \left(\sum_{h=1}^{K} \Delta_h \left(\overline{x}_h + P_{WMh}(x) \right) \right)^{-1} \end{array}\right\}$$

$$(3.10)$$

3.2. Second calibration scheme proposed

To obtain the second class of the proposed estimators, we let

$$\overline{y}_{AS} = \sum_{h=1}^{K} \Pi_{hi}^* \overline{y}_h, \quad i = 1, 2, 3. \quad (3.11)$$

where Π_{hi}^* is the new calibration weight such that the chi-square function U^* is defined as

$$\left.\begin{array}{l} \min \quad U^* = \sum_{h=1}^{K} \frac{\left(\Pi_{hi}^* - \Delta_h \right)^2}{\Delta_h \phi_h} \\[20pt] s.t \sum_{h=1}^{K} \Pi_{hi}^* \left(1 + \overline{x}_h + \lambda_{hi}(x) \right) = \sum_{h=1}^{K} \Delta_h \left(1 + \overline{X}_h + \lambda_{hi}(x) \right), \ i = 1, 2, 3 \end{array}\right\}$$

$$(3.12)$$

Solving for Π_h^* using the Lagrange multipliers technique and putting $\phi_h = \left(1 + \overline{X}_h + \lambda_{hi}(x) \right)^{-1}$, we have the new calibrated weight given by Eq. (2.14).

$$\Pi_{hi}^* = \Delta_h + \Delta_h \left(\sum_{h=1}^{K} \Delta_h \left(1 + \overline{X}_h + \lambda_{hi}(x) \right) - \sum_{h=1}^{K} \Delta_h \left(1 + \overline{x}_h + \lambda_{hi}(x) \right) \right) \left(\sum_{h=1}^{K} \Delta_h \left(1 + \overline{x}_h + \lambda_{hi}(x) \right) \right)^{-1}$$

$$(3.13)$$

By putting Eq. (3.13) in Eq. (3.11) and letting $\lambda_{hi}(x)$ be either $G_{MDh}(x)$ or $D_{Mh}(x)$ or $P_{WMh}(x)$, the new estimators are obtained as

$$
\left.
\begin{aligned}
\bar{y}_{AS1} &= \sum_{h=1}^{K}\Delta_h \bar{y}_h \sum_{h=1}^{K}\Delta_h\left(1+\bar{X}_h + G_{MDh}(x)\right)\left(\sum_{h=1}^{K}\Delta_h\left(1+\bar{x}_h + G_{MDh}(x)\right)\right)^{-1} \\
\bar{y}_{AS2} &= \sum_{h=1}^{K}\Delta_h \bar{y}_h \sum_{h=1}^{K}\Delta_h\left(1+\bar{X}_h + D_{Mh}(x)\right)\left(\sum_{h=1}^{K}\Delta_h\left(1+\bar{x}_h + D_{Mh}(x)\right)\right)^{-1} \\
\bar{y}_{AS3} &= \sum_{h=1}^{K}\Delta_h \bar{y}_h \sum_{h=1}^{K}\Delta_h\left(1+\bar{X}_h + P_{WMh}(x)\right)\left(\sum_{h=1}^{K}\Delta_h\left(1+\bar{x}_h + P_{WMh}(x)\right)\right)^{-1}
\end{aligned}
\right\}
$$

$$(3.14)$$

3.3. Properties (bias and MSE) of the proposed estimators

To obtain bias and MSE of the suggested estimators $\bar{y}_{ARi}$, $\bar{y}_{ASi}$, the following error terms are defined: $e_0 = \left(\bar{y}_{st} - \bar{Y}\right)/\bar{Y}$, $e_1 = \left(\bar{x}_{st} - \bar{X}\right)/\bar{X}$ with expected values defined in Eq. (3.15)

$$
\left.
\begin{aligned}
E(e_0) = E(e_1) = 0, E(e_0^2) = Var(\bar{y}_{st})/\bar{Y}^2, \\
E(e_1^2) = Var(\bar{x}_{st})/\bar{X}^2, E(e_0 e_1) = C\,\mathrm{ov}(\bar{y}_{st}\bar{x}_{st})/\overline{YX}
\end{aligned}
\right\}
$$
$$(3.15)$$

where

$$
Var(\bar{x}_{st}) = \sum_{h=1}^{K}\Delta_h^2\left(n_h^{-1} - N_h^{-1}\right)S_{xh}^2, C\,\mathrm{ov}(\bar{y}_{st}) = \sum_{h=1}^{K}\Delta_h^2\left(n_h^{-1} - N_h^{-1}\right)S_{yxh}.
$$

Expressing Eq. (3.10) and Eq. (3.14) in terms of $e_i, i = 0,1$ and simplifying up to the second degree approximation, we obtained Eq. (3.16) and Eq. (3.17) respectively as

$$
\bar{y}_{ARi} = \bar{Y}\left(1+e_0\right)\sum_{h=1}^{K}\Delta_h\left(\bar{X}_h + \lambda_{ih}(x)\right)/\left(\bar{X}e_1 + \sum_{h=1}^{K}\Delta_h\left(\bar{X}_h + \lambda_{ih}(x)\right)\right)
$$

$$(3.16)$$

$$
\bar{y}_{ASi} = \bar{Y}\left(1+e_0\right)\sum_{h=1}^{K}\Delta_h\left(1+\bar{X}_h + \lambda_{ih}(x)\right)/\left(\bar{X}e_1 + \sum_{h=1}^{K}\Delta_h\left(1+\bar{X}_h + \lambda_{ih}(x)\right)\right)
$$

$$(3.17)$$

Simplifying Eq. (3.16) and Eq. (3.17), we get Eq. (3.18) and Eq. (3.19)

$$
\bar{y}_{ARi} = \bar{Y}\left(1+e_0\right)\left(1+\varpi_i e_1\right)^{-1}
$$
$$(3.18)$$

$$\bar{y}_{ASi} = \bar{Y}\left(1+e_0\right)\left(1+\vartheta_i e_1\right)^{-1} \tag{3.19}$$

where

$$\varpi_i = \sum_{h=1}^{K}\Delta_h\bar{X}_h \Big/ \sum_{h=1}^{K}\Delta_h\left(\bar{X}_h + \lambda_{ih}(x)\right), \quad \vartheta_i = \sum_{h=1}^{K}\Delta_h\bar{X}_h \Big/ \sum_{h=1}^{K}\Delta_h\left(1+\bar{X}_h + \lambda_{ih}(x)\right).$$

Simplifying Eq. (3.18) and Eq. (3.19) up to the first order approximation, we obtained

$$\bar{y}_{ARi} - \bar{Y} = \bar{Y}\left(e_0 - \varpi_i e_1 + \varpi_i^2 e_1^2 - \varpi_i e_0 e_1\right) \tag{3.20}$$

$$\bar{y}_{ASi} - \bar{Y} = \bar{Y}\left(e_0 - \vartheta_i e_1 + \vartheta_i^2 e_1^2 - \vartheta_i e_0 e_1\right) \tag{3.21}$$

Take expectation of Eq. (3.20), Eq. (3.21) and using the results obtained in Eq. (3.15), we obtained the $Bias\left(\bar{y}_{ARi}\right)$ and $Bias\left(\bar{y}_{ASi}\right)$ as

$$Bias\left(\bar{y}_{ARi}\right) = R\bar{X}^{-1}\varpi_i^2 Var\left(\bar{x}_{st}\right) - \bar{X}^{-1}\varpi_i C\,ov\left(\bar{y}_{st}\bar{x}_{st}\right) \tag{3.22}$$

$$Bias\left(\bar{y}_{ASi}\right) = R\bar{X}^{-1}\vartheta_i^2 Var\left(\bar{x}_{st}\right) - \bar{X}^{-1}\vartheta_i C\,ov\left(\bar{y}_{st}\bar{x}_{st}\right) \tag{3.23}$$

where $R = \bar{Y}/\bar{X}$.

Squaring Eq. (3.20) and Eq. (3.21), and taking expectations and substituting the results of Eq. (3.15), we obtained the $MSE\left(\bar{y}_{ARi}\right)$ and $MSE\left(\bar{y}_{ASi}\right)$ as given in Eq. (3.24) and Eq. (3.25) respectively.

$$MSE\left(\bar{y}_{ARi}\right) = Var\left(\bar{y}_{st}\right) + R^2\varpi_i^2 Var\left(\bar{x}_{st}\right) - 2R\varpi_i C\,ov\left(\bar{y}_{st}\bar{x}_{st}\right), \; i = 1,2,3 \tag{3.24}$$

$$MSE\left(\bar{y}_{ASi}\right) = Var\left(\bar{y}_{st}\right) + R^2\vartheta_i^2 Var\left(\bar{x}_{st}\right) - 2R\vartheta_i C\,ov\left(\bar{y}_{st}\bar{x}_{st}\right), \; i = 1,2,3 \tag{3.25}$$

3.4. Properties of the New Weights Ω_{hi}^* and $\Pi_{hi}^*, i = 1,2,3$

Theorem 1: The proposed weights Ω_{hi}^* and $\Pi_{hi}^*, i = 1,2,3$ are consistent.

Proof: As the sample size in each stratum tends to the stratum size, i.e. as $n_n \to N_h$, the stratum sample mean converges to the stratum population mean, i.e. $\bar{x}_h \to \bar{X}_h$.

Then, the expression $\sum_{h=1}^{K}\Delta_h\left(\bar{X}_h + \lambda_{hi}(x)\right) - \sum_{h=1}^{K}\Delta_h\left(\bar{x}_h + \lambda_{hi}(x)\right)$ in

$\Omega_{hi}^*, i = 1, 2, 3$ and expression $\sum_{h=1}^{K} \Delta_h \left(1 + \bar{X}_h + \lambda_{hi}(x)\right) - \sum_{h=1}^{K} \Delta_h \left(1 + \bar{x}_h + \lambda_{hi}(x)\right)$

in $\Pi_{hi}^*, i = 1, 2, 3$ tend to zeros. So,

$$\lim_{n_h \to N_h} \frac{\Omega_{hi}^*}{\Delta_h} \to 1 \tag{3.26}$$

$$\lim_{n_h \to N_h} \frac{\Pi_{hi}^*}{\Delta_h} \to 1 \tag{3.27}$$

Theorem 2: The sum of the proposed weights Ω_{hi}^* and $\Pi_{hi}^*, i = 1, 2, 3$ converged to unity.

Proof: Taking the summation of Ω_{hi}^* and $\Pi_{hi}^*, i = 1, 2, 3$ over K, we obtained

$$\sum_{h=1}^{K} \Omega_{hi}^* = 1 + K \left(\bar{X} - \bar{x}_{st}\right) / \sum_{h=1}^{K} \Delta_h \left(\bar{x}_h + \lambda_{hi}(x)\right) \tag{3.28}$$

$$\sum_{h=1}^{K} \Pi_{hi}^* = 1 + K \left(\bar{X} - \bar{x}_{st}\right) / \sum_{h=1}^{K} \Delta_h \left(1 + \bar{x}_h + \lambda_{hi}(x)\right) \tag{3.29}$$

As $n_n \to N_h$, $\bar{x}_h \to \bar{X}_h$ and $\bar{x}_{st} \to \bar{X}$, then

$$\lim_{n_h \to N_h} \sum_{h=1}^{K} \Omega_{hi}^* = \lim_{n_h \to N_h} \sum_{h=1}^{K} \Pi_{hi}^* = 1 \tag{3.30}$$

Theorem 3: The proposed weights $0 < \Omega_{hi}^* < 1$ and $0 < \Pi_{hi}^* < 1, i = 1, 2, 3$.

Proof: As $n_n \to N_h$, $\bar{x}_h \to \bar{X}_h$ and $\bar{x}_{st} \to \bar{X}$, then

$$\lim_{n_h \to N_h} \Omega_{hi}^* = \lim_{n_h \to N_h} \Pi_{hi}^* = \Delta_h = N_h / N \tag{3.31}$$

Since $N_h > 0, N > 0$ and $N_h < N$, then $0 < \Delta_h < 1$.

4. Empirical study

4.1. Simulation study

In this section, we perform a simulation study to examine the superiority of the proposed estimators over other estimators considered in the study. For this, we generate a bivariate random population of size N=1000 for study population stratified into 3 non-overlapping heterogeneous groups of size 200, 300 and 500 using function defined in Table 4.1. Samples of sizes 20, 30 and 50 were selected 10,000 times by the SRSWOR method from each stratum respectively. The precision (PRE) of the considered estimators was computed using Eq. (4.1).

$$Bias\left(\hat{\theta}\right)=\frac{1}{10000}\sum_{j=1}^{10000}\left(\hat{\theta}-\bar{Y}\right) \tag{4.1}$$

$$MSE\left(\hat{\theta}\right)/Var\left(\hat{\theta}\right)=\frac{1}{10000}\sum_{j=1}^{10000}\left(\hat{\theta}_{j}-\bar{Y}\right)^{2} \tag{4.2}$$

$$PRE\left(\hat{\theta}\right)=\left(Var\left(\bar{y}_{st}\right)/Var\left(\hat{\theta}\right)\right)100 \tag{4.3}$$

where

$$\text{var}\left(\bar{y}_{st}\right)=\frac{1}{10000}\sum_{j=1}^{10000}\left(\bar{y}_{st}-\bar{Y}\right)^{2},\hat{\theta}=\bar{y}_{st},\bar{y}_{CE1},\bar{y}_{CE2},\bar{y}_{RTK1},\bar{y}_{RTK2},\bar{y}_{ARi},\bar{y}_{ASi}$$

Table 4.1. Populations used for Empirical Study

Population	Auxiliary variable x	Study variable y
I	$x_h \square\ N\left(\mu_h,\sigma_h\right), \mu_1=60, \sigma_1=50,$ $\mu_2=50, \sigma_2=70, \mu_3=30, \sigma_3=40$	$y_{hi}=\alpha x_{hi}+x_{hi}^2+\xi_{hi},$ $\alpha=0.5,1,1.5,2.0,2.5,3$ $\xi_h \square\ N\left(0,1\right), h=1,2,3$
II	$x_h \square\ chsq\left(\theta_h\right), \theta_1=1, \theta_2=2, \theta_3=3$	
III	$x_h \square\ \exp\left(\lambda_h\right), \lambda_1=0.2, \lambda_2=0.3, \lambda_3=0.1$	
IV	$x_h \square\ gamma\left(\theta_h,\eta_h\right), \theta_1=3, \eta_1=2,$ $\theta_2=3, \eta_2=1, \theta_3=3, \eta_3=3,$	

    Table 4.2 shows the biases, MSEs and PREs of the traditional, Rao et al. (2016), Clement and Enang (2016) and the proposed estimators using population I defined in Table 4.1. The proposed estimators have smaller MSEs compared to other estimators. This implies that the estimates of the proposed estimators are on average closer to the true estimate than that of other estimators. The PREs of the proposed estimators are higher than that of other estimators. The proposed estimator under has PRE of 326.4 implying 200% and 100% gain in efficiency over and respectively. However, the proposed estimators are averagely more biased compared to other estimators considered in the study.

Table 4.2. PRE of the Proposed and Some Existing Estimators using Pop. BI

Estimators	Values of α								
	0.5			**1.0**			**1.5**		
	Bias	MSE	PRE	Bias	MSE	PRE	Bias	MSE	PRE
$\overline{y}_{st}$	0.1	404126.7	100	0.1	406568.4	100	0.1	409023.1	100
Rao et al. (2016)									
$\overline{y}_{RTK1}$	-0.8	176491.7	229	-0.8	176432.9	230.4	-0.9	176374.1	231.9
$\overline{y}_{RTK2}$	-2	174926	231	-2.1	174866.3	232.5	-2.1	174806.6	234
Clement and Enang (2016)									
$\overline{y}_{CE}$	8.9	192907.5	209.5	8.9	192907.5	210.8	8.9	192907.5	212
Proposed									
$\overline{y}_{AR1}$	-17.2	123802.4	326.4	-17.2	124216.1	327.3	-17.2	124636	328.2
$\overline{y}_{AS1}$	-17.2	125626.8	321.7	-17.2	126052.3	322.5	-17.2	126484.1	323.4
$\overline{y}_{AR2}$	-17.4	119408.2	338.4	-17.4	119731.9	339.6	-17.4	120061.1	340.7
$\overline{y}_{AS2}$	-17.4	121356.7	333	-17.4	121693.5	334.1	-17.4	122035.9	335.2
$\overline{y}_{AR3}$	58.2	125249.8	322.7	58.5	125596.7	323.7	58.8	125949.3	324.8
$\overline{y}_{AS3}$	57.4	127154.8	317.8	57.7	127514.5	318.8	58.0	127879.9	319.8
Estimators	Values of α								
	2.0			**2.5**			**3.0**		
$\overline{y}_{st}$	0.2	411490.7	100	0.2	413971.2	100	0.2	416464.8	100.0
Rao et al. (2016)									
$\overline{y}_{RTK1}$	-0.9	176315.5	233.4	-0.9	176256.9	234.9	-1.0	176198.3	236.4
$\overline{y}_{RTK2}$	-2.1	174747	235.5	-2.2	174687.5	237	-2.2	174628.1	238.5
Clement and Enang (2016)									
$\overline{y}_{CE}$	8.9	192907.5	213.3	8.9	192907.5	214.6	8.9	192907.5	215.9
Proposed									
$\overline{y}_{AR1}$	-17.2	125062	329	-17.3	125494.3	329.9	-17.3	125932.7	330.7
$\overline{y}_{AS1}$	-17.2	126922	324.2	-17.3	127366.1	325	-17.3	127816.4	325.8
$\overline{y}_{AR2}$	-17.4	120395.9	341.8	-17.5	120736.2	342.9	-17.5	121082	344
$\overline{y}_{AS2}$	-17.5	122383.8	336.2	-17.5	122737.4	337.3	-17.5	123096.5	338.3
$\overline{y}_{AR3}$	59.1	126307.7	325.8	59.5	126671.8	326.8	59.8	127041.7	327.8
$\overline{y}_{AS3}$	58.3	128251.1	320.8	58.6	128628.1	321.8	58.9	129010.9	322.8

Table 4.3 also shows the biases, MSEs and PREs of the traditional, Rao et al. (2016), Clement and Enang (2016) and the proposed estimators using population II defined in Table 4.1 The proposed estimators have smaller MSEs compared to other estimators. These results are in conformity with that of population in Table 4.2.

Table 4.3. PRE of the Proposed and Some Existing Estimators using Pop. II

Estimators	Values of α								
	0.5			1.0			1.5		
	Bias	MSE	PRE	Bias	MSE	PRE	Bias	MSE	PRE
$\bar{y}_{st}$	0.02	3.3	100	0.03	3.6	100	0.03	4.0	100
Rao et al. (2016)									
$\bar{y}_{RTK1}$	0.05	1.4	235.7	0.04	1.4	257.1	0.02	1.5	266.7
$\bar{y}_{RTK2}$	0.03	1.7	194.1	0.02	1.8	200	0.04	2.0	200
Clement and Enang (2016)									
$\bar{y}_{CE}$	-0.1.	1.0	330	-0.1	1.0	360	-0.1	1.0	400
Proposed									
$\bar{y}_{AR1}$	-0.1	0.9	366.7	-0.1	0.8	450	-0.1	0.8	500
$\bar{y}_{AS1}$	-0.1	1.2	275	-0.1	1.2	300	-0.1	1.2	333.3
$\bar{y}_{AR2}$	-0.1	0.9	366.7	-0.1	0.8	450	-0.1	0.8	500
$\bar{y}_{AS2}$	-0.1	1.2	275	-0.1	1.2	300	-0.1	1.2	333.3
$\bar{y}_{AR3}$	0.1	0.9	366.7	0.1	0.9	400	0.1	0.9	444.4
$\bar{y}_{AS3}$	0.02	1.2	275	0.01	1.2	300	0.1	1.3	307.7
Estimators	Values of α								
	2.0			2.5			3.0		
$\bar{y}_{st}$	0	4.4	100	0	4.8	100	0	5.2	100
Rao et al. (2016)									
$\bar{y}_{RTK1}$	0	1.5	293.3	0	1.6	300	0	1.7	305.9
$\bar{y}_{RTK2}$	0	2.1	209.5	0	2.2	218.2	0	2.3	226.1
Clement and Enang (2016)									
$\bar{y}_{CE}$	-0.1	1	440	-0.1	1	480	-0.1	1	520
Proposed									
$\bar{y}_{AR1}$	-0.1	0.8	550	-0.1	0.8	600	-0.1	0.8	650
$\bar{y}_{AS1}$	-0.1	1.3	338.5	-0.1	1.3	369.2	-0.1	1.3	400
$\bar{y}_{AR2}$	-0.1	0.8	550	-0.1	0.8	600	-0.1	0.8	650
$\bar{y}_{AS2}$	-0.1	1.3	338.5	-0.1	1.3	369.2	-0.1	1.3	400
$\bar{y}_{AR3}$	0.1	0.9	488.9	0.1	0.9	533.3	0.1	0.9	577.8
$\bar{y}_{AS3}$	0.1	1.3	338.5	0.1	1.3	369.2	0.1	1.4	371.4

Table 4.4. PRE of the Proposed and Some Existing Estimators using Pop. III

Estimators	Values of α								
	0.5			1.0			1.5		
	Bias	MSE	PRE	Bias	MSE	PRE	Bias	MSE	PRE
$\overline{y}_{st}$	-0.1	396	100	-0.1	405.7	100	-0.1	415.5	100
Rao et al. (2016)									
$\overline{y}_{RTK1}$	-1.0	199.4	198.6	-1.0	200.4	202.4	-1.0	201.3	206.4
$\overline{y}_{RTK2}$	-1.0	223.2	177.4	-1.0	225.1	180.2	-1.0	227.1	183
Clement and Enang (2016)									
$\overline{y}_{CE}$	-1.3	175.3	225.9	-1.3	175.3	231.4	-1.3	175.3	237
Proposed									
$\overline{y}_{AR1}$	-1.4	152.7	259.3	-1.4	152.1	266.7	-1.4	151.4	274.4
$\overline{y}_{AS1}$	-1.4	170.1	232.8	-1.4	170	238.6	-1.4	170	244.4
$\overline{y}_{AR2}$	-1.4	153.8	257.5	-1.4	153.2	264.8	-1.4	152.5	272.5
$\overline{y}_{AS2}$	-1.4	172.1	230.1	-1.4	172.1	235.7	-1.4	172.1	241.4
$\overline{y}_{AR3}$	-0.5	156.8	252.6	-0.5	156.2	259.7	-0.5	155.5	267.2
$\overline{y}_{AS3}$	-0.6	175.5	225.6	-0.5	175.6	231	-0.5	175.6	236.6
Estimators	Values of α								
	2.0			2.5			3.0		
$\overline{y}_{st}$	-0.1	425.4	100	-0.1	435.5	100	-0.1	445.7	100
Rao et al. (2016)									
$\overline{y}_{RTK1}$	-1.0	202.3	210.3	-1.0	203.3	214.2	-1.0	204.2	218.3
$\overline{y}_{RTK2}$	-1.0	229	185.8	-1.0	231	188.5	-1.0	233	191.3
Clement and Enang (2016)									
$\overline{y}_{CE}$	-1.3	175.3	242.7	-1.3	175.3	248.4	-1.3	175.3	254.2
Proposed									
$\overline{y}_{AR1}$	-1.4	150.8	282.1	-1.4	150.1	290.1	-1.4	149.5	298.1
$\overline{y}_{AS1}$	-1.4	170	250.2	-1.4	170	256.2	-1.4	170.1	262
$\overline{y}_{AR2}$	-1.4	151.9	280.1	-1.4	151.3	287.8	-1.4	150.7	295.8
$\overline{y}_{AS2}$	-1.4	172.2	247.0	-1.4	172.3	252.8	-1.4	172.3	258.7
$\overline{y}_{AR3}$	-0.4	154.9	274.6	-0.4	154.2	282.4	-0.4	153.6	290.2
$\overline{y}_{AS3}$	-0.5	175.7	242.1	-0.5	175.7	247.9	-0.4	175.8	253.5

      Table 4.4 also shows the biases, MSEs and PREs of the traditional, Rao et al. (2016), Clement and Enang (2016) and proposed estimators using population III. The proposed estimators with the exception of $\overline{y}_{AS3}$, which performed below Clement

and Enang (2016) estimator, have smaller MSEs compared to other estimators. These results are in conformity with that of population in Table 4.2.

Table 4.5. PRE of the Proposed and Some Existing Estimators using Pop. IV

Estimators	Values of α								
	0.5			1.0			1.5		
	Bias	MSE	PRE	Bias	MSE	PRE	Bias	MSE	PRE
$\overline{y}_{st}$	0	0.66	100	0	0.74	100	0	0.83	100
Rao et al. (2016)									
$\overline{y}_{RTK1}$	0	0.32	206.2	0	0.33	224.2	0	0.34	244.1
$\overline{y}_{RTK2}$	0	0.4	165	0	0.43	172.1	0	0.46	180.4
Clement and Enang (2016)									
$\overline{y}_{CE}$	0	0.26	253.8	0	0.26	284.6	0	0.26	319.2
Proposed									
$\overline{y}_{AR1}$	0	0.22	300	0	0.22	336.4	0	0.22	377.3
$\overline{y}_{AS1}$	0	0.31	212.9	0	0.32	231.2	0	0.33	251.5
$\overline{y}_{AR2}$	0	0.22	300	0	0.22	336.4	0	0.22	377.3
$\overline{y}_{AS2}$	0	0.31	212.9	0	0.32	231.2	0	0.34	244.1
$\overline{y}_{AR3}$	0	0.23	287	0	0.23	321.7	0.1	0.23	360.9
$\overline{y}_{AS3}$	0	0.32	206.2	0	0.33	224.2	0	0.34	244.1
Estimators	Values of α								
	2.0			2.5			3.0		
$\overline{y}_{st}$	0	0.92	100	0	1.02	100	0	1.13	100
Rao et al. (2016)									
$\overline{y}_{RTK1}$	0	0.35	262.9	0	0.37	275.7	0	0.39	289.7
$\overline{y}_{RTK2}$	0	0.49	187.8	0	0.52	196.2	0	0.56	201.8
Clement and Enang (2016)									
$\overline{y}_{CE}$	0	0.26	353.8	0	0.26	392.3	0	0.26	434.6
Proposed									
$\overline{y}_{AR1}$	0	0.22	418.2	0	0.22	463.6	0	0.22	513.6
$\overline{y}_{AS1}$	0	0.34	270.6	0	0.36	283.3	0	0.37	305.4
$\overline{y}_{AR2}$	0	0.22	418.2	0	0.22	463.6	0	0.22	513.6
$\overline{y}_{AS2}$	0	0.35	262.9	0	0.36	283.3	0	0.38	297.4
$\overline{y}_{AR3}$	0.1	0.23	400	0.1	0.23	443.5	0.1	0.23	491.3
$\overline{y}_{AS3}$	0	0.36	255.6	0	0.37	275.7	0.1	0.39	289.7

Table 4.5 also shows the biases, MSEs and PREs of the traditional, Rao et al. (2016), Clement and Enang (2016) and proposed estimators using population III. The proposed estimators with the exception of $\overline{y}_{AR3}$ and other estimators are unbiased. The proposed estimators $\overline{y}_{AR1}, \overline{y}_{AR2}, \overline{y}_{AR3}$ performed better compared to other estimators. However, the proposed estimators $\overline{y}_{AS1}, \overline{y}_{AS2}, \overline{y}_{AS3}$, which outperformed Rao et al. (2016) estimators and usual unbiased estimator $\overline{y}_{st}$, performed below the estimator of Clement and Enang (2016).

5. Discussion

Tables 4.2, 4.3, 4.4 and 4.5 report PREs of the sample mean in stratified sampling, Rao et al. (2016), Clement and Enang (2016) and proposed calibration estimators using populations I, II, III and IV (Normal, Chi Square, exponential and gamma distributions) respectively defined in Table 4.1 for different values of $\alpha = (0.5, 1.0, 1.5, 2.0, 2.5, 3.0)$. The results of the PREs reveal that as the values of α (coefficients of linear component of response variable model) increase, the efficiency of the all the estimators increases. The results also revealed that all the proposed estimators have higher PREs compared to their counterparts considered in the study. This implies that the proposed estimators are more efficient in estimation of population mean than other related estimators considered in this study.

6. Conclusion

In this study, we utilized auxiliary information for a heterogeneous population in the form of robust statistical measures based on Gini's mean difference, Downton's method and probability weighted moments. These measures which are not unduly affected by outliers present in the data and provide more efficient estimates of population parameters in the presence of extreme values were used as alternatives for the coefficient of variation used by Rao et al. (2016). From the results of Tables 4.2, 4.3, 4.4 and 4.5, it is observed that in general the estimators proposed under both the calibration schemes are not only robust but more efficient than the usual ratio estimator in stratified sampling, Clement and Enang (2016) and Rao et al. (2016) calibration estimators making them applicable in real life situation when data is somewhat affected by the presence of extreme values. However, the proposed estimators $\overline{y}_{AS1}, \overline{y}_{AS2}, \overline{y}_{AS3}$ performed below the estimator of Clement and Enang (2016) under population IV and generally the efficiency of the proposed estimators is higher when the study variables are characterized by outliers.

References

ADITYA, K., SUD, U., CHANDRA, H., BISWAS, A., (2016). Calibration based regression type estimator of the population total under two stage sampling design. Journal of the Indian Society of Agricultural Statistics, 70, pp.19–24.

BARKTUS, I., PUMPUTIS, D., (2010). Estimation of finite population ratio using calibration of strata weights. In Workshop on survey sampling theory and methodology: August 23–27, 2010, Vilnius, Lithuania, Vilnius.

BREWER, K. R. W., (1995). Combining design based model based inference. Chapter 30 in Business Survey Methods (Eds. B.G. Cox, D.A. Binder, B.N. Chinnapa, A. Christianson, MJ. Colledge and P.S. Kolt). New York:Wiley, 589–606.

BREWER, K. R. W., (1999). Cosmetic calibration with unequal probability sampling. Survey Methodology, 25(2), pp. 205–212.

CLEMENT, E. P., ENANG, E. I., (2016). On the efficiency of ratio estimator over the regression estimator. Communications in Statistics - Theory and Methods, 46(11), pp. 5357–5367.

DEVILLE, J. C., SARNDAL, C. E., (1992). Calibration estimators in survey sampling. Journal of American statistical Association, 87(418), pp. 376–382.

ESTEVAO, V. M., SARNDAL, C. E., (2000). A functional form approach to calibration. Journal of Official Statistics, 16(4), pp. 379–399.

ESTEVAO, V. M., SARNDAL, C. E., (2002). The ten cases of auxiliary information for calibration in two-phase sampling. Journal of Official Statistics, 18(2), pp.233–255.

ESTEVAO, V. M., SARNDAL, C. E., (2006). Survey estimates by calibration on complex auxiliary information. International Statistical Review, 74, pp. 127–147.

HORVITZ, D. G., THOMPSON, D. J., (1952). A generalisation of sampling without replacement from a finite universe. *J.* Amer. Statist. Assoc., 47, pp. 663–685.

KIM, J. M., SUNGUR, E. A., HEO, T. Y., (2007). Calibration approach estimators in stratified sampling. Statistics & probability letters, 77(1), pp.99–103

NIDHI, SISODIA, B. V. S., SINGH, S., SINGH, S. K., (2017). Calibration approach estimation of the mean in stratified sampling and stratified double sampling. Communications in Statistics - Theory and Methods, 46(10), pp. 4932–4942.

RAO, D. K., TEKABU, T., KHAN, M. G., (2016). New calibration estimators in stratified sampling. Asia-Pacific World Congress on Computer Science and Engineering, pp. 66–69.

SALINAS, V. I., SEDORY, S. A., SINGH, S., (2019). Calibrated estimators in two-stage sampling. Communications in Statistics - Theory and Method, 48(6), pp. 1–21

SARNDAL, C.E., WRIGHT, R. L., (1984). Cosmetic form of estimators in survey sampling. Scand.   J. Statist., II, pp. 146–156.

SINGH, S., (2003). Advanced Sampling theory with applications. Dorrdrechi: Kluwer Academic Publisher.

SINGH, A. C., MOHL, C. A., (1996).Understanding Calibration estimators in Survey Sampling. Survey Methodology, 22, pp. 107–111.

SUBZAR, S., MAQBOOL, S., RAJA, T. A., BHAT, M. A., (2018). Estimation of finite population mean in stratified random sampling using non-conventional measures of dispersion. Journal of Reliability and Statistical Studies, 11(1), pp. 83–92.

SUD, U.C., CHANDRA, H., GUPTA, V. K., (2014). Calibration-based product estimator in single- and two-phase sampling. Journal of Statistical theory and Practice, 8(1), pp.1–11.

TRACY, D.S., SINGH, S., ARNAB, R., (2003). Note on calibration in stratified and double sampling. Survey Methodology, 29(1), pp.99–104.

Spatial sampling methods modified by model use [1]

Tomasz Bąk[2]

ABSTRACT

Recent years have seen an intensive development in the field of spatial sampling methods, which generally focus on a balanced distribution of the sample in space. Adaptive sampling methods constitute another dynamic direction in the sampling theory. The issue raised in this article involves the combination of these directions. Five of the commonly known spatial sampling methods have been analysed. The experiment was designed to include statistical model in the sampling procedure. As in the case of adaptive methods, it serves to modify drawing probabilities during sampling. The necessary theory of this sampling modification has been developed and presented. An experiment using artificial data was conducted in order to analyse the efficiency of the model modification in comparison with the primary methods.

Key words: spatial sampling, drawn-by-drawn sampling, kriging, employees distribution

1. Introduction

Spatially balanced samples are samples in which units are well spread throughout the study area. They can be obtained by avoiding or reducing the number of contiguous units. For a long time the main aim in spatial surveys has been to achieve spatially balanced samples (Fattorini et al., 2015).

Wywiał (1996) proposed a sampling design based on the neighbourhood matrix. In this design the number of contiguous units in the sample is reduced using the information about neighbourhoods. After that several spatial designs were proposed (i.e. Bryant et al. (2002)). Stevens Jr and Olsen (2004) introduced the Generalized Random-Tessellation Stratified method (GRTS), which was based on the idea of transformation 2-dimensional space into 1-dimensional space. The Spatially Correlated Poisson Sampling method (SCPS) proposed by Grafström (2012) was an alternative to the GRTS method. It is a drawn-by-drawn sampling method which is a modification of the Correlated Poisson Sampling. Another drawn-by-drawn method, referred to as the Local Pivotal method, was proposed by Grafström et al. (2012). The authors presented two variants of this method, LPM1 and LPM2. Subsequently, the combination of LPM2 and Cube method (proposed by Deville and Tillé (2004)) was proposed by Grafström and Tillé (2013) and it was referred to as the Doubly Balanced Spatial Sampling (DBSS). GRTS, SCPS, LPM1 and LPM2 exploit the information provided by the location of the units in the study area with the purpose of achieving spatially balanced samples. DBSS uses the location in the space and, additionally, the information from auxiliary

---

[2]Department of Statistics, Econometrics, and Mathematics Faculty of Management University of Economics in Katowice, Poland. E-mail: tbak88@wp.pl. ORCID: https://orcid.org/0000-0002-6851-0192.

variables. It leads to a sample which is doubly balanced: spatially balanced and balanced with respect to the auxiliary variables.

Another approach to sampling from the space was proposed by (Thompson and Seber, 1996), who developed an adaptive sampling. In adaptive sampling methods, the sampling design may depend on the values of the survey variable which are observed during sampling. Primarily adaptive sampling designs were divided into two stages of sampling (Thompson and Seber, 1996). In the first stage a 'classical' sampling design, i.e. simple random sampling or stratified sampling is used. In the second stage, some additional elements which fulfilled a certain condition are added to the sample. The second stage is the adaptive part of sampling in the strict sense. Thompson (2006) introduced a new flexible class of adaptive sampling designs - Adaptive Web Sampling. In this method the second stage of sampling is redefined. With a certain probability the adaptive procedure is performed, otherwise a fixed (non-adaptive) design is used.

Model modification of spatial sampling is divided into two stages, as well as the adaptive sampling design. In the first stage it uses the primary sampling method in the way it was defined. In the experiment well-known sampling methods, like simple random sample, GRTS, SCPS, LPM1, LPM2 or DBSS, are considered. In the second stage, just like in the Adaptive Web Sampling, with a certain probability the adaptive procedure is performed, otherwise an initial design is used. For sampling modified by model use, the adaptive procedure is nothing more than rescaling probabilities for some elements and zeroing for others. The choice of whether the element's probability will be rescaled or zeroed depends on the predictions from the spatial interpolation model (i.e. kriging).

The main aim of this paper is to verify the potential of the presented modification. To achieve this, theoretical aspects are introduced in Section 2. Then, in Section 3 the model modification is evaluated using GRTS, SCPS, LPM1, LPM2 and DBSS on artificial data of employees' distribution. The method of data simulation was consistent with the real spatial characteristics of the employee population observed in the research. It should be mentioned that the approach to sampling from the space discussed in this paper is, to a certain extent, the development of the triangular method of spatial sampling (Bąk, 2014). It allows to stabilize the selected elements around a predetermined value.

Model modification of spatial sampling methods can be also applied to other spatial design-based sampling methods than the five chosen for the experiment. The theory presented in Section 2 can be simply adapted to other design-based sampling methods. Therefore this method can expand the range of design-based methods that can be used in research. It may prove useful in planning sample selection in some spatial studies.

2. Construction of spatial sampling modified by model use

Let us consider a finite population of N spatial locations in the study area, labelled by $1, \ldots, N$. Moreover denote by y_i the value of the characteristic Y under study and by x_i the value of an auxiliary variable X well correlated with Y corresponding to location $i = 1, \ldots, N$. Values of the Y characteristic which are observed during sampling are fixed but they are treated as the outcome of a random process just to perform a model-dependent

prediction of these values:

$$\{Y(i) : i \in \{1,\ldots,N\}\}. \tag{1}$$

Two stages of spatial sampling modified by model use may be distinguished, as in the case of adaptive sampling. In the first stage an initial sample $\{i_1,\ldots,i_{n_0}\}$ consisted of n_0 elements is selected. Each of n_0 sampled elements is selected with drawing probabilities:

$$p = [p_1,\ldots,p_N]. \tag{2}$$

Furthermore, n_0 realizations of the Y process are obtained. Denote them by $y_{i_1},\ldots,y_{i_{n_0}}$. It means that for initial sample elements both values and locations are known. Therefore, it is possible to build statistical model to predict values of Y process for the whole population. Denote by $\hat{Y}_{n_0}(i)$, $i \in \{1,\ldots,N\}$ the predictions of the values of Y process which are based on the initial sample. These predictions modify drawing probabilities for next m elements. Then, the new predictor $\hat{Y}_{n_0+m}(i), i \in \{1,\ldots,N\}$ is constructed and then used to modify probabilities for elements $n_0 + m + 1, ..., n_0 + 2m$.

Let us consider that the predictor $\hat{Y}_{n_0+(k-1)m}(i), i \in \{1,\ldots,N\}$ was constructed. Then, the sampling can be made in one of two ways: depending on the predictor $\hat{Y}_{n_0+(k-1)m}(i)$ or in the same way as selection of the elements $1,\ldots,n_0$ - using drawing probabilities (2). In other words a mixture of schemes is used (Thompson, 2006). Next, let us introduce the probability d_{k-1}. Then, the condition that determinates the sampling scheme choice in the k-th drawing is:

$$\exists_{i \in \{\{1,\ldots,N\}\setminus\{i_1,\ldots,i_{n_0+(k-1)m}\}\}} f(x_i,\hat{y}_i,\bar{x},\bar{\hat{y}}) \in A, \tag{3}$$

where $\hat{y}_i$, $i \in \{1,\ldots,N\}$ are predictions of Y characteristic from the model $\hat{Y}_{(k-1)m}(i)$, $i \in \{1,\ldots,N\}$, $\bar{\hat{y}}$ is the average value of those predictions and A is the subset of the range of f function. The construction of the function f is crucial for the efficiency of presented solution. For the purposes of description of the construction of spatial sampling modified by model use, we will limit ourselves to the general form of the f function. Some examples will be presented in the next section.

Depending on the satisfiability of the condition (3), the k-th drawing is conducted in one of two ways:

- Condition (3) is false. Then, a sampling method analogous to the one that has been used in the initial sampling stage is used. Drawing probabilities for k-th element are defined as $p_k = p$.

- Condition (3) is true. Then, two situations are possible. With probability $1 - d_{k-1}$, the k-th element is sampled in the same way as when condition (3) is false. Otherwise, with probability d_{k-1}, the k-th element is sampled among the elements of the set

$$H_{k-1} = \{i \in \{1,\ldots,N\}/\{i_1,\ldots, \quad i_{k-1}\}: \quad f(x_i,\hat{y}_i,\bar{x},\bar{\hat{y}}) \in A\}. \tag{4}$$

It means that the next element could be selected only among elements fulfilling the

condition (3). Then, vector $p'_k = \left[p'_{k,1}, \ldots, p'_{k,N} \right]$ is defined as follows:

$$
p'_{k,i} =
\begin{cases}
\frac{p_{k,i}}{\sum_{i \in H_{k-1}} p_{k,i}}, & \text{when } i \in H_{k-1}, \\
0, & \text{when } i \notin H_{k-1}.
\end{cases}
\tag{5}
$$

In other words, probabilities in p'_k vector are proportional to probabilities in p_k for the elements of the set H_{k-1} and equal to 0 for other elements of the population. The probabilities p'_k are used to sample the k-th element.

As we can see, the sampling plan which was used at the initial stage is used at the second stage of sampling too. Moreover, probabilities p'_k, $k = n_0 + 1, \ldots, N$ are based on the probabilities p_k, $k = n_0 + 1, \ldots, N$. Therefore, the choice of initial sampling has great impact on the spatial sampling modified by model use.

It should be emphasized that the construction of vector (5) implies that sampling without replacement is considered. However, spatial sampling modified by model use can be easily transformed into sampling with replacement. However, it was not done in this paper in order to keep the structure of the paper more transparent.

The basic feature of the d_k, $k = n_0, \ldots, n-1$ is that the higher the value of d_k, the greater 'adaptability' (ability to learn on already sampled elements) of the sampling scheme. On the other hand, the precision of the 'adaptability' is based on the precision of the model, which is mainly conditioned by the number of already sampled elements. It is well known that the precision of the spatial model increases with increasing sample size (Cressie, 1993). Therefore, a sequence of probabilities $d_k, k = n_0, \ldots, n-1$ should be increasing. In principle, the same assumption about the sequence $d_k, k = n_0, \ldots, n-1$ was made by Thompson (2006) in the Adaptive Web Sampling.

The sampling plan can be defined using the probabilities defined above. Let us denote final sample by $s = \{i_1, i_2, \ldots, i_n\}$. Then, the sampling plan is defined as:

$$
P(s) = \sum_{\{j_1, \ldots, j_n\} \in S(s)} \prod_{m=1}^{n_0} p_{m, j_m}
\tag{6}
$$

$$
\prod_{k=n_0+1}^{n} \left[P(\bar{\bar{H}}_{k-1} = 0) p_{k, j_k} + P(\bar{\bar{H}}_{k-1} \neq 0) \left(p'_{k, j_k} d_{k-1} + p'_{k, j_k} (1 - d_{k-1}) \right) \right],
$$

where $S(s)$ is a set of all permutation of s and $\bar{\bar{H}}_k$ is cardinality of the H_k set. The sampling plan (6) is conditioned by probabilities $P\left(\bar{\bar{H}}_k \neq 0 \right)$ and $P\left(\bar{\bar{H}}_k = 0 \right)$. In other words, the sampling plan is primarily defined by model predictions on unsampled elements.

The use of model modification results in unequal first-order probabilities of inclusion. They depend on the results of modelling and cannot be defined explicitly. Fattorini (2006) proposed the method of using Horvitz-Thompson estimator in the case when an explicit derivation of the first-order probabilities of inclusion is prohibitive. This approach was later developed in several papers (Thompson and Wu, 2008; Gamrot, 2014).

Let us assume that M samples from the population $\{1, \ldots, N\}$ are selected independently and by repeating the same rules. An invariably positive estimator of the first-order

probabilities of inclusion π_j, $j = 1, \ldots, N$ is

$$\hat{\pi}_j = \frac{m_j + 1}{M + 1}, \quad j = 1, \ldots, N, \tag{7}$$

where m_j is the total number of samples in which the j-th element was drawn. Since $M \to \infty$, then the asymptotically unbiased modification of the primary Horvitz-Thompson estimator is

$$\hat{T} = \sum_{j=1}^{n} \frac{y_j}{\hat{\pi}_j}. \tag{8}$$

Another, much simpler method of first order probabilities of inclusion will be evaluated too. First order probabilities of inclusion will be proportional to the auxiliary variable X:

$$\pi_{X_j} = \frac{x_j}{\sum_{i=1}^{N} x_i}, \quad j = 1, \ldots, N. \tag{9}$$

3. Example of spatial sampling modified by model use

Let us consider a spatial research of the average number of employees by district. Elements under study are districts of a region. An additional characteristic observed during research is the number of inhabitants. Artificially generated data were prepared to illustrate the usefulness of spatial sampling modified by model use in such research. In the first step the X matrix of the size 200×200 was generated. It contains simulated numbers of inhabitants. Each element of this matrix was sampled using one of three normal distributions:

$$\sim \begin{cases} \mathcal{N}(2500, 120^2), & \text{when } (30 \le i \le 55 \text{ or } 155 \le i \le 190) \text{ and } 170 \le j \le 200, \\ \mathcal{N}(1500, 80^2), & \text{when } (1 \le i \le 30 \text{ or } 165 \le i \le 200) \text{ and } 1 \le j \le 45, \\ \mathcal{N}(2000, 100^2), & \text{in other cases,} \end{cases} \tag{10}$$

where i and j are row and column index in X matrix respectively. Then, the Y matrix of the size 200×200 was generated. Elements of the Y matrix are simulated values of the number of the employees. Values of the Y process were simulated in such a way that the correlation between Y and X was high (Pearson correlation coefficient equal to 0.928). Moreover, both matrices are identified with two-dimensional space $X_1 \times X_2$ size of $[0, 200] \times [0, 200]$. Each element of both matrices is related to the fragment of the two-dimensional space. The size of each fragment is 1×1 unit.

In practice, the spatial distribution of the employees, agglomerations is not as regular as Y matrix (Combes and Overman, 2004; Glaeser and Kerr, 2009). Therefore, from the two-dimensional space 300 fragments were sampled using uniform distribution. These 300 elements were treated as the population under study. Each population element is related to an appropriate element (realization of the random process) of X and Y matrices.. The average value of X characteristic in the final population was equal to 1990.794 inhabitants. It is assumed that the average value of X is known before sampling, but without the knowledge

of the value of a X variable in specific elements. In the case of Y characteristic the average value was equal to 501.477 employees and is not known before sampling. The spatial distribution of the population under study is shown in Figure 1. The distribution of X and Y characteristics in the population is shown in Figure 2 and Figure 3 respectively. For both characteristics two aggregations of lower values and two aggregations of higher values can be observed. The purpose of these aggregations is to simulate spatial heterogeneity.

Spatial sampling modified by model use was used to draw a sample from this population. Five sampling methods were considered as the initial sampling method. Three of them were: Spatially Correlated Poisson Sampling method (SCPS) and both Local Pivotal methods (LPM1 and LPM2). Each of them exploits the information about spatial location of the population elements. The fourth of the methods was Cube method (CM), which is an example of balanced sampling, and the fifth was Unequal Probability Sampling (UPS). The last two methods use the information from auxiliary variables. The first-order inclusion probabilities were defined proportionally to the auxiliary variable for each initial sampling methods.

Ordinary kriging was chosen as a spatial data modeling method. The kriging model was built using *automap* package in R language (Hiemstra et al., 2008). Already sampled elements were used to variogram estimation. Spherical, exponential, Gaussian and Matern family models were considered as potential shape of the variogram. Finally, the sampling algorithm picked the variogram model that has the smallest residual sum of squares and used it in kriging modelling.

Kriging refers to making inferences on unobserved values of random process $X(i) : i \in \{1,\dots,N\}$ from data (Cressie, 1993)

$$\{x_{i_1},\dots,x_{i_n}\} \tag{11}$$

observed at n known spatial locations

$$\{i_1,\dots,i_n\}. \tag{12}$$

Ordinary kriging refers to spatial predictor $\hat{X}$, which fulfils the following two assumptions:

$$1.\ X(i) = \mu + \varepsilon(i), \tag{13}$$

where $i \in \{1,\dots,N\}$, $\varepsilon(i)$ is the error process with the expected value equal to 0 and μ is unknown.

$$2.\ \hat{X}(i) = \sum_{j=1}^{n} \lambda_j x_{i_j}, \tag{14}$$

where $\sum_{j=1}^{n} \lambda_j = 1$ and $i \in \{1,\dots,N\}$. Weights $\lambda_j, j = 1,\dots,n$ are determined by Lagrange multipliers so as to minimize the mean square error of the $\hat{X}$ predictor. As a result, the best linear unbiased predictor is obtained (Cressie, 1993).

Condition (3) was defined generally. To make it easier to interpret and use, let us con-
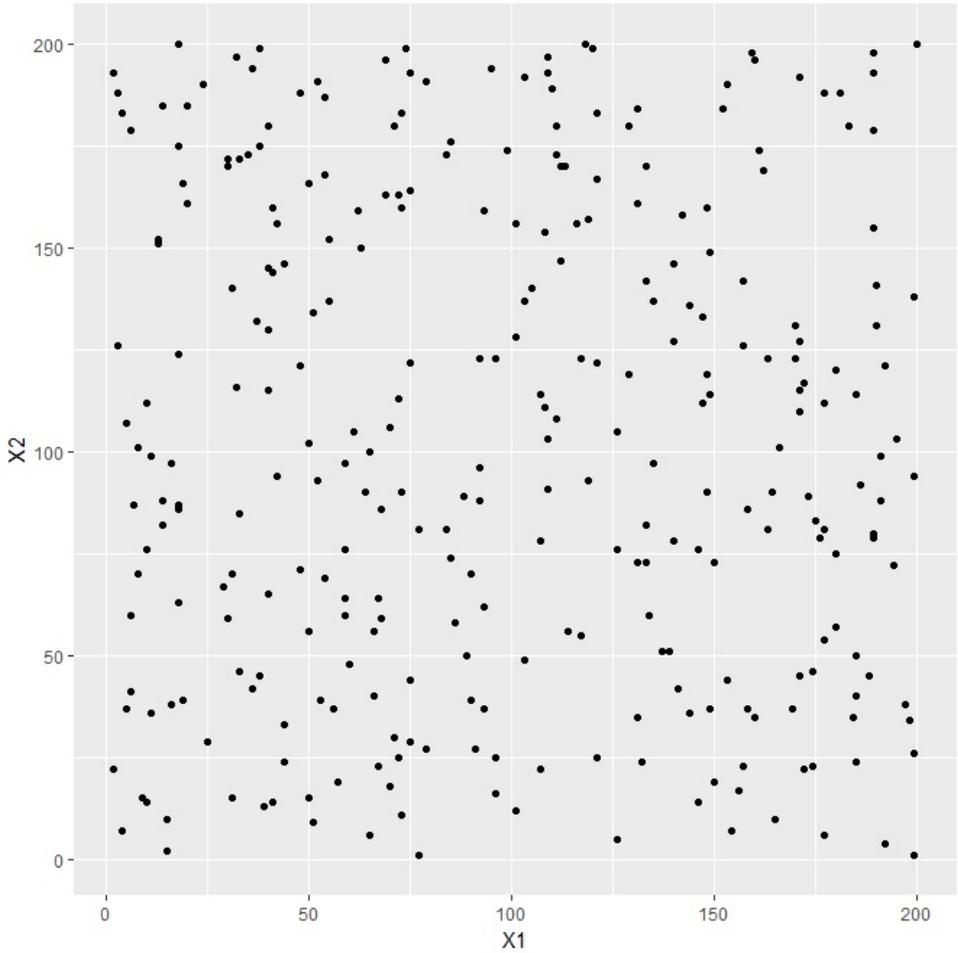
Figure 1: Population distribution.
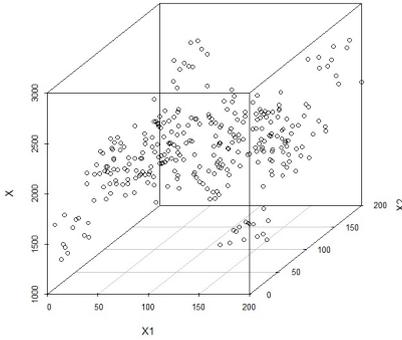
Figure 2: X characteristic distribution.
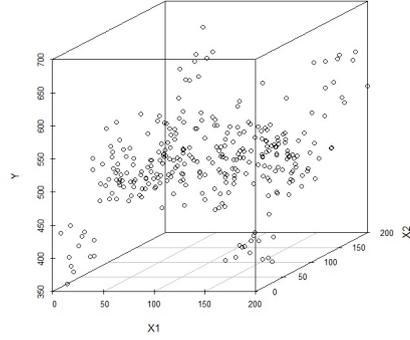
Figure 3: Y characteristic distribution.

sider $f(x_i, \hat{y}_i, \bar{x}, \bar{\hat{y}}) = \|\hat{x}_i - \bar{x}\|$ and $A = [0, c]$. Then, condition (3) can be represented as:

$$\|\hat{x}_i - \bar{x}\| \le c, \tag{15}$$

and can be explained as a tendency to prefer elements for which the value of X is close to the average.

The sample size of $n = 100$, and the initial sample size of $n_0 = 50$ were considered. The sequence $\{d_k\}, k = 50, \ldots, n-1$ was defined as

$$d_k = \frac{k}{100}, k = 50, \ldots, 99. \tag{16}$$

The value of c coefficient was equal to 5, which was about 25% of standard deviation of X. Hereby, increased probabilities of the values of X characteristic which are close to the average value $\bar{x}$ was achieved. For each of five initial sampling methods, sampling of 100 elements was repeated 10 000 times to achieve, through the Monte Carlo method, estimation of the first-order inclusion probabilities.

Two different approaches to the first-order probabilities of inclusion definition were considered. The first was the empirical approach (7), which was based on Monte Carlo simulation. In this approach, the first-order probability of inclusion tends to increase when the $|X - \bar{x}|$ variable decreases. In the second approach, the first-order probabilities of inclusion were proportional to the auxiliary variable (9). For both approaches and each type of initial sampling method 1 000 samples were selected.

Results of spatial sampling modified by model use were compared to primary forms of sampling methods (the one which were used as initial sampling methods). Therefore, 1 000 samples, each consisting of 100 elements, were selected using all five sampling methods used for initial sampling. Estimation was based on the first-order probabilities of inclusion proportional to the auxiliary variable.

For each sampling method the Horvitz-Thompson estimators were calculated. Effi-

ciency of spatial sampling modified by model use was verified using rRMSE of the estimator of the mean of Y. Results are presented in Table 1.

Table 1: rRMSE for different sampling methods, c=5.

Method	Spatial sampling modified by model usage - Probabilities of inclusion based on Monte Carlo	Spatial sampling modified by model usage - Probabilities of inclusion proportional to X variable	Spatial sampling method in primary form
CM	0.467%	0.298%	0.371%
LPM1	0.489%	0.306%	0.279%
LPM2	0.464%	0.275%	0.280%
SCPS	0.478%	0.276%	0.324%
UPS	0.494%	0.296%	0.310%

rRMSE for estimators based on the Monte Carlo first-order probabilities of inclusion was significantly higher than for the other two approaches. Spatial sampling modified by model use with the first-order probabilities of inclusion proportional to the auxiliary variable delivered lower rRMSE than the primary form for all methods except LPM1.

Simulation was repeated for c equal to 10, 15, 20 and 25. Results are presented in Tables from 2 to 5 respectively.

Table 2: rRMSE for different sampling methods, c=10.

Method	Spatial sampling modified by model usage - Probabilities of inclusion based on Monte Carlo	Spatial sampling modified by model usage - Probabilities of inclusion proportional to X variable	Spatial sampling method in primary form
CM	0.541%	0.299%	0.371%
LPM1	0.512%	0.276%	0.279%
LPM2	0.547%	0.267%	0.280%
SCPS	0.559%	0.285%	0.324%
UPS	0.491%	0.296%	0.310%

Table 3: rRMSE for different sampling methods, c=15.

Method	Spatial sampling modified by model usage - Probabilities of inclusion based on Monte Carlo	Spatial sampling modified by model usage - Probabilities of inclusion proportional to X variable	Spatial sampling method in primary form
CM	0.507%	0.302%	0.371%
LPM1	0.522%	0.342%	0.279%
LPM2	0.494%	0.270%	0.280%
SCPS	0.483%	0.273%	0.324%
UPS	0.477%	0.295%	0.310%

The results show that for Cube method, Spatially Correlated Poisson Sampling method and Unequal Probability Sampling model modification increases sampling efficiency (in terms of rRMSE reduction) and it is independent from the c value. Modification of UPS method delivered quite stable rRMSE values for different c values. Model modification of SCPS and CM methods had generally higher rRMSE values for higher c values. LPM1

Table 4: rRMSE for different sampling methods, c=20.

Method	Spatial sampling modified by model usage - Probabilities of inclusion based on Monte Carlo	Spatial sampling modified by model usage - Probabilities of inclusion proportional to X variable	Spatial sampling method in primary form
CM	0.485%	0.336%	0.371%
LPM1	0.549%	0.286%	0.279%
LPM2	0.538%	0.322%	0.280%
SCPS	0.497%	0.316%	0.324%
UPS	0.465%	0.298%	0.310%

Table 5: rRMSE for different sampling methods, c=25.

Method	Spatial sampling modified by model usage - Probabilities of inclusion based on Monte Carlo	Spatial sampling modified by model usage - Probabilities of inclusion proportional to X variable	Spatial sampling method in primary form
CM	0.452%	0.331%	0.371%
LPM1	0.507%	0.326%	0.279%
LPM2	0.408%	0.304%	0.280%
SCPS	0.472%	0.309%	0.324%
UPS	0.441%	0.293%	0.310%

modified by model use achieved a better rRMSE value than primary LPM1 only for $c = 10$ and the difference was negligible. As a rule, model modification was inefficient for this sampling method. In the case of LPM2, model modification delivered better efficiency for low values of c parameter. However, the gain was smaller than the ones obtained on CM, SCPS and UPS model modifications.

4. Conclusions

The choice of modelling method determines the efficiency of the presented sampling modification. The precision of the model is a key to achieve the main advantage of this sampling modification - stabilization of sampling elements around the global average value of the additional variable. In the presented example a very basic type of model - ordinary kriging, was chosen. It did not give a full picture of the influence of modelling method on sampling efficiency. The aim of the example was, however, to evaluate the potential of the presented modification rather than in-depth analysis of how changes in values of different parameters (n_0, m, d_k) impact the rRMSE on different populations. Such an analysis would require rather a more multi-aspect approach than the one used in the above example and need more research.

The aim of this paper was to evaluate the potential of model modification. Considering the results of the experiment in which the model modification often improved the quality of the underlying method, it can be concluded that model modification can also be used for other design-based methods. From a theoretical point of view, the presented solution can easily be translated into other methods. As a results, we obtained quite an effective method, which expands the range of design-based methods, which can be used in spatial research.

Spatial sampling modified by model use requires ongoing access to statistical program which allows to construct kriging or other spatial models. This requirement could be fulfilled in two ways: by using mobile devices with the access to statistical software or by sending the information about sampled elements to a PC which works as a computational station. Both solutions increase the cost of research. However, both solutions should be treated as long-term investments in research equipment. Then, learning on the elements selected to the sample could be ongoing. Generally, real-time observation and analysis of the sample seems to be an interesting direction in the development of sampling methods.

The presented sampling modification gives a possibility to adjust the sampling method to the analysed population and its different characteristics. Adjustment could be introduced quite straightforwardly, by changes in the coefficients of spatial sampling modified by model use. It could also have other, more complex aspects, such as definition of probabilities of inclusion or modelling method choice. One can also think about further modification, which have not been discussed yet. One of more interesting in the author's opinion is to substitute $\bar{x}$ value of the additional variable, which is known before sampling, by an average value $\bar{y}$ of the characteristic under study, which is calculated during sampling process. After this change the sampling could be conducted without using the additional variable.

References

Bąk, T., (2014). Triangular method of spatial sampling. *Statistics in Transition*, Vol. 15/2, pp. 9–22.

Bryant, D., Chang, Y., Rodger, C., and Wei, R., (2002). Two-dimensional balanced sampling plans excluding contiguous units. *Communications in Statistics - The- ory and Methods*, Vol. 31/8, pp. 1441–1455.

Combes, P.-P., Overman, H., (2004). The spatial distribution of economic activities in the European union. *Handbook of Regional and Urban Economics*, Vol. 4, pp. 2845–2909.

Cressie, N., (1993). Statistics for Spatial Data. *Wiley*.

Deville, J.-C., Tillé, Y., (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91(4), pp. 893–912.

Fattorini, L., (2006). Applying the Horvitz-Thompson criterion in complex designs: a computer-intensive perspective for estimating inclusion probabilities. *Biometrika*, 93, pp. 269–278.

Fattorini, L., Corona, P., Chirici, G., and Pagiarella, C., (2015). Design-based strategies for sampling spatial units from regular grids with applications to forest sur- veys, land use, and land cover estimation. *Environmetrics*, 26, pp. 216–228.

Gamrot, W., (2014). Estimators for the Horvitz-Thompson statistic based on some posterior distributions. *Mathematical Population Studies*, 21(1), pp. 12–29.

Glaeser, E., Kerr, W., (2009). Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain? *Journal of Economics & Management Strategy*, 18, pp. 623–663.

Grafström, A., (2012). Spatially correlated poisson sampling. *Journal of Statistical Planning and Inference*, 142, pp. 139–147.

Grafström, A., Lundström, P., and Schelin, L., (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68, pp. 514–520.

Grafström, A., Tillé, Y., (2013). Doubly spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics*, 24, pp. 120–131.

Hiemstra, P., Pebesma, E., Twenhöfel, C., and Heuvelink, G., (2008). Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network. *Computers & Geosciences*.

Stevens Jr, D., Olsen, A., (2004). Spatially balanced sampling of natural re- sources. *Journal of the American Statistical Association*, 99, pp. 262–278.

Thompson, M., Wu, C., (2008). Simulation-based randomized systematic PPS sampling under substitution of units. *Survey Methodology*, 34, pp. 3–10.

Thompson, S., (2006). Adaptive web sampling. *Biometrics*, Vol. 62, No. 4, pp. 1224–1234.

Thompson, S., Seber, G., (1996). Adaptive Sampling. *John Wiley & Sons, Inc*.

Wywiał, J., (1996). On space sampling. *Statistics in Transition*, Vol.7, No. 7, pp. 1185–1191.

Small area estimates of the low work intensity indicator at voivodeship level in Poland

Łukasz Wawrowski[1], **Maciej Beręsewicz**[2]

ABSTRACT

The EU Statistics on Income and Living Conditions (EU-SILC) has provided annual estimates of the number of labour market indicators for EU countries since 2003, with an almost exclusive focus on national rates. However, it is impossible to obtain reliable direct estimates of labour market statistics at low levels based on the EU-SILC survey. In such cases, model-based small area estimation can be used. In this paper, the low work intensity indicator for the spatial domains in Poland between 2005-2012 was estimated. The Rao and You (1994), Fay and Diallo (2012), and Marhuenda, Molina and Morales (2013) models were applied. The bootstrap MSE for the discussed methods was proposed. The results indicate that these models provide more reliable estimates than direct estimation.

Key words: EU-SILC, low work intensity, labour market, small area estimation, temporal models, spatio-temporal models.

1. Introduction

Sample surveys conducted by National Statistical Institutes (NSIs) are in most cases representative at the national or region level (in particular at NUTS 1 level). In more detailed domains, such as states/voivodeships (NUTS 2) or subregions (NUTS 3), a small sample size does not allow for obtaining precise and accurate estimates of socio-economic indicators. Therefore, one needs methods that may provide more reliable estimates. For that purpose small area estimation (SAE) is often used. SAE makes it possible to estimate characteristics even if the sample is small, direct estimation is not reliable or domains are not observed in the sample. The underlying idea of SAE is to account for random effects in studied domains and „borrow strength" from auxiliary variables, over time or in space.

Small area estimation methods are widely used in many statistical domains. Social sciences examples include the labour market (López-Vizcaíno, Lombardía and Morales 2013), poverty (Molina and Rao 2010; Szymkowiak, Młodak and Wawrowski 2017) and business statistics (Chandra, Chambers and Salvati 2012; Dehnel and Wawrowski 2020). Due to limited access to data many applications cover estimation for only one year.

The main goal of the study described in this article was to estimate the low work intensity indicator (LWI) in the domains defined by the level of voivodeships (NUTS 2) in

---

[1]Computer Science Research Centre, Research Network Łukasiewicz - Institute of Innovative Technologies EMAG, Poland. E-mail: lukasz.wawrowski@emag.lukasiewicz.gov.pl. ORCID: https://orcid.org/0000-0002-1201-5344.

[2]Department of Statistics, Poznań University of Economics and Business. Statistical Office in Poznań, Centre for Small Area Estimation, Poland. E-mail: maciej.beresewicz@ue.poznan.pl. ORCID: https://orcid.org/0000-0002-8281-4301.

Poland between 2005 and 2012 with acceptable precision measured by coefficient of variation (CV). The LWI indicator, at-risk-of-poverty and material deprivation indicators are required by Eurostat as part of Europe 2020 strategy. The current official information is available only at the national and the NUTS 1 level in Poland and other EU states. At the more detailed domains small sample size results in big variances of obtained estimates. To achieve the main goal we discuss three recent small area models — Rao and You (1994), Fay and Diallo (2012), and Marhuenda, Molina and Morales (2013) — and then apply them. The first two models take into account temporal effects, while the third also incorporates spatial effects.

The article has the following structure. First, we present the variable of interest — the low work intensity indicator. The third section provides the notation for direct and model-based estimation. We also calculate mean square error (MSE) and model diagnostics, and present Generalized Inflation Factors in the context of SAE. The fourth section describes the EU-SILC survey and data from 2005 to 2012. In the fifth section we present the results and model diagnostics. The article ends with a summary.

2. Low work intensity

2.1. EU-SILC survey

The survey to collect EU Statistics on Income and Living Conditions (EU-SILC) was launched in 2003. The main aim of the survey was to deliver comparable data about income, poverty and living conditions of households in EU Member States. EU-SILC data are collected using a questionnaire in face-to-face interviews covering demography, education, health, housing conditions, economic activity, and more importantly, the level and sources of household incomes. EU-SILC is a sample-based, representative survey, in which a household is the basic statistical unit. In addition, every household member above 16 is also surveyed.

Various social cohesion indicators are estimated based on the EU-SILC survey. Several of them are used to monitor Europe 2020 strategy and to calculate the fraction of people living in households with very low work intensity (Statistics Poland 2014).

2.2. Low work intensity indicator

According to Eurostat, "the indicator of persons living in household with low work intensity is defined as the number of persons living in a household where work intensity is below a threshold set at 20%". Intensity of work is defined as the number of months that all working age household members (aged between 18 and 64) worked during the reference year divided by the total number of months that could theoretically be worked within the household. This means that households with low work intensity caused by different factors do not utilize their available time for work. Time spent at work is defined by Eurostat as:

- months in paid employment (full-time or part-time),
- paid internships and trainings,
- self-employment, with or without employees,
- unpaid work in a family business (helping family members).

To calculate the low work intensity indicator the total number of hours worked per week for each respondent is computed. For part-time employees, "the number of months in terms of full time equivalents is estimated on the basis of the number of hours usually worked at the time of the interview" (Mélina and Emilio, 2012). Eurostat set a threshold for the low work intensity at the level of 20%. This value refers to the expected risk of poverty in households with low work intensity. Nevertheless, Ward and Ozdemir (2013) argued that the threshold should be set slightly higher. Equation (1) presents the last stage of calculating the low work intensity indicator for the domains.

$$\text{LWI}_{dt} = \frac{\sum_{i=1}^{n_{dt}} I(WI_{i,dt} < 0.2) d_{i,dt}}{\sum_{i=1}^{n_{dt}} d_{i,dt}}, \tag{1}$$

where: $WI_{i,dt}$ is work intensity of i-th household member in d-th domain at time t, $d_{i,dt}$ is calibrated weight of i-th household member, $I(\bullet)$ is an indicator function with two values $\{0,1\}$.

3. Notation for estimators and diagnostics

The classic Fay and Herriot (1979) area-level model does not take into account temporal nor spatial random effects. Therefore, when a panel survey data are used for estimation, the correlation between years should not be neglected. Thus, for the purpose of estimating LWI we applied two area-level small area estimators that take into account temporal random effects (Rao and You, 1994; Fay and Diallo, 2012) and spatio-temporal random effects (Marhuenda, Molina and Morales, 2013) for NUTS 2 level. The motivation for choosing these estimators is the observed strong temporal effect for NUTS 2 (voivodeships) in Poland. In addition, we would like to verify whether including the spatial effect in the model leads to better estimates.

3.1. Direct estimator

Let $\Omega = \{1, \ldots, N\}$ denote the target population of size N. From this population we draw a sample according to the sampling scheme $s \subseteq \Omega$ of size n. Let Ω_{dt} denote target population in domain (e.g. area), $d = 1, ..., D$ denote a domain and $t = 1, ..., T$ denote the time when the survey was conducted. Next, π_{dti} denotes the inclusion probability of i-th unit in d-th domain at time t in the corresponding domain sample s_{dt} and $d_{dti} = \pi_{dti}^{-1}$ the corresponding sampling weight. The EU-SILC survey uses the calibration approach proposed by Deville and Särndal (1992) to account for non-response. Thus, $w_{dti} = \lambda_{dti} d_{dti}$ denotes a calibration weight and λ_{dti} is the scaling factor for sampling weights d_{dti}. Let y denote the target variable (low work intensity) defined as follows:

$$y_{dti} = \begin{cases} 1 & \text{if the household suffers from low work intensity,} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Therefore, a design-unbiased direct estimator of $\bar{y}_{dt}$ is the Horvitz-Thompson (HT) estimator for the subpopulation Ω_{dt}, given by:

$$\hat{\bar{y}}_{dt} = \sum_{i \in s_{dt}} w_{dti} y_{dti} / \sum_{i \in s_{dt}} w_{dti}. \tag{3}$$

Because NUTS 2 level was used for stratification, the variance of $\hat{\bar{y}}_{dt}$ was estimated using a nonparametric bootstrap method as follows. Separately for each time t according to the sampling scheme, in particular taking into account strata $h = 1, ..., H$, draw a sample with replacement B times. For each sample b calculate the bootstrapped weight defined by the equation (4):

$$w_{dti}^{b} = w_{dti} \frac{n_{h,dt}}{n_{h,dt} - 1} m_{dti}^{b}, \tag{4}$$

where $n_{h,dt}$ denotes the number of sampled units in stratum h, domain d, at time t in the original sample and m_{dti}^{b} denotes the number of times that i-th unit was included in sample b. Finally, the bootstrap estimator of the variance of $\hat{\bar{y}}_{dt}$ for the domain Ω_{dt} is derived by:

$$\hat{V}(\hat{\bar{y}}_{dt}) = \hat{\psi}_{dt} = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\bar{y}}_{dt}^{b} - \hat{\bar{y}}_{dt})^{2}, \tag{5}$$

where $\hat{\bar{y}}_{dt}^{b} = \sum_{i \in s_{dt}} w_{dti}^{b} y_{dti} / \sum_{i \in s_{dt}} w_{dti}^{b}$. For the sake of clarity, we will use ψ_{dt} for the known sampling variance.

3.2. Rao and Yu (1994) model

Rao and You (1994) proposed an extension of Fay and Herriot (1979) model, which accounts for domains defined as time-series and cross-sectional classification. The model assumes two random effects — the domain effect, which is constant in time, and autocorrelation of domain effects in time. The autocorrelation is assumed to be the same between domains.

To enable comparison, we will apply the notation used in Marhuenda, Molina and Morales (2013). Therefore, in the first stage, Rao and You (1994) model assumes the following sampling model:

$$\bar{y}_{dt} = \mu_{dt} + e_{dt} \tag{6}$$

where $e_{dt} \overset{ind.}{\sim} N(0, \psi_{dt})$, where ψ_{dt} is the known sampling variance. The second stage (the linking model) μ_{dt} is assumed to follow a linear mixed model given by:

$$\mu_{dt} = X_{dt}' \beta + u_{1d} + u_{2dt} \tag{7}$$

where X_{dt} is the matrix of auxiliary variables (fixed effects), $u_{1d} \overset{ind.}{\sim} N(0, \sigma_1^2)$ denotes the random effect for domain at time $t = 1$ and constant in time $u_{1d} = u_{1d,t=1} = u_{1d,t=2} = ... = u_{1d,t=T}$. The second random component denoted by u_{2dt} is assumed to follow the autoregressive process $AR(1)$ with σ_2^2 and ρ_2, and is given by:

$$u_{2dt} = \rho_2 u_{2d,t-1} + \varepsilon_{2dt}, \tag{8}$$

where $|\rho_2| < 1$ and $\varepsilon_{2dt} \overset{ind.}{\sim} N(0, \sigma_2^2)$. We use ρ_2 for autocorrelation to be consistent with the number of random effects and for the consistency with the other models. In addition, let $\theta = (\sigma_1^2, \sigma_2^2, \rho_2)'$ be the vector of unknown parameters involved in the covariance structure of the model. Finally, the BLUP estimator of $\bar{y}_{dt}$ obtained by Rao and You (1994) through the method of moments is given by:

$$\mu_{dt} = X'_{dt}\tilde{\beta} + (\sigma_1^2 1'_T + \sigma_2^2 \gamma_T)(\Sigma_d + \sigma_2^2 \Gamma + \sigma_1^2 1_T 1'_T)^{-1}(y_d - X_d\tilde{\beta}), \tag{9}$$

where, for simplicity, we use $u_{1d} = u1$, $u_{2dt} = u2$ and $\rho_2 = \rho$,

- Γ is a symmetric matrix $T \times T$ with elements $\rho^{|i-j|}/1(-\rho^2)$,
- $V_d = \Sigma_d + \sigma_2^2\Gamma + \sigma_1^2 1_T 1'_T = Cov(y_d)$,
- $V = diag(V_d) = Cov(y)$,
- $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$,

When $\hat{\theta} = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\rho}_2)$ is known, the EBLUP is given by

$$\hat{\mu}_{dt} = X'_{dt}\tilde{\beta} + (\hat{\sigma}_1^2 1'_T + \hat{\sigma}_2^2 \hat{\gamma}_T)(\hat{\Sigma}_d + \hat{\sigma}_2^2 \hat{\Gamma} + \hat{\sigma}_1^2 1_T 1'_T)^{-1}(y_d - X_d\tilde{\beta}). \tag{10}$$

The notation of EBLUP (10) can be simplified in equation (7) and is given by (11). Moreover, rewriting the equation (10) as (7) enables comparison with Marhuenda, Molina and Morales (2013) and specifies that the model can be estimated using the Henderson (1975) approach.

$$\hat{\mu}_{dt} = X'_{dt}\beta + \hat{u}_{1d} + \hat{u}_{2dt}. \tag{11}$$

3.3. Fay and Diallo (2012) model

Another extension of Fay and Herriot (1979) was proposed by Fay and Diallo (2012) and Fay, Planty and Diallo (2013). Fay and Diallo (2012) proposed a univariate and Fay, Planty and Diallo (2013) a multivariate dynamic small area model that takes into account auto-correlation of random effects for domains. The Fay and Diallo (2012) model also extends Rao and You (1994) by assuming nonstationarity of the domain effect, thus the effect is not constant over time. Fay and Diallo (2012) in the first stage assume a sampling model given by:

$$\bar{y}_{dt} = \mu_{dt} + e_{dt} \tag{12}$$

where $e_{dt} \overset{ind.}{\sim} N(0, \psi_{dt})$, where ψ_{dt} is known sampling variance. The second stage (the linking model) assumes a linear mixed model given by the following equation:

$$\mu_{dt} = X'_{dt}\beta + u_{1dt} + u_{2dt} \tag{13}$$

where $u_{1dt} = \rho_2^{t-1}u_{1d}$ and $u_{1d} \overset{ind.}{\sim} N(0, \sigma_1^2)$ is the random effect for d-th domain at time $t = 1$. The random effect u_{1d} is scaled by ρ_2, which denotes the autocorrelation for the

second random effect u_{2dt}. u_{2dt} is assumed to follow $AR(1)$ process, as does the Rao and You (1994) model, and is defined below:

$$u_{2dt} = \rho_2 u_{2d,t-1} + \varepsilon_{2dt}, \tag{14}$$

where $|\rho_2| < 1$ and $\varepsilon_{2dt} \overset{ind.}{\sim} N(0, \sigma_2^2)$. The main difference between the Fay and Diallo (2012) and Rao and You (1994) approach is that the former does not constrain $|\rho_2| < 1$ and avoids discontinuity at $\rho_2 = 1$. When $\rho_2 > 1$ a divergence between domains is observed. Let $\theta = (\sigma_1^2, \sigma_2^2, \rho_2)'$ be the vector of unknown parameters involved in the covariance structure of the model. The BLUP estimator for μ_{dt} is calculated in the same fashion as (9):

$$\mu_{dt} = x_{dt}'\tilde{\beta} + (\sigma_1^2 \gamma_{T,u1} + \sigma_2^2 \gamma_{T,u2})V_d^{-1}(y_d - X_d\tilde{\beta}), \tag{15}$$

where the elements are defined as follows (for simplicity $u_1 = u_{1d}$, $u_2 = u_{2dt}$ and $\rho_2 = \rho$ is used):

- Γ_{u1} is a symmetric matrix $T \times T$ where $\Gamma_{u1(1,j)} = 0$ and $\Gamma_{u1(i,j)} = \rho^{(j-i)} \sum_{i'=1}^{i-1} \rho^{(2i'-2)}$ for $1 < i \le j$,
- Γ_{u2} is a symmetric matrix $T \times T$ of elements ρ^{i+j-2},
- $V_d = \Sigma_d + \sigma_1^2 \Gamma_{u1} + \sigma_2^2 \Gamma_{u2} = Cov(y_d)$,
- $V = diag(V_d) = Cov(y)$,
- $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$,
- $\gamma_{T,u1}$ is T column of matrix Γ_{u1},
- $\gamma_{T,u2}$ is T column of matrix Γ_{u2}.

Finally, when $\hat{\theta} = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\rho}_2)'$ is known, the EBLUP of μ_{dt} is given by

$$\hat{\mu}_{dt} = x_{dt}'\tilde{\beta} + (\hat{\sigma}_1^2 \hat{\gamma}_{T,u1} + \hat{\sigma}_2^2 \hat{\gamma}_{T,u2})\hat{V}_d^{-1}(y_d - X_d\tilde{\beta}), \tag{16}$$

or, following Henderson (1975) and Marhuenda, Molina and Morales (2013), can also be written as:

$$\hat{\mu}_{dt} = X_{dt}'\beta + \hat{u}_{1dt} + \hat{u}_{2dt}, \tag{17}$$

where $\hat{u}_{1dt} = \hat{\rho}_2^{t-1}\hat{u}_{1d}$. For the proof of (15) and mathematical details of the model (16) refer to Fay and Diallo (2012).

3.4. Marhuenda, Molina and Moralez (2013) model

Finally, in order to verify whether to include the spatial effect, we applied the spatio-temporal model proposed by Marhuenda, Molina and Morales (2013). The model assumes two random effects — spatially correlated and temporally correlated domain effect. As in the previous models, in the first stage it assumes:

$$\bar{y}_{dt} = \mu_{dt} + e_{dt} \tag{18}$$

where $e_{dt} \overset{ind.}{\sim} N(0, \psi_{dt})$, where ψ_{dt} is the known sampling variance. In the second stage (the linking model) a linear mixed model is assumed and is given by:

$$\mu_{dt} = X'_{dt}\beta + u_{1d} + u_{2dt} \tag{19}$$

where u_{1d} denotes a spatial random effect that follows the SAR(1) process with variance σ_1^2, spatial autocorrelation ρ_1 and row-standardized proximity matrix $W = (w_{d,k})$. Such a proximity matrix is created based on neighbours matrix W^0. The matrix W is derived from the matrix W^0 by dividing each row element by the row total (Bivand, Pebesma and Gomez-Rubio, 2013). We assume that the spatial representation of domains does not change over time (borders are the same). The SAR(1) process is given by:

$$u_{1d} = \rho_1 \sum_{d \neq k} w_{d,k} u_{1k} + \varepsilon_{1d}, \tag{20}$$

where $|\rho_1| < 1$, and $\varepsilon_{1d} \overset{ind.}{\sim} N(0, \sigma_1^2)$. The second random effect u_{2dt} is assumed to follow the AR(1) process with σ_2^2 and ρ_2 and is given by the following equation:

$$u_{2dt} = \rho_2 u_{2d,t-1} + \varepsilon_{2dt}, |\rho_2| < 1, \varepsilon_{2dt} \overset{ind.}{\sim} N(0, \sigma_2^2). \tag{21}$$

Let $\theta = (\sigma_1^2, \sigma_2^2, \rho_1, \rho_2)'$ be the vector of unknown parameters involved in the covariance structure of the model. After the estimation of $\hat{\theta} = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\rho}_1, \hat{\rho}_2)'$ the EBLUP estimator (19) of $\bar{y}_{dt}$ proposed by Marhuenda, Molina and Morales (2013) is given by:

$$\hat{\mu}_{dt} = X'_{dt}\beta + \hat{u}_{1d} + \hat{u}_{2dt}. \tag{22}$$

In contrast to Rao and You (1994) and Fay and Diallo (2012), Marhuenda, Molina and Morales (2013) estimated the parameters using the Henderson (1975) approach instead of the method of moments. Details about the estimation of the model (19) can be found in Marhuenda, Molina and Morales (2013) and Molina and Marhuenda (2015).

Shortly summarizing the models presented, the following differences can be indicated. Rao and You model assumes stationarity for time series and two uncorrelated random effects. In Fay and Diallo model a time series is non-stationary and random effects are correlation. Marhuenda, Molina and Morales model takes into account SAR(1) process for the first random effect and AR(1) process for the second random effect.

3.5. MSE calculation

Rao and You (1994) and Fay and Diallo (2012) obtained MSE for estimators (16) and (10) by deriving a direct formula using the method of moments based on the Prasad and Rao (1990) approach. In contrast, Marhuenda, Molina and Morales (2013) proposed a parametric bootstrap to estimate MSE of (22). The motivation for such an approach is based on the González-Manteiga et al. (2008) and Molina, Salvati and Pratesi (2009) papers, which discussed estimation of MSE through the parametric bootstrap.

Therefore, to make MSE comparable between the models we applied the parametric bootstrap approach for each model. In the case of Marhuenda, Molina and Morales (2013)

model, the parametric bootstrap was available. For Rao and You (1994) and Fay and Diallo (2012) we developed a procedure to estimate MSE under the parametric bootstrap. The details can be found in Table 1. The notation used in the table is consistent with the Marhuenda, Molina and Morales (2013) article. The steps (3), (5), (7) and (8) are the same for all models.

3.6. Diagnostics measures for models

3.6.1   Model comparison measures

In order to compare and evaluate the models we applied several measures. Firstly, we used cAIC criterion (Greven and Kneib, 2010), pseudo-R^2 and Wald statistic. These measures were used to compare and verify which model is the most suitable for estimation of the low work intensity indicator. In addition, for practical and descriptive reasons, pseudo-R^2 for each model was computed and is given in (23). The inclusion of the pseudo-R^2 measure is motivated by the ease of interpretation as a measure of goodness of fit and end users' experiences with linear models. However, this measure is rarely presented in the context of small area models. For other pseudo-R^2 measures for linear mixed models, see Nakagawa and Schielzeth (2013), and for Wald statistic denoted by W refer to Brown et al. (2001). Calculated information criteria are given in (23):

$$
\begin{aligned}
cAIC &= -2 \times LogLik + 2 \times (trace(H) + 1), \\
pseudo - R^2 &= \text{Var}(\hat{\mu}_{dt}) / \text{Var}(\hat{\tilde{y}}_{dt}), \\
W &= \sum (\hat{\tilde{y}}_{dt} - \hat{\mu}_{dt})^2 / (\text{Var}(\hat{\tilde{y}}_{dt}) + \text{Var}(\hat{\mu}_{dt})),
\end{aligned}
\tag{23}
$$

where *LogLik* is the value of log-likelihood estimated through REML estimation of the variance components, p denotes the number of model parameters (fixed and for random effects), n denotes the number of observations, $trace(H)$ trace of hat matrix given by equation (24) and Var denotes simple random sampling variance.

$$
\begin{aligned}
trace(H) =& trace((X'_{dt} V(\theta)^{-1} X_{dt})^{-1} X'_{dt} V(\theta)^{-1} V_e V(\theta)^{-1} X_{dt}) \\
& + n - trace(V_e V(\theta)^{-1})
\end{aligned}
\tag{24}
$$

Bias correction of conditional Akaike information criterion is given by equation (24). V_e in this equation denotes variance matrix of random error. Calculation of this term is possible with cAIC4 R package written by Saefken et al. (2018). Conditional Akaike information criterion depends on the structure of the model used so two other metrics in (23) were proposed.

3.6.2   Collinearity diagnostics

To evaluate the models we investigated collinearity measures using generalized variance inflation factors (GVIF) proposed by Fox and Monette (1992). The GVIF measure is limited

Table 1: Calculation of parametric bootstrap MSE in Rao and Yu (1994), Fay and Diallo (2012) and Marhuenda, Molina and Moralez (2013) models

Step	Rao and Yu (1994)	Fay and Diallo (2012)	Marhuenda, Molina and Moralez (2013)
1	Using the available data $\{(\hat{\bar{y}}_{dt}, X_{dt}), t = 1, \ldots, T, d = 1, \ldots, D\}$, fit the Rao and You (1994) model to obtain model parameter estimates $\hat{\beta}$, $\hat{\sigma}_2^1$, $\hat{\sigma}_2^2$ and $\hat{\rho}_2$.	Using the available data $\{(\hat{\bar{y}}_{dt}, X_{dt}), t = 1, \ldots, T, d = 1, \ldots, D\}$, fit the Fay and Diallo (2012) model to obtain model parameter estimates $\hat{\beta}$, $\hat{\sigma}_2^1$, $\hat{\sigma}_2^2$ and $\hat{\rho}_2$.	Using the available data $\{(\hat{\delta}_{dt}^{DIR}, \mathbf{x}_{dt}), t = 1, \ldots, T, d = 1, \ldots, D\}$, fit the Molina, Marhuenda, Molina and Morales (2013) model to obtain model parameter estimates $\hat{\beta}$, $\hat{\sigma}_1^1$, $\hat{\rho}_1$, $\hat{\sigma}_2^2$ and $\hat{\rho}_2$.
2	Generate bootstrap area effects $\{u_{1d}^{*(b)}, d = 1, \ldots, D, t = 1, \ldots, T\}$ using $\hat{\sigma}_1^2$ as true values of parameters σ_1^2 that $\{u_{1d}^{*(b)} = u_{1d,t=1}^{*(b)} = \ldots = u_{1d,t=T}^{*(b)}\}$.	Generate bootstrap area effects $\{u_{1d}^{*(b)}, d = 1, \ldots, D, t = 1\}$ with known $\hat{\sigma}_1^2$ as true value of parameter σ_1^2. Then, compute $\{u_{1dt}^{*(b)} = \rho_2^{t-1} u_{1d}^{*(b)}, t = 2, \ldots, T\}$ where $\hat{\rho}_2$ is the true value of ρ_2.	Generate bootstrap area effects $\{u_{1d}^{*(b)}, d = 1, \ldots, D, t = 1, \ldots, T\}$, from the SAR(1) process given in (20), using $(\hat{\sigma}_1^2, \hat{\rho}_1)$ as true values of parameters (σ_1^2, ρ_1) and $u_{1d}^{*(b)} = u_{1d,t=1}^{*(b)} = \ldots = u_{1d,t=T}^{*(b)}$.
3	Independently of $\{u_{1d}^{*(b)}\}$ and independently for each d, generate bootstrap time effects $\{u_{2dt}^{*(b)}, t = 1, \ldots, T\}$, from the AR(1) process given in (8), with $(\hat{\sigma}_2^2, \hat{\rho}_2)$ acting as true values of parameters (σ_2^2, ρ_2).		
4	Calculate true bootstrap quantities, $\mu_{dt}^{*(b)} = X'_{dt}\hat{\beta} + u_{1d}^{*(b)} + u_{2dt}^{*(b)}$.	Calculate true bootstrap quantities, $\mu_{dt}^{*(b)} = X'_{dt}\hat{\beta} + u_{1dt}^{*(b)} + u_{2dt}^{*(b)}$.	Calculate true bootstrap quantities, $\mu_{dt}^{*(b)} = X'_{dt}\hat{\beta} + u_{1d}^{*(b)} + u_{2dt}^{*(b)}$.
5	Generate errors $e_{dt}^{*(b)} \overset{ind.}{\sim} N(0, \psi_{dt})$ and obtain bootstrap data from the sampling model, $\hat{\bar{y}}_{dt}^{*(b)} = \mu_{dt}^{*(b)} + e_{dt}^{*(b)}$		
6	Using the new bootstrap data $\{(\hat{\bar{y}}_{dt}^{*(b)}, X_{dt}), t = 1, \ldots, T, d = 1, \ldots, D\}$, fit the Rao and You (1994) model (7) - (11) to obtain the bootstrap EBLUPs, $\hat{\mu}_{dt}^{*(b)}$	Using the new bootstrap data $\{(\hat{\bar{y}}_{dt}^{*(b)}, X_{dt}), t = 1, \ldots, T, d = 1, \ldots, D\}$, fit Fay and Diallo (2012) model (13) - (17) to obtain the bootstrap EBLUPs, $\hat{\mu}_{dt}^{*(b)}$	Using the new bootstrap data $\{(\hat{\bar{y}}_{dt}^{*(b)}, X_{dt}), t = 1, \ldots, T, d = 1, \ldots, D\}$, fit Marhuenda, Molina and Morales (2013) model (19) - (22) to obtain the bootstrap EBLUPs, $\hat{\mu}_{dt}^{*(b)}$
7	Repeat steps (1)-(6) for $b = 1, \ldots, B$, where B is a large number.		
8	Calculate parametric bootstrap MSE according to the following formula: $MSE(\hat{\mu}_{dt}) = \frac{1}{B} \sum_{b=1}^{B} (\hat{\mu}_{dt}^{*(b)} - \mu_{dt}^{*(b)})^2$		

to fixed effects (X_{dt}) and does not account for the variance structure of random effects. Thus, it overestimates the collinearity between auxiliary variables X_{dt}. Other approaches to

estimate VIF in the context of complex surveys are discussed by Liao and Valliant (2012) and Li and Valliant (2015) assuming a linear model with known sampling variances.

Therefore, we modified GVIF to be conditional on the Fay-Herriot small area model covariance matrix of y given by: $V(\theta) = ZV(\theta)_u Z' + V_e$, where Z is a matrix of random effects, $V(\theta)_u$ denotes block-diagonal covariance structure for random effects and V_e is a diagonal matrix of known sampling variances. Let $\Sigma_x(\theta)$ denote the variance-covariance matrix for the fixed effect X_{dt} defined by the equation (25)

$$\Sigma_x(\theta) = (X'_{dt}V(\theta)^{-1}X_{dt})^{-1}, \tag{25}$$

and the estimator of (25) is given by

$$\hat{\Sigma}_x = \tilde{\Sigma}_x(\hat{\theta}) = (X'_{dt}V(\hat{\theta})^{-1}X_{dt})^{-1}, \tag{26}$$

where $V(\hat{\theta})$ is an estimated covariance structure of the small area model. The $V(\hat{\theta})$ can differ between the models and depends on the assumed underlying structure of random effects. To calculate conditional GVIF $\hat{\Sigma}_x$ need to be transformed into a correlation matrix, which we denote as $R(\theta)$. The estimator of $R(\theta)$ is given by the following transformation of $\hat{\Sigma}_x$

$$R(\hat{\theta}) = D^{-1}\hat{\Sigma}_x D^{-1}, \tag{27}$$

where $D = diag(\sqrt{diag(\hat{\Sigma}_x)})$. Finally, the GVIF conditional on $V(\hat{\theta})$ for each variable of the fixed effect is given by

$$GVIF(x_k|V(\hat{\theta})) = \frac{det(R(\hat{\theta})_{k,k}) \times det(R(\hat{\theta})_{-k,-k})}{det(R(\hat{\theta}))} \tag{28}$$

where x_k denotes k-th variable from the auxiliary matrix X_{dt}, det denotes the determinant of a matrix, $R(\hat{\theta})_{k,k}$ denotes matrix with k-th variable and $R(\hat{\theta})_{-k,-k}$ without k-th variable. According to Chatterjee and Price (1991), it is assumed that values $GVIF(x_k|V(\hat{\theta}))$ exceeding 10 are to be highly correlated with other fixed effects. Thus, a given variable should be removed from the small area model.

4. Data utilized in the study

4.1. EU-SILC data

The study was based on EU-SILC data from 8 years: 2005 to 2012. As mentioned earlier, the EU-SILC survey is conducted to collect information on income, poverty and other aspects of living conditions of households in European countries. The sample size is set to be representative at the national level. However, in Poland the sample size is big enough to publish information about households at the regional level (NUTS 1) as well.

The number of households in the sample varies from 317 (Opolskie Voivodeship in 2009) to 2,212 (Slaskie Voivodeship in 2005). According to the sampling scheme applied, the sample size was distributed proportionally to the domains in the voivodeship. It should

be noticed that the sample of households in the survey decreases from year to year. An average decrease compared to the base year 2005 is 20%. The change is due to several causes. First of all, EU-SILC is a panel and thus requires respondents to participate in the survey multiple times. In addition, non-response is present, which decreases the sample size. Coefficient of variation for direct estimates varies from 5.4% for Slaskie Voivodeship in 2005 to 37.4% for Podlaskie in 2010. For these reasons the sample size in the domains of interest is not acceptable for deriving direct estimates.

4.2. Auxiliary variables

Small area estimation at area-level requires auxiliary information about study domains. Rao and Molina (2015) recommend using register or census data that are free from sampling errors. Therefore, to estimate models, we collected socio-economic data from the Local Data Bank maintained by Statistics Poland. The main criteria for the choice of variables were availability at NUTS 2 level for the years 2005-2012 and the source of data, in particular registers. Several variables were considered and finally the following ones were chosen: registered unemployment rate, working and post-working age people and the number of people in NUTS 2 regions.

The registered unemployment rate is calculated as the ratio of the number of registered unemployed persons to the economically active civilian population. Working and post-working age was used to create two ratios. First, the number of people of working age (aged 15-64) divided by the number of people of post-working age (65 and over). This measure can be interpreted as describing how many independent workers have to provide for one pensioner. The second ratio has the same numerator but the denominator is the number of people without additional criterion (the whole population).

5. Estimation of low work intensity indicator at voivodeship level

In this section we describe the results and provide diagnostics for each model. All the calculations were done in R using the following packages: `sae` (Molina and Marhuenda, 2015), `sae2` (Fay and Diallo, 2015), `metafor` (Viechtbauer, 2010). For the sake of simplicity, we will use `RY` for Rao and You (1994), `FD` for Fay and Diallo (2012) and `MMM` for Marhuenda, Molina and Morales (2013) model.

5.1. Comparison of models

Table 2 contains a comparison of the parameters and statistics for all the models. `RY` and `FD` had 7 parameters, while `MMM` had a total of 8 parameters. The fixed effects in all the models are significant and have expected signs. Slight differences can be observed in the level of the intercept in `FD`. In all the models registered unemployment rate is positively correlated with the LWI indicator: a rise in the level of registered unemployment is associated with higher LWI. When the ratio of the post-working age to working age population rises, the LWI also rises and the ratio of the working-age population to the whole population has the expected sign: if the ratio grows, the LWI decreases. Therefore, we can conclude that the

auxiliary variables are good predictors for the LWI indicator and do not differ between the models.

The second group of parameters are variances of random effects. For the sake of simplicity, standard deviations σ_{u*} are used instead of variances σ_{u*}^2. In RY the AR(1) effect dominates the domain effect and is responsible for almost all the variance of random effects. In contrast, in FD and MMM the domain effect is higher than the AR(1) process of random effects. In the case of the FD model, this means that the domain effect is not constant over time (is nonstationary) and is higher in the first year of the survey but decreases over time by 0.9407^{t-1}. On the other hand, in the MMM model the domain effect is spatially correlated and the variance of this random effect is higher than the AR(1) effect.

In all of the models the AR(1) effect has a strong autocorrelation (ρ_2), which means that the effects within domains depend strongly on what happened in the previous year. The RY and FD models indicate that this autocorrelation is over 0.9 while, in the case of MMM, we can observe a slightly smaller value. In the case of the MMM model, this is due to the second autocorrelation parameter (ρ_1), which is associated with the spatial effect (SAR(1)). The value of $\rho_1 = 0.4866$ indicates that a moderate spatial effect between NUTS 2 is observed, which is smaller than the AR(1) autocorrelation.

If we compare the model statistics concerning information criteria and R^2 we can observe slight differences between the models. All the models explained almost 85% of the variance of the direct estimator. The RY and FD models have similar information criteria, while the MMM model differs slightly. However, the differences between the model statistics do not clearly indicate which model should be recommended. Nonetheless, if we compare all the statistics in Table 2 the model proposed by Marhuenda, Molina and Morales (2013) seems to be the most reasonable due to the significant spatial effect.

The comparison of EBLUPs for the RY, FD and MMM models with the direct estimator indicates that the model-based estimation is coherent with direct estimation. Pearson correlation coefficients for all estimates are above 0.9. EBLUPs obtained for the models do not differ significantly; however, compared to direct estimates, we can observe differences between estimates.

The differences between model-based and direct estimates are visible in Figure 1. LWI decreases over time from over 15% to below 10%. The solid line indicates direct estimates and dashed lines represent model-based estimates. In general, we can observe a similar trend in all NUTS 2 regions in Poland, but at different levels of intensity. In addition, model-based estimates are more stable over time than direct estimates. In some voivodeships (Lubuskie, Podlaskie or Zachodniopomorskie) there is a clearly visible rise in LWI after 2008, which can be associated with the start of the 2008 crisis.

The biggest differences in the LWI indicator can be observed for Lubuskie and Opolskie Voivodeships. Direct estimates for Lubuskie indicate that from 2008 to 2010 LWI increased, while model-based estimates indicate that the increase was smaller and was only present between 2009 and 2010. These differences, however, may be due to the sampling error, which is higher at NUTS 2 level. It is possible that in the case of Lubuskie specific units were included in the sample in 2008 and took part in the EU-SILC survey until 2010. In Opolskie Voivodeship, there was a considerable increase in direct estimates between 2009 and 2010, followed by a decrease. These differences may also be due to the sampling error,

Table 2: Summary of the estimated model parameters and statistics. Standard deviations are given in parentheses after the mean values.

Parameters/Models	RY	FD	MMM
Model parameters – fixed effects			
Intercept	1.28 (0.27)	1.47 (0.30)	1.28 (0.29)
Register Unemployment Rate	0.37 (0.07)	0.35 (0.07)	0.38 (0.07)
Working / Post-Working Ratio	0.09 (0.01)	0.09 (0.01)	0.09 (0.01)
Working age / Population Ratio	-2.49 (0.46)	-2.80 (0.49)	-2.49 (0.46)
Model parameters – random domain effects variances			
σ_1 Domain effect	0.00 (0.11)	0.04 (0.03)	0.03 (0.02)
σ_2 AR(1)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Model parameters – random domain effects autocorrelation			
ρ_1 SAR(1)	-	-	0.47 (0.00)
ρ_2 AR(1)	0.98 (0.26)	0.94 (0.02)	0.88 (0.00)
Model statistics			
REML LL	327.72	329.13	336.69
cAIC	-582.90	-591.53	-581.96
pseudo R^2	83.97	84.57	84.43
$W(\chi^2_{0.05} = 155.40)$	43.76	47.49	43.45
Degrees of freedom	7	7	8

especially given that the region of Opolskie is characterized by the highest level of the standard error of direct estimates.

According to Brown et al (2001) the difference between direct estimates and model-based estimates should be not significant. Figure 1 shows that these differences are rather small. Pearson correlation coefficient for direct and RY model estimates vary from 0.3856 to 0.9886 with average equal to 0.9226. For FD model correlations are in the range [0.3600;0.9873] (average 0.9192) while for estimates derived from MMM the model minimum value is equal to 0.3906, maximum to 0.9894 and average to 0.9239. In all cases the lowest values were observed in Lubuskie voivodeship and the highest in Śląskie Voivodeship. The highest similarity of estimates measured by average correlation coefficient was obtained for Marhuenda, Molina and Moralez (2013) model. These results show consistency of direct and small area estimates.

5.2. Comparison of coefficient of variations of estimates

The distribution of the CV is given in Table 3. An increase in CV over time was observed, which is due to increasing non-response and the respondent burden in the EU-SILC survey. On average, CV for direct estimates is equal to 15.77%.

In comparison to model-based estimation, the CV for direct estimation increases more rapidly, while the CV for RY, FD and MMM models increase more steadily. Moreover, CVs differ depending on the NUTS 2 unit. For example, in Opolskie and Podlaskie CV is significantly higher in comparison to other NUTS 2 units in Poland, mainly owing to smaller sample sizes. Therefore, especially for these regions, the direct estimator is not reliable.
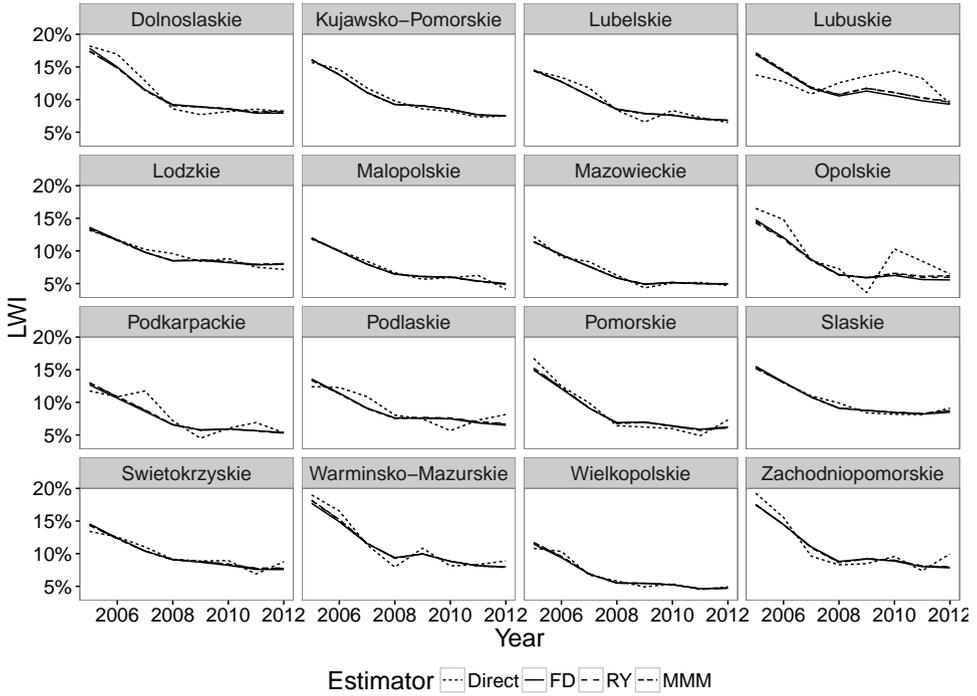
Figure 1: Comparison of direct, Rao and Yu (1994), Fay and Diallo (2012) and Marhuenda, Molina and Moralez (2013) models estimates

Table 3: Comparison of direct, Rao and Yu (1994), Fay and Diallo (2012) and Marhuenda, Molina and Moralez (2013) models coefficient of variations

Model	Min	Q1	Median	Mean	Q3	Max
Direct	5.37	11.62	14.76	15.77	18.55	37.38
RY	4.01	6.91	8.38	8.99	10.39	20.26
FD	3.99	6.92	8.36	9.04	10.33	21.74
MMM	4.19	6.84	8.55	9.09	10.45	20.57

CVs for all the models of interest are lower in comparison to the direct estimator. On average, the CV for each model is approximately 9%, which indicates that the model-based approach provides more reliable estimates. However, as was the case with model diagnostics, models RY, FD and MMM provide similar level of precision and, on average, the RY model is slightly better in comparison to the other models. The lowest CV can be observed for Slaskie and Mazowieckie Voivodeships and the highest for Podlaskie and Opolskie. What is worth noticing is model-based estimation provides more reliable estimates over time even if the non-response increases.

5.3. Diagnostics of the models

Table 4 contains information about GVIF calculated using formula (28). The first three columns refer to the model in question and the last one, denoted by WOLS, refers to weighted ordinary least square regression, where weights are the inverse of sampling errors. Results indicate that GVIF values for all variables in the RY, FD and MMM models are close to 1.3, which is lower than the threshold of 4. Moreover, as expected, the values are smaller than those observed in weighted OLS. The inclusion of the estimated covariance matrix accounts for the uncertainty.

Table 4: Generalized variance inflation measures for auxiliary variables used in Rao and Yu (1994), Fay and Diallo (2012) and Marhuenda, Molina and Moralez (2013) models

Variable	RY	FD	MMM	WOLS
Register Unemployment Rate	1.30	1.31	1.29	1.42
Working / Post-Working Ratio	1.31	1.33	1.29	2.05
Working age / Population Ratio	1.44	1.51	1.37	1.54

6. Conclusions

Application of three proposed models — Rao and You (1994), Fay and Diallo (2012), and Marhuenda, Molina and Morales (2013), allows to obtain more reliable (in the sense of CV) estimates in previously unpublished domains. All models take into account auxiliary variables, temporal effect, however Marhuenda, Molina and Morales (2013) also deal with spatial information. The registered unemployment rate showed the strongest relation with the indicator. Based on the results and strong spatial autocorrelation, we choose Marhuenda, Molina and Morales (2013) model as the most suitable for the estimation of the low work intensity indicator. The final results are presented in Figure 2.

Based on Figure 2, we noticed spatial regimes in the low work intensity in the West (Zachodniopomorskie, Lubuskie and Dolnoslaskie Voivodeships) and Central (Lodzkie, Swietokrzyskie and Slaskie Voivodeships) Poland between 2005 and 2012. Mazowieckie (with Warsaw) and Wielkopolskie (with Poznań) regions are characterized by the lowest level of the indicator.

Future works will focus on estimation of Europa 2020 indicators at more detailed levels of spatial aggregation, i.e. NUTS 3 or LAU 1. Local authorities demand such information to conduct adequate social policy. However, due to sample sizes at such low level as LAU 1 (380 areas in Poland) area-level models might not be adequate. Instead, unit-level models might be useful, but require access to population unit-level data, e.g. from registers or census.
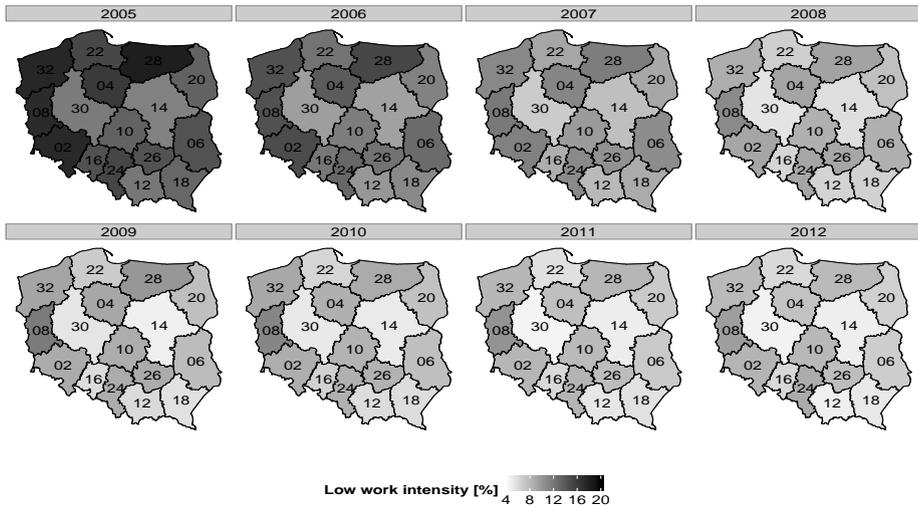
Figure 2: Spatial distribution of low work intensity indicator in Poland between 2005 and 2012 estimated using Marhuenda, Molina and Moralez (2013) model.

Codes on map: 02 – Dolnoslaskie, 04 – Kujawsko-Pomorskie, 06 – Lubelskie, 08 – Lubuskie, 10 – Lodzkie, 12 – Malopolskie, 14 – Mazowieckie, 16 – Opolskie, 18 – Podkarpackie, 20 – Podlaskie, 22 – Pomorskie, 24 – Slaskie, 26 – Swietokrzyskie, 28 – Warminsko-Mazurskie, 30 – Wielkopolskie, 32 – Zachodniopomorskie.

References

Brown, G., Chambers, R., Heady, P., Heasman, D., (2001). Evaluation of small area estimation methods – an application to unemployment estimates from the UK LFS. *Proceedings of Statistics Canada Symposium*.

Bivand, R., Pebesma, E., Gomez-Rubio, V., (2013). *Applied spatial data analysis with R, Second edition.* Springer. New York.

Chandra, H., Chambers, R., Salvati, N., (2012). Small area estimation of proportions in business surveys. *Journal of Statistical Computation and Simulation*, 82(6), pp. 783–795.

Chatterjee, S., Price, B., (1991). *Regression Diagnostics*, New York: John Wiley.

Dehnel, G., Wawrowski, Ł., (2020). Robust estimation of wages in small enterprises: the application to Poland's districts. *Statistics in Transition New Series*, 21(1), pp. 137–157.

Deville, J. C., Särndal, C. E., (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376–382.

Fay, R. E., Diallo, M. S., (2012). Small Area Estimation Alternatives for the National Crime Victimization Survey. *Proceedings of the Section on Survey Research Methods. American Statistical Association*, pp. 3742–3756.

Fay, R. E., Diallo, M. S., (2015). *sae2: Small Area Estimation: Time-series Models*. R package.

Fay, R. E., Herriot, R. A., (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, pp. 269–277.

Fay, R. E., Planty, M., Diallo, M. S., (2013). Small area estimates from the national crime victimization survey. *Proceedings of the Section on Survey Research Methods. American Statistical Association*, pp. 1544–1557.

Fox, J., Monette, G., (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 87, pp. 178–183.

GonzáLez-Manteiga, W., LombardíA, M., Molina, I., Morales, D., SantamaríA, L., (2008). Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation*, 78, pp. 443–462.

Greven, S., Kneib, T., (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, 97(4), pp. 773–789.

Henderson, C. R., (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, pp. 423–447.

Li, J., Valliant, R., (2015). Linear Regression Diagnostics in Cluster Samples. *Survey Methodology*, 31, pp. 61–75.

Liao, D., Valliant, R., (2012)., Variance inflation factors in the analysis of complex survey data. *Survey Methodology*, 38, pp. 53–62.

López-Vizcaíno, E., Lombardía, M. J., Morales, D., (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling,13(2), pp. 153–178, doi:10.1177/1471082X13478873.

Marhuenda, Y., Molina, I., Morales, D., (2013). Small area estimation with spatio-temporal Fay–Herriot models. *Computational Statistics and Data Analysis*, 58, pp. 308–325.

Mélina, A., Emilio, D. M., (2012). 23% of EU citizens were at risk of poverty or social exclusion in 2010. *Statistics in Focus*, 9, pp. 1–7.

Molina, I., Marhuenda, Y., (2015). sae: An R package for small area estimation. *The R Journal*, 7(1), pp. 81–98.

Molina, I., Rao, J. N. K., (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, 38(3), pp. 369–385.

Molina, I., Salvati, N., Pratesi, M., (2009). Bootstrap for estimating the MSE of the Spatial EBLUP. *Computational Statistics*, 24, pp. 441–458.

Nakagawa, S., Schielzeth, H., (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4, pp. 133–142.

Prasad, N., Rao, J. N. K., (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, pp. 163–171.

Rao, J. N. K., Molina, I., (2015). *Small Area Estimation*. Wiley and Sons.

Rao, J. N. K., Yu, M., (1994). Small-area estimation by combining time-series and cross-sectional data. *Canadian Journal of Statistics*, 22, pp. 511–528.

Saefken, B., Ruegamer, D., Kneib, T., Greven, S., (2018). Conditional Model Selection in Mixed-Effects Models with cAIC4, *ArXiv e-prints 1803.05664*.

Statistics Poland, (2014). *Incomes and living conditions of the population of Poland —- report from the EU-SILC survey of 2012*. Statistical Publishing Establishment. Warsaw.

Szymkowiak, M., Młodak, A., Wawrowski, Ł., (2017). Mapping poverty at the level of subregions in Poland using indirect estimation. *Statistics in Transition new series*, 18(4), pp. 609–635.

Viechtbauer, W., (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), pp. 1–48.

Ward, T., Ozdemir, E., (2013). Measuring low work intensity — an analysis of the indicator. *ImPRovE Discussion Paper*, 13, pp. 1–37.

A dynamic MST-deltaCoVaR model of systemic risk in the European insurance sector

Anna Denkowska[1], Stanisław Wanat[2]

ABSTRACT

This work is a response to the EIOPA paper entitled 'Systemic risk and macroprudential policy in insurance', which asserts that in order to evaluate the potential systemic risk (SR), the build-up of risk, especially risk arising over time, should be taken into account, as well as the interlinkages occurring in the financial sector and the whole economy. The topological indices of minimum spanning trees (MST) and the deltaCoVaR measure are the main tools used to analyse the systemic risk dynamics in the European insurance sector in the years 2005-2019. The article analyses the contribution of each of the 28 largest European insurance companies, including those appearing on the G-SIIs list, to systemic risk. Moreover, the paper aims to determine whether the most important contribution to systemic risk is made by companies with the highest betweenness centrality or the highest degree in the obtained MST.

Key words: systemic risk, minimum spanning trees, deltaCoVaR, insurance sector.

1. Introduction

The subject of this work is part of the current research on the interlinkages of large insurance companies and their contribution to systemic risk (SR) in the insurance sector. In the international literature we find many studies related to the systemic risk analysis in the banking sector but little is done regarding systemic risk in the insurance sector. The novelty we bring to the literature is to identify the relationship between the contribution to systemic risk and the structure of the minimum spanning tree (MST) described by topological network indicators, which can be used in further research to construct models whose task is to predict the possibility of systemic risk. The use of a hybrid approach, combining statistical and econometric tools with network modelling and predictive analysis tools, improves both the explanatory and predictive

[1] Department of Mathematics, Cracow University of Economics, Kraków, Poland.
  E-mail: anna.denkowska@uek.krakow.pl. ORCID ID: https://orcid.org/0000-0003-4308-8180.
[2] Department of Mathematics, Cracow University of Economics, Kraków, Poland. E-mail: wanats@uek.krakow.pl.
  ORCID ID: https://orcid.org/0000-0001-6225-4207.

power of models used to assess the impact of global system disturbances on the development of the insurance sector.

Our work constitutes an answer to the recommendations contained in the 2017 report of the European Insurance and Occupational Pensions Authority (EIOPA), an independent EU advisory body to the European Parliament, the Council of Europe and the European Commission, which shows that when analyzing systemic risk in the insurance sector one should take into account the dynamics of interconnectedness between institutions, among other things. The present article is another study of the authors in this subject. In a previous paper (Wanat and Denkowska, 2019), selected topological indicators of Minimum Spanning Trees were analysed and interrelationships between major European insurance institutions were examined.

The purpose of this work is to confront the contribution to the systemic risk of each of the 28 insurance institutions analysed with the results of the previous study. We examine whether interlinkages between insurers, and thus the level of possibility of contagion with a potential crisis, is related to the creation of systemic risk by individual insurance companies. Therefore, we examine whether insurance companies, which in MST play a significant role in the structure of the entire network (they are the so-called HUBs) and would pose a threat to the linked companies in the event of bankruptcy, have an equally significant contribution to SR.

The global economic crisis on financial markets, the peak of which was in the years 2008-2009, initiated on the market of high-risk mortgage loans in the USA as a result, among other things, of the deregulation of the financial market in 1999, when the ban on combining two types of banking: investment (high risk) and deposit-credit banking, which had been in force since 1929 and was to protect citizens in the event of losses in investment banking, ceased to apply. On September 15th 2008 Lehman Brothers – the fourth largest investment bank, after a fruitless attempt to obtain help from the US central bank, declared bankruptcy. A week earlier, the Fed took over two insurance and loan companies with huge debts: Fannie Mae and Freddie Mac. The Fed and the Ministry of the Treasury have recapitalized the largest insurance and financial holding company AIG in the amount of USD 85bn, as the loss of liquidity and, consequently, its collapse would mean a rapid spread of the crisis. The American International Group (AIG) is the first example of an insurance company that required (and received) funding because it was considered systemically important. When dividing financial institutions into three groups according to their activity, i.e. investment, depository and risk-dissipating institutions, insurance companies should be in the third group. Therefore, they should not generate systemic risk as long as they deal with taking over, dispersing and redistributing the financial effects of risk. But if they take over credit risk, e.g. financial insurance, insurance guarantees, derivative trading, in particular like AIG Credit Default swap (CDS) involved in trading financial instruments given as

collateral in case of default on repayment obligations, they generate systemic risk (see William and Sjostrum, 2009)

After the subprime crisis, all financial supervising authorities drew attention to the need for macro-prudential policy, which would take into account the dynamics of structure changes and linkages between financial institutions. The situation of AIG in 2008 was surprising, when it was announced in February that in 2007 it achieved profits of USD 6.20 billion (USD 2.39 per share). The stock was closed on that day at USD 50.15 per share. Less than seven months later, the company was on the verge of bankruptcy and had to be funded by the US government. This is the first and very important event that indicates that the insurance sector plays a large role when assessing systemic risk.

Therefore, in 2013 and 2016, International Association of Insurance Supervisors (IAIS), when developing a method of identifying insurance institutions of particular importance for financial stability, takes into account the following five dimensions: size of the insurance institution, range of global activity, assessment of the degree of direct and indirect linkages of the institution within the financial system, non-traditional and non-insurance activity of the insurer, product substitutability – the significance of the institution for the financial system increases along with the lack of real substitution possibilities for the services provided by the insurer. In line with the IAIS recommendations, the Financial Stability Board (FSB) announced in 2016 a list of systemically important insurers (G-SIIs): Aegon N.V., Allianz SE, American International Group, Inc. (AIG), Aviva plc, Axa S.A., MetLife, Inc., Ping An Insurance (Group) Company of China, Ltd., Prudential Financial, Inc., Prudential plc.

On the other hand, in 2017, EIOPA publishes the document "Systemic risk and macroprudential policy in insurance", which first analyzes the discussion to date on SR, then gives lessons learned from the financial crisis and the banking sector crisis. It draws attention to the need to complement micro-prudential policy with a macro-prudential approach that requires transnational coordination. EIOPA considers SR in a broader, but not contradictory to IAIS, context. According to EIOPA, system events can be generated in two ways. The first is the "direct" effect, which through a failure is initiated by a systemically important insurer or a collective failure of several insurers generating a cascade effect. The second is the "indirect" effect, which involves behaviour. It is based on the actions and/or responses of insurers to external shocks. Potential externalities generated by direct and indirect sources are transferred to the rest of the financial system and the real economy through dedicated channels (the transmission channel) and can have a significant impact on SR. The transmission channels listed in Table 5, page 29, by EIOPA are: Exposure channel, asset liquidation channel, lack of supply of insurance products, Bank-like channel, Expectations and information asymmetries – 'Soft' or 'indirect contagion' channel, linked to issues such as irrational panics and re-evaluation of expectations. It also includes reputational

issues. One source involving behaviour is the collective actions of insurers, which can influence market prices and excessive risk-taking. In our analysis, we use the weekly logarithmic returns of insurance companies as a reflection of the sentiment in the financial market. They are a response to the behaviour of insurers, thanks to which we catch their reaction to potential shocks. Our emphasis on market returns is motivated by the desire to incorporate the most current information in our measures; market returns reflect information more rapidly than non-market-based measures, such as accounting variables.

Another document where we find the legal definition of SR on the UE level is Regulation (EU) 2019/2176 of the European Parliament and the Council of 18 December 2019 amending Regulation (EU) No 1092/2010 on European Union macro-prudential oversight of the financial system and establishing a European Systemic Risk Board in Article 1, point 1: "Systemic risk means a risk of disruption in the financial system with the potential to have serious negative consequences for the real economy of the Union or of one or more of its Member States and for the functioning of the internal market. All types of financial intermediaries, markets and infrastructure may be potentially systemically important to some degree". According to Article 2, point (b) of Regulation (EU) No 1092/2010 "the financial system means all financial institutions, markets, products and market infrastructures".

Therefore, when analyzing various definitions of SR, we find three common elements in them - namely the fact that SR is associated with an undesirable event that occurs in the financial market, has a systemically important cause and the consequences have an impact on the real economy. The rates of return reflect the condition of insurance companies. In the event of a stock market turmoil, we are able to spot it very quickly. Moreover, these are data that can be seen as a measure of the effects of insurers' activities and policies in the real economy. In the literature, we can find many works in which direct links between insurers and between insurers and other financial institutions, such as banks, are analyzed. One example is Alves (2015). This article explores some direct relationships and recommends that you perform an indirect relationship analysis via *market price channels*.

2. Literature review

Before the subprime (2007–2009) and excessive public debt in the euro area countries (2010–2013) crises, there was a strong belief that the insurance market is systemically insignificant. In the international literature that emerged as a consequence of the crisis, many studies have maintained their previous beliefs, but we also find numerous works confirming the possibility of the insurance sector creating systemic risk. Examples include works in which authors believe that insurance companies have become an unavoidable source of systemic risk (e.g. Billio et al., 2012; Weiß and

Mühlnickel, 2014) and those in which they claim that they can be systematically significant, but this is due to their non-traditional (banking) activities (e.g. Baluch et al., 2011; Cummins and Weiss, 2014) and the overall systemic importance of the insurance sector as a whole is still subordinated to the banking sector (Chen et al., 2013). In turn, in (Bierth et al., 2015), the authors, after examining a very large sample of insurers in the long term, believe that the contribution of the insurance sector to systemic risk is relatively small, however, they claim that it reached its peak during the financial crisis in 2007–2008, which we also confirm in our analysis. Analysts also report that significant factors affecting the insurers' exposure to systemic risk are the strong linkages of large insurance companies, leverage, losses and liquidity (four L's). However, the complete lack of evidence of the systemic importance of the insurance industry is indicated, among others, by the following papers (Harrington, 2009; Bell and Keller, 2009; Geneva Association, 2010).

The problem of the insurance sector's ability to create systemic risk has also been the subject of consideration in Polish scientific literature in recent years. In general, as above, two main positions are represented. In (Czerwińska, 2014), based on research of the insurance sector in European Union countries covering the period 2005–2012, it was found that along with the increase in the level of linkages between insurers and various financial system segments, mainly the capital market and the banking sector, the importance of insurance institutions for the stability of the entire system increases. In turn, in (Bednarczyk, 2013), the author, assessing insurance institutions as a potential creator of systemic risk, concluded that dispersing and taking over insurance risk ultimately does not create systemic risk. It indicates a relatively low level of interconnectedness and draws attention to the fact that insurers are not highly dependent on external financing, so they should not be included in the group of systemically important institutions. At the same time, it was mentioned that insurers engaged in non-insurance activities pose a threat to the system by taking over credit risk. Measures of the impact (contribution) of an individual financial institution on the systemic risk of a given market and measures of the institution's sensitivity to this risk are CoVaR (Acharya et al., 2010; Bierth et al., 2015; Jobst, 2014) and deltaCoVaR (Adrian and Brunnermeier, 2011)

3. Methodology

In order to identify the possible relationship linking the structure of the interconnections between insurance companies to the creation of systemic risk by these companies, we proceed in two steps. In the first one, we analyze the dynamics of the structure of interlinkages between insurers. For this purpose, we use the time series of the following selected topological network indicators (Wang et al., 2014):
  – Average Path Length (APL);

- Maximum Degree - Max.Degree;
- parameters α of the power distribution of vertex degrees: $P(s) = C \cdot s^{-\alpha}$, $\alpha > 0$;
- betweenness centrality (BC).

The Average Path Length (APL) is defined as the average number of steps taken along all the shortest paths connecting all possible pairs of network nodes.

The Maximum Degree: in graph theory it is defined as the maximal number of edges coming out from a vertex (where each loop counts for two). In other words, it measures the number of connections to the central vertex.

The Betweenness Centrality (BC) measures the centrality of a vertex: we consider the ratio between the number of shortest paths connecting two vertices and passing through the given one, and the number of all the shortest paths between pairs of distinct vertices. It indicates thus the most important nodes of a network based on shortest paths (e.g. the most influential insurer).

We obtain these series based on the determined minimum spanning trees MST_t for each period studied. We construct the MST_t trees using conditional correlations between each pair of analysed insurance companies determined using the copula-DCC-GARCH model, using Kruskal's algorithm. Details of the construction are presented in the work (Wanat and Denkowska, 2019).

In the second step, we examine the contribution of a single insurer to the systemic risk of the European insurance sector using the deltaCoVaR model. In this model, the basis for measuring risk is the CoVaR measure. Formally, $CoVaR_{\beta,t}^{j|i}$ is defined as the value at risk (VaR) of an institution j under the condition that another institution i is at risk of crisis in a given period t, i.e. its rate of return is less than its value at risk:

$$P(r_{j,t} \leq CoVaR_{\beta,t}^{j|i} \,|\, r_{i,t} \leq VaR_{\alpha,t}^{i}) = \beta \tag{1}$$

Using the formula for the conditional probability we have:

$$\frac{P(r_{j,t} \leq CoVaR_{\beta,t}^{j|i}, r_{i,t} \leq VaR_{\alpha,t}^{i})}{P(r_{i,t} \leq VaR_{\alpha,t}^{i})} = \beta \tag{2}$$

In addition, the definition of value-at-risk for institutions implies that

$$P(r_{i,t} \leq VaR_{\alpha,t}^{i}) = \alpha \tag{3}$$

From equations (2) and (3) we get:

$$P(r_{j,t} \leq CoVaR_{\beta,t}^{j|i}, r_{i,t} \leq VaR_{\alpha,t}^{i}) = \alpha\beta \tag{4}$$

From the relationship (4) we can estimate $CoVaR_{\beta,t}^{j|i}$, but first we need to determine the two-dimensional distribution F_t of the rate of return vector $(r_{j,t}, r_{i,t})$. This distribution can be represented using the copula in the following way:

$$F_t(r_{j,t}, r_{i,t}) = C_t(F(r_{j,t}), F(r_{i,t})) \tag{5}$$

From formula (5) we can numerically calculate $CoVaR_{\beta,t}^{j|i}$ by solving the equation:

$$C_t\left(F_t\left(CoVaR_{\beta,t}^{j|i}\right), \alpha\right) = \alpha\beta \tag{6}$$

Then, knowing the value of the measure $CoVaR_{\beta,t}^{j|i}$, we can calculate the measure deltaCoVaR ($\Delta CoVaR_{\beta}^{j|i}$). The value of this measure is the difference between the value at risk of the insurance sector (institution j), provided that the insurer (institution i) is in a state of financial crisis and the value at risk of the insurance sector in the event that the financial standing of the entity i is normal (average), i.e.

$$\Delta CoVaR_{\beta}^{j|i} = CoVaR_{\beta}^{j|X^i \leq VaR_{\alpha}^i} - CoVaR_{\beta}^{j|X^i \leq Median^i}. \tag{7}$$

The value of this measure represents the contribution of the institution i to systemic risk. The lower this value, the greater the institution's share in generating systemic risk. In our analysis we estimate the distributions F_t of the vectors $(r_{j,t}, r_{i,t})$ and determine $CoVaR_{\beta,t}^{j|i}$ using the two-dimensional copula-DCC-GARCH models with the t-Student copula.

In these models, the average rate of return was modelled using the following ARIMA process:

$$r_{i,t} = \mu_{i,t} + y_{i,t} \tag{8}$$

$$\mu_{i,t} = E(r_{i,t}|\Omega_{t-1}) \tag{9}$$

$$\mu_{i,t} = \mu_{i,0} + \sum_{j=1}^{p}\varphi_{ij}r_{i,t-j} + \sum_{j=1}^{q}\theta_{ij}y_{i,t-j} \tag{10}$$

$$y_{i,t} = \sqrt{h_{i,t}}\varepsilon_{i,t}, \tag{11}$$

where Ω_{t-1} denotes the collection of information available until the moment $t-1$, while $\varepsilon_{i,t}$ are independent random variables with identical distributions. We model the conditional variance $h_{i,t}$ using the exponential GARCH (eGARCH) model:

$$\log(h_{i,t}) = \omega_i + \sum_{j=1}^{p}\left(\alpha_{ij}\varepsilon_{i,t-j} + \gamma_{ij}\left(|\varepsilon_{i,t-j}| - \right.\right.$$
$$\left.\left. E|\varepsilon_{i,t-j}|\right)\right) + \sum_{j=1}^{q}\beta_{ij}\log(h_{i,t-j}) \tag{12}$$

where $\varepsilon_{i,t} = \frac{y_{i,t}}{\sqrt{h_{i,t}}}$ stands for the standardized errors.

To model the relationship between the rates of return we use Student's t-copula, whose parameters are the conditional correlations R_t, obtained using the DCC(m, n) model:

$$H_t = D_t R_t D_t \tag{13}$$

$$D_t = diag(\sqrt{h_{1,t}}, ..., \sqrt{h_{k,t}}) \tag{14}$$

$$R_t = (diag(Q_t))^{-\frac{1}{2}}Q_t(diag(Q_t))^{-\frac{1}{2}} \tag{15}$$

$$Q_t = (1 - \sum_{j=1}^{m}c_j - \sum_{j=1}^{n}d_j)\bar{Q} + \sum_{k=1}^{m}c_j\left(\varepsilon_{t-j}\varepsilon'_{t-j}\right) + \sum_{k=1}^{n}d_jQ_{t-j} \tag{16}$$

$\bar{Q}$ is the unconditional covariance matrix for standardized errors ε_t; c_j, d_j, $j = 1, ...,$ are scalar values, with c_j describing the impact on current correlations of earlier shocks, and d_j takes into account the impact on current correlations of earlier conditional correlations. We estimate the parameters of the above copula-DCC-GARCH model using the inference function for the margins method.

4. Data and results of empirical analysis

The basis of the analysis are stock quotes of 28 European insurance institutions selected from among the 50 largest companies, for which quotations are available in Thomson Reuters database, in the period adopted for the analysis, according to https://www.relbanks.com/top-insurance-companies/europe, all those that were listed in the period studied, namely: AXA, Allianz, Prudential plc, Legal & General, Generali, Aviva, Aegon, CNP Assurances, Zurich Insurance, Munich Re, Old Mutual, Swiss Life, Chubb Ltd, Ageas, Phoenix, Unipol Gruppo, Mapfre, Hannover Re, Storebrand, XL.Group, Helvetia Holding, Vienna Insurance, SCOR SE, Mediolanum, Sampo Oyj, RSA Insurance Group, Società Cattolica di Assicurazione, Topdanmark A/S. Five of them, AXA, Allianz, Prudential plc, Aviva, Aegon appear as systemically relevant on the current list of G-SIIs published by the FSB in 2016. We estimate the deltaCoVaR measure assuming that the European insurance sector is represented by the STOXX 600 Europe Insurance index. We analyse weekly logarithmic rates of return from 07.01.2005 to 26.04 2019.

Time series of topological network indicators, determined according to the first stage of the presented empirical strategy, are presented in Figure 1 and 2. The first figure (Figure 1) shows the average path length (APL), average maximum degree (Max.Degree) and estimated parameters α of the power distribution for minimum spanning trees from 07/01/2005 to 26/04/2019. The analysis of the charts shows that in the periods of June 2nd, 2006 – August 17th, 2007 and December 5th, 2008 – September 17th, 2010, APL decreases while Max.Degree increases, the α index is close to 2, which means that the network is shrinking and its structure is "scale-free", that is, it takes a form in which there are few vertices with numerous edges (hubs) and many vertices with a low degree (betweenness centrality). In turn, Figure 2 shows BC for AXA, Allianz and Phoenix. The first two companies have the highest average values of this measure in the examined period (see Figure 3), while the third one (Phoenix) is one of the five insurers for which BC is equal to zero in each of the examined weeks. Considering the fact that BC is an indicator on the basis of which we assess the importance of a given insurer in the context of the possibility of risk contagion, we note that the time series for these companies are "complementary" behaving graphs. During the entire analysed period, if BC for AXA increases, BC for Allianz decreases and vice

versa. If BC for AXA remains stable, the BC level for Allianz does not change either. Figure 3 shows that in the subprime crisis state the French AXA company was clearly the dominant one on the European insurance market, while the German Allianz took over during the phase of excessive public debt.
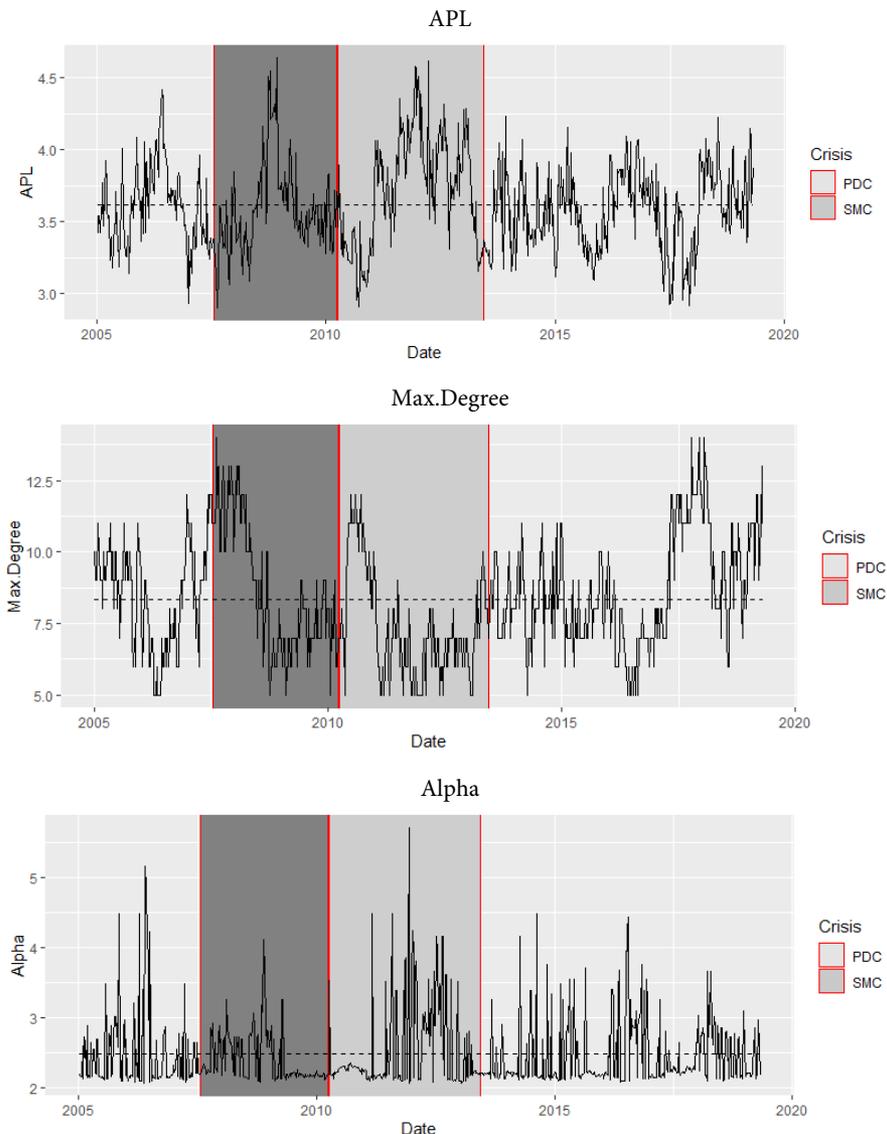

Figure 1. Average distances (APL), maximum degrees (Max.Degree) and estimated parameters α of power distribution α for MST in the period 07.01.2005 – 26.04.2019 with the subprime mortgage crisis (SMC) and the public debt crisis (PDC) highlighted
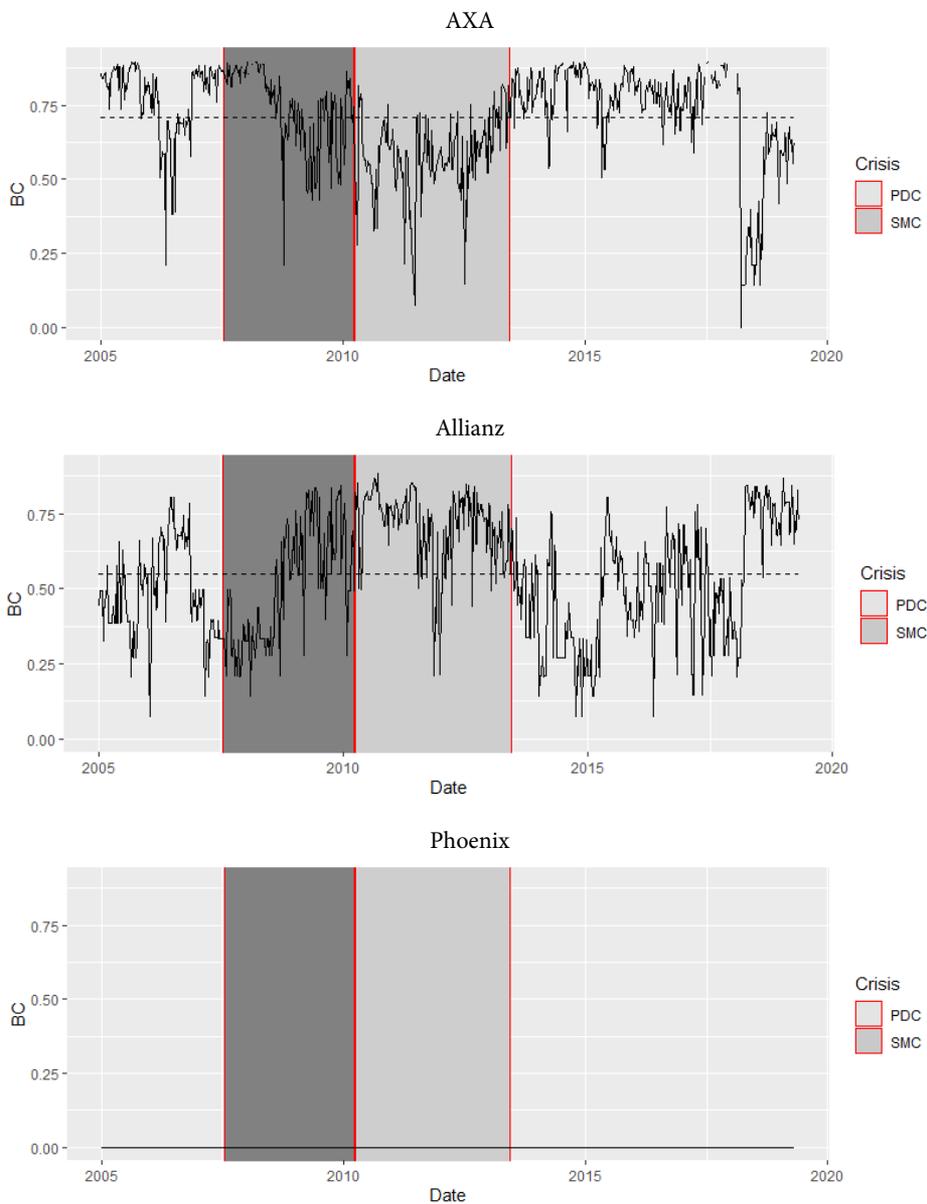
Source: Own study.

Figure 2. BC for selected insurance institutions (AXA, Allianz and Phoenix) during the period 07.01.2005 – 26.04.2019 with the subprime mortgage crisis (SMC) and the public debt crisis (PDC) highlighted
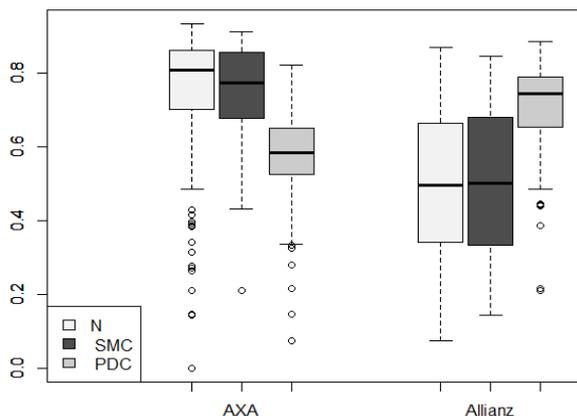
Source: Own study.

Figure 3. BC distribution for AXA and Allianz in the normal state of the market (N), during the subprime mortgage crisis (SMC) and the public debt crisis (PDC)

Source: Own study.

We examine the relationship between the structure of the network of connections (MST) and the contribution to systemic risk based on time series deltaCoVaR measures, determined for individual insurers in accordance with the empirical strategy of the second step of the study. In each analysed period, we determine the deltaCoVaR measure for each insurer, assuming $\alpha = \beta = 0.05$. Figure 4 shows the average BC and Figure 5 shows the average deltaCoVaR over the period under consideration for all 28 companies analysed. Comparison of these diagrams shows that BC and deltaCoVaR levels are related in extreme situations. In clear, for institutions with high BC, the deltaCoVaR value is the smallest, which means the largest contribution to systemic risk, for institutions with BC at zero level, the deltaCoVaR is the highest (i.e. the lowest contribution to systemic risk). So AXA and Allianz contribute to systemic risk to a much greater extent than Phoenix. However, it cannot be argued that companies that have a low BC do not generate systemic risk. The diagram shows that the contribution to the risk of companies with low BC is comparable to that of companies with high BC. Figure 6 shows how the deltaCoVaR depends on the mean BC. It can be seen that with the increase of BC of insurer vertices, the deltaCoVaR decreases (the contribution to systemic risk increases). Figure 7 presents a summary of MST and a diagram of the deltaCoVaR dependence on BC in the period when all analysed companies have the smallest deltaCoVaR, which happens to coincide with the middle of the crisis, that is October 17th, 2008. The tree during this period is such that AXA, Allianz and Aegon,

having the largest BC and the largest contribution to systemic risk are directly related to each other. In Figure 8, we compile the contribution of AXA and Phoenix to systemic risk, i.e. the two companies that have the lowest and highest average deltaCoVaR, respectively. We observe large differences between these companies in the determined market states: the normal one and two crisis states.
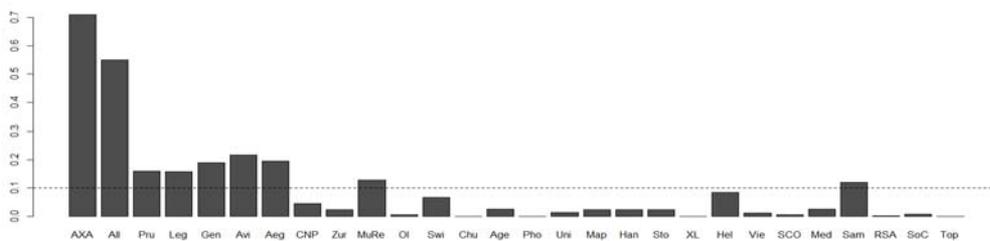

Figure 4. Average value of BC in the period under consideration for individual insurance institutions, the line showing the arithmetic mean of all the average values
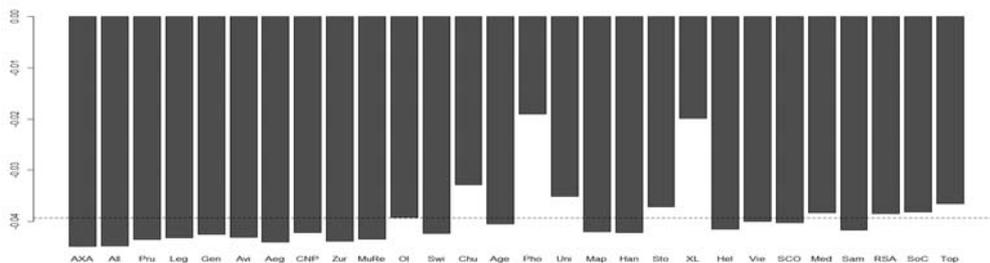
Source: Own study.

Figure 5. The average deltaCoVaR value over the period under consideration for individual insurers, the line showing the arithmetic mean of all the average values
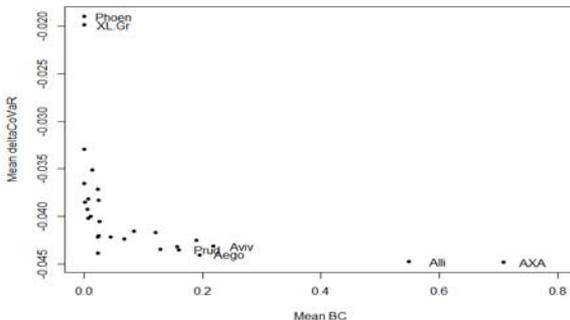
Source: Own study.

Figure 6. The relationship between the average BC and deltaCoVaR values over the period under consideration for individual insurers

Source: Own study.
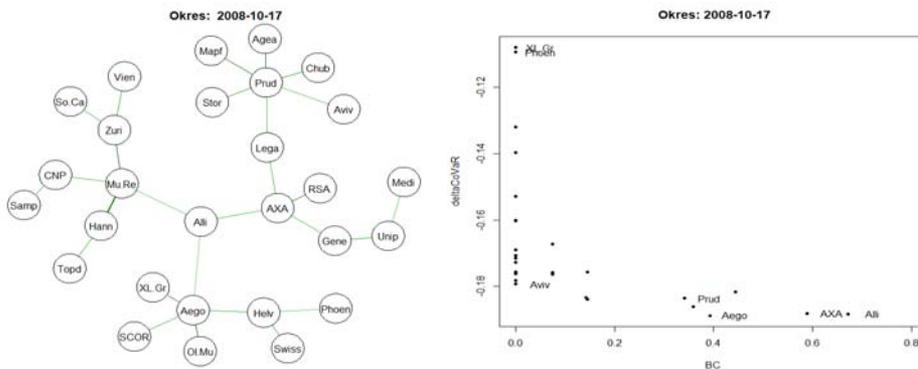
Figure 7. MST[3] and the relationship between BC and deltaCoVaR in the period with the lowest deltaCoVaR for AXA and Allianz

Source: Own study.

---

[3] AXA, Allianz (Alli), Prudential plc (Prud), Legal & General (Lega), Generali (Gene), Aviva (Aviv), Aegon (Aego), CNP Assurances (CNP), Zurich Insurance (Zuri), Munich Re (Mu Re), Old Mutual (Ol Mu), Swiss Life (Swiss), Chubb Ltd (Chub), Ageas (Agea), Phoenix (Phoen), Unipol Gruppo (Unip), Mapfre (Mapf), Hannover (Hann) Re, Storebrand (Stor), XL.Group (Xl Gr), Helvetia Holding (Helv), Vienna Insurance (Vien), SCOR SE (SCOR), Mediolanum (Medi), Sampo Oyj (Samp), RSA Insurance Group (RSA), Società Cattolica di Assicurazione (So Ca), Topdanmark A/S (Topd).
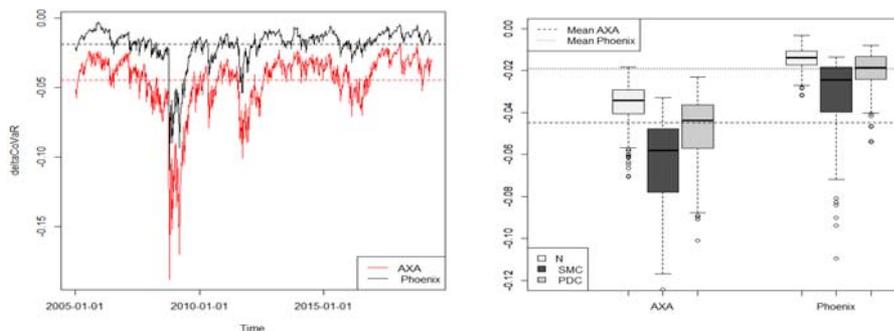
Figure 8. DeltaCoVaR for AXA and Phoenix in the period under consideration (left panel) and its distribution in the normal state of the market (N), during the subprime mortgage crisis (SMC) and the public debt crisis (PDC) (right panel)

Source: Own study.

5. Conclusions

Analyzing the time series of topological APL, MD, BC indicators and the MST alpha indicator and the MST structure (Wanat and Denkowska, 2019), we conclude that when the market is in a normal state, the series show volatility, but they do not have a large amplitude. MSTs change their structure and the linkages between the companies are different depending on time. We also note that to assess the potential risk of a rapid spread of the crisis, it is necessary to analyze all the four indicators. Indeed, with high correlation, the tree becomes very dispersed, which means that assessing the linkages only on the basis of the correlation coefficient could lead to erroneous conclusions. From the MST analysis, a clear shrinkage of the network can be seen in the periods of June 2nd, 2006 - August 17th, 2007, i.e. just before the subprime crisis and during its first phase, and in the period December 5th, 2008 – September 17th, 2010, i.e. before and at the beginning of the European public debt crisis. If MSTs are shrunk, it promotes potential propagation of financial problems. However, during the subprime crisis itself, the trees changed their appearance. They were relaxed: APL increased, MaxDegree decreased.

In this study, we analysed 28 largest insurance companies from the point of view of the contribution of each institution to systemic risk in accordance with the currently used deltaCoVaR measure. The analysis of time series shows that in the period from 2005 to 2019 for each of the companies there is an obvious relation between its contribution to systemic risk and the structure of the network of connections (MST). During the entire period, the contribution of each company remains at the same level, save for the clearly apparent period during which the deltaCoVaR decreases and,

consequently, the contribution to the systemic risk increases, and this happens at the very centre of the subprime crisis, October 17th, 2008. As the deltaCoVaR changes, the APL ratio increases. We should emphasize that for the entire analysed period, it reaches its maximum exactly on December 5th, 2008.

The identified relationship between the contribution to systemic risk and the minimum spanning tree structure described by topological network indicators can be used in the construction of models whose task is to predict the possibility of systemic risk. The construction of this type of predictive models is the subject of further research by the authors.

Acknowledgements

References

Acharya, V. V., Pedersen, L. H., Philippon, T. and Richardson, M. P., (2010). Measuring Systemic Risk, SSRN Electronic Journal, doi:10.2139/ssrn.1573171

Adrian, T., Brunnermeier, M., (2011). Covar. Princeton University Working Paper.

Alves, I., Brinkhoff, J., Georgiev, S., Heam, J-C., Moldovan, I., and di Carlo M.S., (2015). Network analysis of the EU insurance sector, ESRB Occasional Paper Series 07, European Systemic Risk Board.

Baluch, F., Mutenga, S. and Parsons, C., (2011). Insurance, Systemic Risk and the Financial Crisis, "The Geneva Papers on Risk and Insurance - Issues and Practice", Vol. 36(1), pp. 126–163, doi:10.1057/gpp.2010.40.

Bednarczyk, T., (2013). Czy sektor ubezpieczeniowy kreuje ryzyko systemowe?, „Studia Oeconomica Posnaniensia", Vol. 11 (260), pp. 7–17.

Bell, M., Keller, B., (2009). Insurance and Stability: The Reform of Insurance Regulation, Zurich Financial Services Group Working Paper.

Bierth, C., Irresberger, F. and Weiß, G. N., (2015). Systemic risk of insurers around the globe. "Journal of Banking & Finance", 55, pp. 232–245, doi:10.1016/j.jbankfin.2015.02.014.

Billio, M., Getmansky, M., Lo, A. W. and Pelizzon, L., (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors, "Journal of Financial Economics", Vol. 104(3), pp. 535–559, doi:10.1016/j.jfineco.2011.12.010.

Chen, H., Cummins, J. D., Viswanathan, K. S. and Weiss, M. A., (2013). Systemic Risk and the Interconnectedness Between Banks and Insurers: An Econometric Analysis, "Journal of Risk and Insurance", Vol. 81(3), pp. 623–652, doi:10.1111/j.1539-6975.2012.01503.x.

Cummins, J. D., Weiss, M. A., (2014). Systemic Risk and The U.S. Insurance Sector, "Journal of Risk and Insurance", Vol. 81(3), pp. 489–528, doi:10.1111/jori.12039.

Czerwińska, T., (2014). Systemic risk in the insurance sector, „Problemy Zarządzania", Vol. 12(48), pp. 41–63, doi:10.7172/1644-9584.48.3.

EIOPA, (2017). Systemic risk and macroprudential policy in insurance. Luxembourg: Publications Office of the European Union.

Geneva Association, (2010). Systemic risk in insurance: an analysis of insurance and financial stability, Technical report, Special Report of The Geneva Association Systemic Risk Working Group, Switzerland.

Harrington, S. E., (2009). The Financial Crisis, Systemic Risk, and the Future of Insurance Regulation, "Journal of Risk and Insurance", Vol. 76(4), pp. 785–819.

IAIS, (2013). Global Systemically Important Insurers: Final Initial Assessment Methodology, International Association of Insurance Supervisors, Basel.

IAIS, (2016). Global Systemically Important Insurers: Updated Assessment Methodology, International Association of Insurance Supervisors, Basel.

Mantegna, R. N., (1999). Hierarchical structure in financial markets, Eur. Phys. J. B 11(1), pp. 193–197.

Weiß, G. N., Mühlnickel, J., (2014). Why do some insurers become systemically relevant?. "Journal of Financial Stability", Vol. 13, pp. 95–117.

Wanat, S., Denkowska, A., (2019). Linkages and systemic risk in the European insurance sector: Some new evidence based on dynamic spanning trees, arXiv:1908.01142.

William, K., Sjostrum, Jr., (2009). The AIG Bailout, 66 Wash. & LeeL. Rev. 943.

A comparative study of a class of direct estimators for domain mean with a direct ratio estimator for domain mean using auxiliary character

Brij Behari Khare[1], Ashutosh[2], Piyush Kant Rai[3]

ABSTRACT

Estimation techniques for a domain parameter play a very significant role in the theory of sample surveys. In the recent years many advanced methodologies have been developed for domain estimation. In particular, direct and synthetic estimators are applied for the estimation of domain mean in the government and private sectors under certain assumptions as to the size of the samples relating to particular domains. The findings demonstrate that the direct estimator fails to perform more efficiently as compared to the synthetic estimator when reliable units are not directly accessible in the studied domains. Moreover, due to the fact that small units belong to the sample of the studied domain, the direct estimator produces an unacceptably large standard error. In contrast, if a sufficient number of units are available in the studied domain, the direct estimator produces effective results. This paper presents the theoretical aspects of the proposed class of direct estimators for domain mean with the use of a single auxiliary character, compared with an existing direct ratio estimator for domain mean (given in section 3.2). In addition, an empirical study has been provided to support the validity of the proposed estimators. The findings prove that the proposed estimators outperform the direct ratio estimator for domain mean using a single auxiliary character in the case of two studied populations and their analysed domains considered from Sarndal et al. (1992).

Key words: domain, auxiliary character, direct ratio estimator, class of estimators, mean square error (MSE).

1. Introduction

If we are interested in the estimation of subpopulations also called domains like a block, a county and a village, etc., instead of whole population. It has been seen in recent years that the accelerated demand for policy implementation and decision-

[1] Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi-221005, India.
E-mail: bbkhare56@yahoo.com.
[2] Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi-221005, India.
E-mail: kumarashubhustat@gmail.com. ORCID: https://orcid.org/0000-0002-0183-6083.
[3] Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi-221005, India.
E-mail: raipiyush5@gmail.com. ORCID: https://orcid.org/0000-0001-8462-4707.

makers, different types of estimation methods have been developed, which may solve these types of problems. There are two types of methods direct method and indirect method of estimations are used in the estimation of domain parameter. The direct method is generally used when the number of sufficient units is accessible in the study domain. In the direct method, we take a sample from the study domain and it is applied in the estimator which may improve the efficiency of the estimator. The estimator based on the sole of direct method has been illustrated in the book by Rao (2003). Whenever, accessible numbers of units do not sufficiently in the study domain, we prefer an indirect method based estimator. In this situation, a sample is selected from the whole population instead of subpopulation. Eminent works have been already done using the indirect method of estimation, e.g. synthetic estimators using auxiliary character have been illustrated by Gonzalez (1973), Tikkiwal and Ghiya (2000), Rai and Pandey (2013) and Khare and Ashutosh (2017), among many others in sample surveys. We mainly collect the fact from the surrounding value of auxiliary and study information and employ them in the estimator to improve the efficiency of the estimator. An idea of class of estimator to estimate the population mean has been given by Srivastava (1971).

In this paper, he proposed an estimator $\bar{y}_h = \bar{y}\, h(u)$, where $u = \dfrac{\bar{x}}{\bar{X}}$, under certain regularity conditions, the asymptotic MSE is same for all its members. In his unpublished work, he did the extension of his own above paper of a wider class of estimator, which is $\bar{y}_g = g\left(\bar{y}, \dfrac{\bar{x}}{\bar{X}}\right)$ where the function $g(.,.)$ satisfies the regularity conditions. Furthermore, another work related to the class of estimator for population mean was discussed by Srivastava and Jhajj (1981). They also developed a class of estimator for finite population mean using single auxiliary character x according to some parametric function h(.), which satisfied certain regularity conditions along with the limitation of $h_1=1$, and also it was shown that the lower bound of the asymptotic MSE of the estimator is equal as the asymptotic MSE of the linear regression estimator, which is itself not a member of the class of estimator developed by Srivastava and Jhajj (1981). Another work was done by Srivastava (1983) which may get an improve version of the above paper, in which he incorporated another parameters called variance, and he suggested $\bar{y}_g = g\left(\bar{y}, u, v\right)$ where $u = \dfrac{\bar{x}}{\bar{X}}$ and $v = \dfrac{s_x^2}{S_x^2}$.

where, $\bar{y}$ = Sample mean of study character,

$\bar{x}$ = Sample mean of auxiliary character,

$\bar{X}$ = Population mean of auxiliary character,

s_x^2 = Sample mean square of auxiliary character x and

S_x^2 = Population mean square of auxiliary character x.

Other works have also been done which are related to the class of estimator using two phase sampling scheme for estimation of the population mean discussed by Srivastava and Khare (1993), Khare and Pandey (2000), and Khare and Sinha (2009). In the coming year work has been done by Khare et al. (2018). They have proposed a class of synthetic estimators for domain parameter like mean, and a function which combined value of u and v, the estimator is given by

$$T_{SC,a} = f(u,v)$$

where $u = \bar{y}$ and $v = \dfrac{\bar{x}}{\overline{X}_a}$,

$\bar{y}$ = Sample mean of population of study character.

$\bar{x}$ = Sample mean of population of auxiliary character.

$\overline{X}_a$ = Population mean of ath Domain of auxiliary character.

In the present paper, we obtained MSE of the members of the proposed estimators for domain mean $\overline{Y}_a$ is equal under the certain regularity conditions but their values of constants are different. And we proposed a class of direct estimator for domain mean using auxiliary character, which is given by

$$T_{D,C,a} = h(u,v)$$

where $u = \bar{y}_a$ and $v = \dfrac{\bar{x}_a}{\overline{X}_a}$, which is given in the further section.

The particular cases of the proposed class of direct estimator are also discussed for domain mean. A comparative study of the proposed estimator for domain mean $(T_{D,C,a})$ with direct ratio estimator for domain mean $(T_{D,RS,a})$ has been given by using the real data of Swedish municipalities (Sarndal et al. (1992)).

2. Formulation of the problem and notations for domains

Suppose that non-overlapping domains U_a of size N_a such that (a=1, 2, 3,.., A). Now, our interest is in the estimation of the parameter of the domain mean $\overline{Y}_a$ of ath domain with size N_a. Later, we selected a sample $'s'$ through simple random sampling without replacement (SRSWOR) in which come from a^{th} domain have size n_a from

domain population size N_a. We represent the study character and auxiliary character by y and x respectively.

We denote the population mean and sample mean for domain of y and x as follows:

$\overline{Y}_a$: a^{th} domain meanof y based on size N_a observations.

$\overline{y}_a$: Sample mean of y of ath domain based on n_a observations.

$\overline{X}_a$: a^{th} domain mean of x based on size N_a observations.

$\overline{x}_a$: Sample mean of x of ath domain based on n_a observations.

Let us denote y_{a_i} as the ith observation of ath domain of the study character y for domain U_a ($a = 1, 2,..., A,\ i = 1, 2,..., N_a$) and x_{a_i} is the ith observation of ath domain of the auxiliary character x for domain U_a ($a = 1, 2,...A,\ i = 1, 2,..., N_a$).

We further use the following notations:

$$S_{Y_a}^2 = \frac{1}{(N_a - 1)}\sum_{i=1}^{N_a}\left(y_{a_i} - \overline{Y}_a\right)^2, \quad S_{X_a}^2 = \frac{1}{(N_a - 1)}\sum_{i=1}^{N_a}\left(x_{a_i} - \overline{X}_a\right)^2,$$

$$S_{X_aY_a} = \frac{1}{(N_a - 1)}\sum_{i=1}^{N_a}\left(x_{a_i} - \overline{X}_a\right)\left(y_{a_i} - \overline{Y}_a\right), \quad C_{Y_a} = \frac{S_{Y_a}}{\overline{Y}_a},$$

$$C_{X_a} = \frac{S_{X_a}}{\overline{X}_a} \text{ and } C_{X_aY_a} = \frac{S_{X_aY_a}}{\overline{X}_a\overline{Y}_a}.$$

$$(2.1)$$

3. Direct estimator for domain mean using single auxiliary character

There are several direct estimators which are used for estimation of population parameters of different segments. Auxiliary characteristics are used to improve the existing estimator for the domains. Here, in our case, we are considering the direct ratio estimator for estimating domain mean. Thus, let us consider the case of the direct ratio estimator under the above design and obtain the expressions of Bias and MSE in the next subsection.

3.1. Direct ratio estimator for domain mean

$$T_{D,RS,a} = \frac{\overline{y}_a}{\overline{x}_a}\overline{X}_a \quad \text{Tikkiwal and Ghiya (2000)} \tag{3.1.1}$$

$$Bias(T_{D,RS,a}) = \frac{(N_a - n_a)}{N_a n_a}\overline{Y}_a(C_{X_a}^2 - C_{X_aY_a}) \tag{3.1.2}$$

$$MSE(T_{D,RS,a}) = \frac{(N_a - n_a)}{N_a n_a} \overline{Y}_a^2 (C_{Y_a}^2 + C_{X_a}^2 - 2C_{X_a Y_a}) \qquad (3.1.3)$$

3.2. proposed class of direct estimators for domain mean using single auxiliary character $T_{D,C,a}$

We proposed a class of direct estimators for domain using auxiliary character, which is given as:

$$T_{D,C,a} = h(u,v) \qquad (3.2.1)$$

where, $u = \overline{y}_a$ and $v = \dfrac{\overline{x}_a}{\overline{X}_a}$ and the function $h(u,v)$ satisfied the following regularity conditions:

1. The function $h(u,v)$ exists for all the values of (u,v) and it contains the points $(\overline{Y}_a, 1)$ in a bounded subset D of two dimensional real spaces.

2. The first and second order partial derivatives of $h(u,v)$ exist and are bounded also.

Members of the estimator for C=1, 2 and 3 are given as follows:

$$T_{D,1,a} = uv^\alpha \qquad (3.2.2)$$

$$T_{D,2,a} = \alpha_1 u + (1-\alpha_1) u v^{\alpha_2} \qquad (3.2.3)$$

$$T_{D,3,a} = u e^{\alpha_3 \frac{(v-1)}{(v+1)}} \qquad (3.2.4)$$

Now, expanding the proposed class of estimators $T_{D,C,a}$ using Taylor series expansion about the point $(\overline{Y}_a, 1)$ up to the second order, we have

$$T_{D,C,a} = h(u,v)_{(\overline{Y}_a,1)} + (u - \overline{Y}_a) h_1 + (v-1) h_2 +$$
$$+ \frac{1}{2} \left[(u - \overline{Y}_a)^2 + (v-1)^2 h_{22} + (v-1) h_{12} \right] \qquad (3.2.5)$$

where

$$h_1 = \left(\frac{\partial h(u,v)}{\partial u} \right)_{(\overline{Y}_a,1)}, h_2 = \left(\frac{\partial h(u,v)}{\partial v} \right)_{(\overline{Y}_a,1)}, h_{11} = \left(\frac{\partial^2 h(u,v)}{\partial u^2} \right)_{(\overline{Y}_a,1)},$$

$$h_{22} = \left(\frac{\partial^2 h(u,v)}{\partial v^2} \right)_{(\overline{Y}_a,1)}$$

and $h_{12} = \left(\frac{\partial^2 h(u,v)}{\partial u \partial v} \right)_{(\overline{Y}_a,1)}.$

$$(3.2.6)$$

Now, we put $h_1 = 1$ and $h_{11} = 0$ in the equation (3.2.5), and we have

$$T_{D,C,a} = h(u,v)_{(\overline{Y}_a,1)} + (v-1)h_2 + \frac{1}{2}[(v-1)^2 h_{22} + 2(v-1)(u-\overline{Y}_a)h_{12}].$$

$$(3.2.7)$$

For large sample approximations, we assume that

$$\overline{y}_a = \overline{Y}_a(1+\varepsilon_0), \overline{x}_a = \overline{X}_a(1+\varepsilon_1), \text{ such that } E(\varepsilon_0) = 0, E(\varepsilon_1) = 0,$$

$$E(\varepsilon_0^2) = \frac{(N_a - n_a)}{N_a n_a}C_{Y_a}^2 \quad E(\varepsilon_1^2) = \frac{(N_a - n_a)}{N_a n_a}C_{X_a}^2 \quad \text{and}$$

$$E(\varepsilon_0\varepsilon_1) = \frac{(N_a - n_a)}{N_a n_a}C_{X_a Y_a}.$$

$$(3.2.8)$$

The Bias and MSE of the proposed class of the estimators for domain mean using auxiliary character is obtained as:

$$Bias(T_{D,C,a}) = \frac{(N_a - n_a)S_{X_a}^2 h_2}{2N_a n_a \overline{X}_a^2}\left(\frac{h_2}{\overline{Y}_a} - 1\right) + \frac{(N_a - n_a)S_{X_a Y_a}}{N_a n_a \overline{X}_a \overline{Y}_a}h_2 \qquad (3.2.9)$$

$$MSE(T_{D,C,a}) = E\left[(\overline{y}_a - \overline{Y}_a)^2 + \left(\frac{\overline{x}_a}{\overline{X}_a} - 1\right)h_2 + \frac{1}{2}\left\{\left(\frac{\overline{x}_a}{\overline{X}_a} - 1\right)^2 h_2^2\right.\right.$$

$$\left.\left. + 2\left(\frac{\overline{x}_a}{\overline{X}_a} - 1\right)(\overline{y}_a - \overline{Y}_a)\frac{h_2}{\overline{Y}_a}\right\}\right]$$

$$(3.2.10)$$

Now, for optimum value of h_2, we partially differentiate equation (3.2.10) w.r.to h_2 and equating to zero, we have

$$h_{2,opt} = -\frac{\overline{X}_a \rho_a \sqrt{S_{Y_a}^2}}{\sqrt{S_{X_a}^2}} \qquad (3.2.11)$$

After substituting the value of $h_{2,opt}$ in the equation (3.2.10) the optimum MSE of $T_{D,C,a}$ is given by

$$MSE(T_{D,C,a,opt}) = (1 - \rho_a^2)\frac{(N_a - n_a)}{N_a n_a}S_{Y_a}^2 \qquad (3.2.12)$$

Theorem 1. The values of the constants (given in equations 3.2.2, 3.2.3 and 3.2.4) of the member of the proposed estimators, which are included in $T_{D,1,a}$, $T_{D,2,a}$ and $T_{D,3,a}$ after minimizing their individual MSE expressions, are given as follows:

$$\alpha_{opt} = -\frac{\overline{X}_a \rho_a \sqrt{S_{Y_a}^2}}{\overline{Y}_a \sqrt{S_{X_a}^2}} \tag{3.2.13}$$

$$\alpha_{2,opt} = -\frac{\overline{X}_a \rho_a \sqrt{S_{Y_a}^2}}{(1-\alpha_1)\overline{Y}_a \sqrt{S_{X_a}^2}} \quad \text{where,} 0<\alpha_1<1 \tag{3.2.14}$$

$$\alpha_{3,opt} = -\frac{2\overline{X}_a \rho_a \sqrt{S_{Y_a}^2}}{\overline{Y}_a \sqrt{S_{X_a}^2}} \tag{3.2.15}$$

The minimum values of the MSE of the estimators $T_{D,1,a}$, $T_{D,2,a}$ and $T_{D,3,a}$ for optimum values of the constants α_{opt}, $\alpha_{2,opt}$ and $\alpha_{3,opt}$ are the same and given in the equation (3.2.12), the optimum value of the constants α_{opt}, $\alpha_{2,opt}$ and $\alpha_{3,opt}$ are given in the form of the parameters in the equations (3.2.13), (3.2.14) and (3.2.15), it may be possible use of the optimal values using the past data regarding parameters given by Reddy (1978), and it has been seen that in the terms of order n^{-1}, the minimum value of the MSE of the estimator does not change when we estimate the optimal value of the constants using the sample values of idea given by Srivastava and Jhajj (1981).

3.3. Comparison between proposed class of estimators and direct ratio estimator for domain mean using auxiliary character

Let us consider $MSE(T_{D,RS,a}) - MSE(T_{D,C,a,opt}) \geq 0$

$$= \frac{(N_a - n_a)}{N_a n_a}\overline{Y}_a^2\left(C_{Y_a}^2 + C_{X_a}^2 - 2C_{X_a Y_a}\right) - \left(1 - \rho_a^2\right)\frac{(N_a - n_a)}{N_a n_a}S_{Y_a}^2$$

$$= \frac{(N_a - n_a)}{N_a n_a}\overline{Y}_a^2\left(C_{X_a}^2 - 2C_{X_a Y_a} + \rho_a^2 C_{Y_a}^2\right)$$

$$= \frac{(N_a - n_a)}{N_a n_a}\overline{Y}_a^2\left(\rho_a C_{Y_a} - C_{X_a}\right)^2 \tag{3.3.1}$$

Since $\left(\rho_a C_{Y_a} - C_{X_a}\right)^2$ must be positive. Hence,

$$MSE(T_{D,RS,a}) - MSE(T_{D,C,a,opt}) \geq 0$$

$$\text{i.e. } MSE(T_{D,RS,a}) \geq MSE(T_{D,C,a,opt}) \tag{3.3.2}$$

4. Empirical study

For the purpose of empirical study, we considered the data from Sarndal et al. (1992) in the appendix B. The population of Swedish municipalities is classified into eight non-overlapping domains, but we consider only four domains i.e. 2, 3, 4 and 5 have sizes (48, 32, 38 and 56). The empirical study of the two populations (1 and 2), the information about population 1 and population 2 is given as follows:

Population 1

y= Revenues from the 1985 municipal taxation (in millions of kronor)

x= Real estate values according to 1984 assessment (in millions of kronor).

Table 4.1. The parameter values of the domains (1, 2, 3 and 4)

Domain Values	Domain			
	1	2	3	4
N_a	48	32	38	56
$\overline{Y}_a$	233.69	176.13	265.74	273.30
$\overline{X}_a$	2970.958	2498.75	2915.526	3046.946
$S^2_{X_a}$	11118969	4474735	27860176	27861139
$S^2_{Y_a}$	93788.43	32183.08	311726.60	788518.90
$S_{X_a Y_a}$	990772.90	344998.30	1621192.00	4518431.00
ρ_a	0.970	0.942	0.938	0.964

Population 2

y=Real estate values according to 1984 assessment (in millions of kronor).

x= Number of municipal employees in 1984.

Table 4.2. The parameter values of the different domains (1, 2, 3 and 4)

Domain Values	Domain			
	1	2	3	4
N_a	48	32	38	56
$\overline{X}_a$	1658.708	1316.938	1937.5	1950. 393
$\overline{Y}_a$	2970.96	2498.75	2915.53	3046.95
$S^2_{X_a}$	4601899	1989177	15986523	38786393
$S^2_{Y_a}$	11118969	4164522	27860176	27861139
$S_{X_a Y_a}$	6920432	2681882	11697923	31770622
ρ_a	0.967414	0.9317951	0.9454677	0.9664827

Table 4.3. MSE of the direct ratio estimator for domain mean using auxiliary character ($T_{D,RS,a}$) and MSE of the proposed estimators for domain mean using auxiliary character ($T_{D,C,a,opt}$) for the optimum values of $h_{2,opt}$ for all domains 1, 2, 3 and 4, also the value of different constants which exist in the proposed estimators (**for population 1**):

Estimator		Domains			
		1	2	3	4
$T_{D,RS,a}$		1203.66	1786.603	21417.37	36810.38
$T_{D,C,a,opt}$		986.148	1088.301	8333.17	10151.570
$h_{2,opt}$		-264.732	-207.002	-493.607	-494.144
α		-1.133	-1.175	-1.857	-1.808
α_1	0.25	(-1.510)	(-1.567)	(-1.896)	(-2.411)
	0.50	(-2.266)	(-2.351)	(-3.715)	(-3.616)
	0.75	(-4.531)	(-4.701)	(-7.430)	(-7.232)
α_3		-2.266	-2.351	-3.715	-3.616

() shows α_2 constant included in the estimator ($T_{D,2,a}$)

Table 4.4. MSE of the direct ratio estimator for domain mean using auxiliary character ($T_{D,RS,a}$) and the proposed estimators for domain mean using auxiliary character ($T_{D,C,a,opt}$) at optimum values of function $h_{2,opt}$ for all domains 1, 2, 3 and 4, also the value of constants which are included in the member of the estimators for population 2:

Estimator		Domains			
		1	2	3	4
$T_{D,RS,a}$		61034.97	96377.97	1044100	1735007
$T_{D,C,a,opt}$		52858.67	79172.95	379373.3	465640
$h_{2,opt}$		-1189.497	-1112.12	-4602.01	-4099.073
α		-0.717	-0.845	-2.375	-2.102
α_1	0.25	(-0.956)	(-1.126)	(-3.167)	(-2.802)
	0.50	(-1.434)	(-1.689)	(-4.750)	(-4.203)
	0.75	(-2.868)	(-3.378)	(-9.501)	(-8.407)
α_3		-1.434	-1.690	-4.751	-4.203

() shows α_2 constant included in estimator ($T_{D,2,a}$)

From the table (4.3) it is seen that the amount of MSE of the class of direct estimators for domain mean ($T_{D,C,a,opt}$) is less than the amount of MSE of the direct ratio estimator for domain mean ($T_{D,RS,a}$) for domain 1 and the value of $h_{2,opt}$=-264.732 but the value of the member of the constants α =-1.133, α_1 =0.25, α_2 =-1.510 and α_3 =-2.266 is different, and for domain 2, 3, and 4, the value of $h_{2,opt}$ is fixed while the value of constant is different for population 1.

From the table (4.4) it is seen that the amount of MSE of the class of direct estimator for domain mean ($T_{D,C,a,opt}$) is less than the amount of MSE of the direct ratio estimator for domain mean ($T_{D,RS,a}$) for domain 1 and the value of $h_{2,opt}$ =-1189.497 but the value of the member of the constants α =-0.717, α_1 =0.25, α_2 =-0.956 and α_3 =-1.434 is different. This pattern is also seen for others domains 2, 3 and 4 for population 2.

Table 4.5. Percentages Relative Efficiency (PRE) of the proposed estimator for domain mean ($T_{D,C,a,opt}$) to direct ratio estimator for domain mean ($T_{D,RS,a}$) for different domains 1, 2, 3 and 4 (**for population 1 and population 2**):

Population	Estimators	Domains			
		1	2	3	4
1	$T_{D,RS,a}$	100.000	100.000	100.000	100.000
	$T_{D,C,a,opt}$	122.057	164.164	257.014	362.608
2	$T_{D,RS,a}$	100.000	100.000	100.000	100.000
	$T_{D,C,a,opt}$	115.468	121.731	275.217	372.607

From the table (4.5), it is observed that the value of PRE of the proposed estimator for domain mean ($T_{D,C,a,opt}$) is higher than the PRE of the direct ratio estimator ($T_{D,RS,a}$) for all domains 1, 2, 3 and 4 for population 1 and population 2.

5. Conclusion and recommendations

It is emphasized that the MSE of the proposed class of estimators for domain mean is less than the corresponding MSE of the direct ratio estimator for the domain mean in both populations considered for empirical analysis for nearly all domains. Also, the results for MSE supported the superiority of the proposed estimators theoretically as compared to the direct ratio estimator. There are some deviations in the result of MSE of the proposed estimator for the first and second domains as compared to the results for the third and fourth domains. It may be due to the variation present in the observations. PRE is also calculated for the proposed estimator and derived the results for family of estimators under certain regularity conditions given in the literature. Also, it is shown that the values of three constants available in the proposed member of family of estimators are different while the function $h_{2\,opt}$ is fixed under certain regularity conditions in the domains for both first and second populations considered for analysis.

Thus, it is recommended that the class of direct estimators proposed in this article for the estimation of domain mean using proper auxiliary information have substantial utility in the domain estimation methodology as compared to the existing direct ratio estimator under the condition that a sufficient member of units fall in the domain concerned.

References

Gonzalez, M.E., (1973). Use and Evaluation of Synthetic Estimators, Proceedings of the Social Statistics, American Statistical Association, pp. 33–36.

Khare, B. B., Pandey, S. K., (2000). A Class of Estimators for Ratio of Two Population Means using Auxiliary Character in the Presence of Non-Response. Journal of Scientific Research, BHU, 50, pp. 115–125.

Khare, B. B., Sinha. R. R., (2009). On Class of Estimators for Population Mean using Multi-Auxiliary Characters in the Presence of Non-Response, Statistics in Transition-New Series, 10, 1, pp. 3–14.

Khare. B. B., Ashutosh., (2017). Synthetic Estimator for Domain Mean Using Auxiliary Character, International Journal of Statistics and Economics, 18, 4, pp. 32–44.

Khare, B. B., Ashutosh and Khare. S., (2018). Comparative Study of Synthetic Estimators with Ratio Estimator for Domain Mean in Survey Sampling using Auxiliary Character. International Journal of Applied Mathematics and Statistics, 57, 3, pp. 152–161.

Rai, P. K., Pandey, K. K., (2013). Synthetic Estimators using Auxiliary Information in Small Domains. Statistics in Transition-New Series, 14, 1, 31–44.

Rao. J. N. K. (2003). Small Area Estimation, Wiley Inter-Science, John Wiley and Sons, New Jersey.

Reddy, V. N., (1978). A Study on the Use of the Prior Knowledge on Certain Population Parameters in Equation. Sankhya. 40, pp. 29–37.

Sarndal, C. E. Swensson, B. and Wretman, J. H. (1992). Model Assisted Survey Sampling, Springer-Verlog, New York.

Srivastava. S. K. (1971). A Generalized Estimators for the Mean of a Finite Population using Multi-Auxiliary Information, International Journal of American Statistical Association. 66, pp. 404–407.

Srivastava. S. K., Jhajj. H. J., (1981). A class of Estimators of the Population Mean in Survey Sampling using Auxiliary Information. Biometrika. 68, 1, pp. 341–343.

Srivastava, S. K., Jhajj, H. J., (1983). A Class of Estimators of the Population Mean using Multi Auxiliary Information. Calcutta Statistical Association Bulletin. 15, pp. 127–134.

Tikkiwal. G. C., Ghiya. A., (2000). A Generalized Class of Synthetic Estimators with Application to Crop Acreage Estimation for Small Domains. Biometrical Journal. 42, 7, pp. 865–876.

APPENDIX

Population 1

Figure1. Mean Square Error for different domains (1,2,3 and 4)
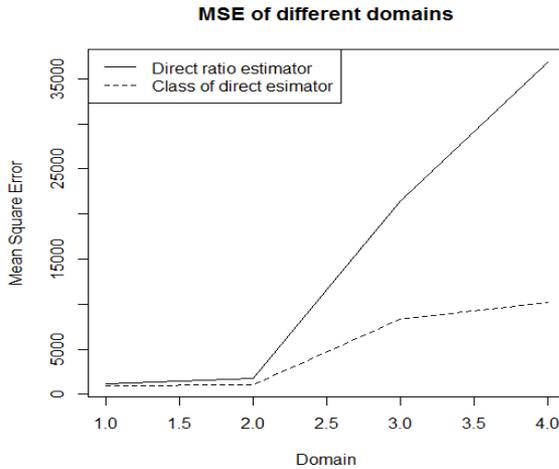
y axis: Mean Square Error, x axis: Domain

Population 2

Figure 2. Mean Square Error for different domains (1,2,3 and 4)
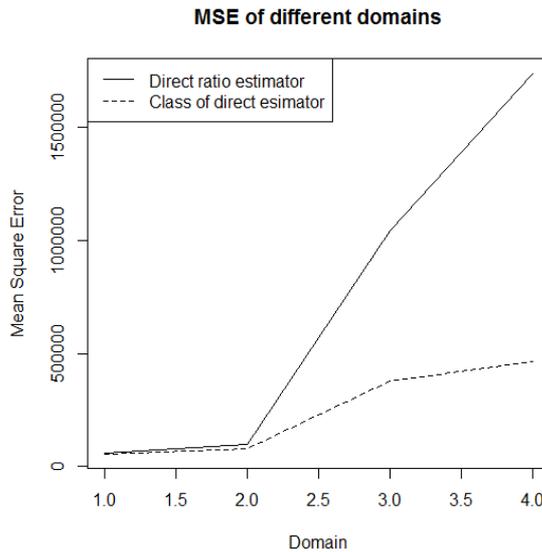
y axis: Mean Square Error, x axis: Domain

On behalf of the IAOS authorities we are pleased to announce the organization
of the conference which will take place on April 26–28, 2022
at the Convention Center in Cracow, Poland

The general theme of the IAOS-2022 Conference is:
Worthy Information for Challenging Times.

For more information on this event look at:

https://www.iaos-isi.org/index.php/latestnews/288-announcing-the-iaos-2022-conference
https://www.iaos-isi.org/index.php/latestnews/289-more-information-on-the-2022-iaos-conference-in-krakow

INTRODUCTION

Today's university faces the challenge of conducting world-class scientific research, fostering international cooperation and disseminating scientific developments. With this end in view it is necessary to follow and acknowledge the best practices adopted by universities and their researchers across the world.

The University of Economics in Katowice, committed to the pursuit of scientific excellence, has resolved to recognise the achievements of an eminent scientist indefatigably involved in the popularisation of science, who works closely with our Alma Mater.

The person to be awarded the highest academic honour – the doctor honoris causa degree – is Professor Malay Ghosh, Distinguished Professor at the University of Florida, a world authority in the field of statistical theory and its applications in, inter alia, economic and social sciences.

Professor Malay Ghosh is our highly valued partner, who has been cooperating with the Department of Statistics, Econometrics and Mathematics at the College of Management. We wholeheartedly welcome the fact that the accomplishments of the Katowice research team have caught Professor's attention, and that our shared scientific interests translate into specific initiatives, such as visiting lectures, international seminars and publications.

Professor Malay Ghosh has thereby become our University's global ambassador. It is, therefore, a great honour and pleasure to confer the degree of doctor honoris causa on Professor Ghosh and thus admit him into the community of the University of Economics in Katowice.

Vivat Academia, Vivant Professores!

Rector of the University of Economics in Katowice **Prof. Celina M. Olszak, PhD,**
D.Sc. Head of the College of Management **Prof. Wojciech Dyduch, PhD**

REVIEWS

Professor Malay Ghosh's outstanding contribution to statistics

Professor Yves G. Berger, PhD
University of Southampton, United Kingdom, 22ⁿᵈ January 2020

Professor Malay Ghosh's contribution to theoretical statistics covers a wide range of area, such as survey statistics, order and nonparametric statistics, sequential analysis, decision theory, Bayesian statistics, and small-area estimation. Malay also contributed to applied research projects on prostate cancer studies, case-control studies, quality assurance, county-level estimation and on detection of exoplanets.

Professor Malay Ghosh obtained a BA in Statistics in 1962 from the University of Calcutta, and a MA in 1964 from the same university. Afterwards, he moved to the United States and obtained his PhD degree in 1969 from the University of North Carolina at Chapel Hill, under the supervision of Pranab K. Sen. His PhD dissertation title is "Asymptotically optimal nonparametric tests for miscellaneous problems of linear regression". After his PhD, Malay took several research assistant positions at the University of North Carolina. He was an associate professor at the Indian Statistical Institute from 1971 to 1978, then became a professor at the Iowa State University in 1978. In 1982, Malay joined the University of Florida, Gainesville, where he became the Distinguished Professor of the Department of Statistics in 1998.

Professor Malay Ghosh has co-authored two books, published about 315 research manuscripts and supervised sixty PhD students, including Partha Lahiri, Gauri S. Datta, Kannan Natarajan and Nitis Mukhopadhyay. His contribution to small area estimation spans over two decades. With his PhD students, Partha Lahiri, Gauri Datta and Kannan Natarajan, Malay was the first to develop a unified Bayesian approach for solving small area estimation problems. He served from 1996 to 2001 in the United States Census Advisory Committee as a representative of the American Statistical Association.

Malay's other methodological contributions to Survey Statistics include development of new empirical Bayes confidence intervals based on Edgeworth expansion, outlier adjustment, and use of measurement error models. Malay has applied Bayesian and empirical Bayesian methods for the adjustment of census counts, estimation of median income of four-person families, estimation of the proportion of

people without health insurance for small domains cross-classified by age, sex, ethnicity and other characteristics.

Malay was the editor of "Sequential Analysis" for eight years, and "Sankhya, B" for four years. He is currently a member of the Editorial Board of "Statistics in Transition". Previously, he was part of the Editorial Board of the "Journal of Statistical Research" and the "Brazilian Journal of Statistics". He acted as Associate Editor for many journals: "Journal of the American Statistical Association", "Sequential Analysis", "Statistics and Decisions", "Communications in Statistics – Theory and Methods", "Journal of Nonparametric Statistics", "Journal of Statistical Planning and Inference", "Annals of Statistics", "Statistics", "American Statistician" and "Metron". Malay published about 315 papers, some in the most prestigious journals. In January 2020, he has 5,638 citations on researchgate.net and a ResearchGate score of 39.28, with an h-index of 36. Malay's score is higher than 95% of all ResearchGate members' scores. His most h-cited papers are Small area estimation: An appraisal (601 citations), Small area estimation: An approach (321 citations), both co-authored with J. N. K. Rao. Other highly cited papers are Penalized regression, standard errors, and Bayesian lassos (271 citations), Multivariate negative dependence (175 citations), Generalized linear models for small area estimation (156 citations), Sequential estimation (132 citations), Some remarks on non-informative priors (100 citations), Bayesian multivariate spatial models for roadway traffic crash mapping (99 Citations), On the invariance of noninformative priors (95 citations), Constrained Bayes estimation with applications (92 citations), Simultaneous estimation of parameters under entropy loss (91 citations), Statistical decision theory and Bayesian analysis (90 citations), Bayesian prediction in linear models: Applications to small area estimation (83 citations), Miscellanea. Second-order probability matching priors (81 citations), Bayesian methods for finite population sampling (78 citations). Five of these papers are published in the "Journal of the American Statistical Association", two papers are published in the "The Annals of Statistics". The other journals are "Bayesian Analysis", "Biometrika", "Statistical Science" and "Technometrics". Malay is also the co-author of two famous books: Sequential estimation (Wiley and Sons) with N. Mukhopadhyay and P.K. Sen, and Bayesian methods for finite population sampling (Chapman and Hall) with G. Meeden. Malay was principal investigator on several projects awarded by the National Science Foundation: "Simultaneous estimation of parameters in exponential families", "Admissibility in multiparameter estimation and in finite population sampling", "Multiparameter estimation and estimation in finite population sampling", "Empirical and hierarchical Bayes estimation in finite population sampling, quality assurance and random effects models", "Hierarchical and empirical Bayes estimation in survey sampling, linear models and quality assurance", "Bayesian methods and inference", "Bayesian Methods for small area estimation and latent structure models", "Parametric

and semiparametric Bayesian methods for small area estimation", "Bayesian and likelihood based multilevel models for small area estimation", "Some topics in small area estimation", "Empirical and hierarchical Bayesian methods with applications to small area estimation", "Case-control studies, new directions and applications", "Some contributions to sampling theory with applications" and "Bayesian empirical likelihood and penalized splines for small area estimation". Other projects are: "Nonparametric estimation and estimation in finite population sampling" (Research Development Award), "Conference on topics in generalized linear models" (US National Security Agency), "Bayesian neural networks for prostate cancer studies" (National Institute of Health), "Validation study of subdomain estimation method" (Centre of Diseases Control and Prevention), "Topics in Bayesian analysis, empirical likelihood and decision theory" (US National Security Agency), "Multivariate empirical and hierarchical Bayes methods for small area estimation", "Developing small area models for improved county-level estimation of agricultural cash rent" (U.S. Department of Agriculture), and "Small area estimation" (ASA Senior Research Fellowship). Malay collaborated in other research projects: "Order statistics and nonparametric statistics", "Conference on topics in generalized linear models" and "Search for Earth-mass planets: Bayesian Algorithms to analyse transit timing observations". Malay is an elected fellow of the American Statistical Association, the Institute of Mathematical Statistics, the International Statistical Institute and the International Society for Bayesian Analysis. In 1993, he received the Eugene Lukacs Distinguished Visiting Professorship of Statistics at the Bowling Green State University. Malay was awarded the Jerzy Spława-Neyman Medal at the 100th Anniversary of the Polish Statistical Society in 2012, the Lifetime Achievement Award from the International Indian Statistical Association in 2017 and the Small Area Estimation Award in 2019.

In May 2014, an international conference in honour of Professor Malay Ghosh, entitled "Frontiers of Hierarchical Modelling in Observational Studies, Complex Surveys, and Big Data", was hosted by the Joint Program in Survey Methodology, University of Maryland at College Park. Several areas to which Ghosh made substantial contributions were represented, including small-area estimation, objective Bayesian inference, hierarchical Bayesian modelling, and statistical inference for case-control studies. More than 200 people (including 16 of his doctoral students) celebrated Ghosh's outstanding contributions to statistics and his dedicated role as researcher, teacher and mentor. At last but not least, Malay taught a wide ranges of course: Intermediate probability and inference, Advanced inference, Sequential analysis, Nonparametric inference, Decision theory, Large sample theory, General theory of linear estimation and hypothesis testing, Multivariate analysis, Descriptive statistics, Statistical models, Statistical methods, Multivariate nonparametric inference,

Reliability theory, Introductory probability theory, Introductory inference, Advanced probability theory, Introductory statistical methods, Applied Bayesian statistics, Bayesian theory and likelihood.

The outstanding academic and research curriculum mentioned above are the reasons to propose Professor Malay Ghosh as a candidate for a Degree of Doctor Honoris Causa. I have no doubt that he is an ideal candidate, given the quality of his research curriculum during his impressive academic career. I strongly request that Professor Malay Ghosh be awarded the Degree of Doctor Honoris Causa by the University of Economics in Katowice.

Scientific achievements of Professor Malay Ghosh

Professor Dr. Ralf T. Münnich,
TrierUniversity, Trier, 28ᵗʰ January 2020

It was in January 2004 when Professor Ghosh entered the conference room at the IMS/ASA-SRMS Joint Mini Meeting, Raichak, West Bengal, India. Of course, I knew who had just entered the room when I was preparing my talk. But it was the incredible aura that surrounded him and made me pause. Quite unpretentiously he greeted me with a "Hi, I'm Malay" – this incredibly impressive researcher immediately showed his very friendly, warm and human manner with which he was no less impressive. This special professional though kind manner was evident in all meetings, conversations and invitations, and I'll mention some of them in a moment. Therefore, it is a special pleasure and honour for me to formulate a laudation for this outstanding researcher, Professor Malay Ghosh, on the occasion of his honorary doctorate from the University of Economics in Katowice.

Professor Ghosh is without any doubt a world-leading researcher in statistics. His contributions to statistics cover a wide range of topics from theoretical findings to very important applications in many different areas. His main areas of interest are Bayesian and empirical Bayesian methodology, resampling methods, and hierarchical modelling, as well as sampling and small area estimation. His amazing 59 page curriculum vitae speaks for itself and you can easily find many other topics that have stimulated major interest.

After completing his studies in Calcutta, Professor Ghosh subsequently took the next steps of his career at UNC in Chapel Hill, the Indian Statistical Institute in Calcutta, and Iowa State University in Ames, until he became full professor in Calcutta. After another period at Iowa State, he finally arrived at the department of statistics at the University of Florida at Gainesville in 1982. Though he completed several prominent visiting professorships far off, he never left Florida. Since 1998, he has been Distinguished Professor of Statistics at the University of Florida, Gainesville, and currently the only one at this institution.

Professor Ghosh has served in almost 20 different roles as an editor and an associate editor for international journals. Amongst these journals are highly prestigious publications as the "Journal of the American Statistical Association" and the "Annals of Statistics".

Relating to his own research, Professor Ghosh can be proud of over 300 peer-reviewed papers with nine more in press, and by the time of writing this laudatory speech surely even more. His research work is published in the most prestigious

statistics journals including the "Annals of Statistics", the "Journal of the American Statistical Association", "Biometrika", the "Journal of the Royal Statistical Society B", as well as many other well-known journals.

Since 1976, Professor Ghosh has supervised and co-supervised more than 60 PhD students. The variety of topics of the theses is as impressive as the range of his own research, covering theoretical and practical findings in so many directions. Amongst his students are so famous researchers who have also become professors, such as Nitis Mukhopadyay, Partha Lahiri, and Gauri Datta. It is not solely his own list of PhD students who have benefitted from his rich set of ideas but also a long list of guests. And whenever I speak to one of his students, I hear only very warm words about his research and mental support. Possibly sometimes so impressive that students face a challenge of making his ideas into reality as quickly as Professor Ghosh develops them. This unbelievable intensity in research ideas and promotion of students may sometimes provide a special challenge for students. However, he has always associated this with particularly positive support to foster a best possible development of his students and guests.

As you might expect that this outstanding re-searcher would likely be less committed to committee services, you will be surprised to see his very long list of contributions to all faculties where he was and is present. He has been active in many directions in the university system, with special emphasis on postgraduate education. In addition to these university committees, he showed major support on many different occasions outside the university. It is unbelievable, how he could manage all these duties besides his amazing research record. These activities encompass various roles in societies such as the Institute for Mathematical Statistics and the American Statistical Association, for which he served in many different positions. Additionally, he was well respected and often invited to participate in programme committees for international conferences. It is self-explanatory that having him on a board was already a major point of attraction for any conference. Especially for the series of Small Area Statistics Conferences, he was always appointed as member of the advisory committee.

As Professor Ghosh was certainly often asked to organise invited sessions, he himself has provided an impressively long list of over 120 invited papers covering so many regions all over the world. Additionally, he has provided 130 invited and special invited lectures in such different areas of his interest. This incredible reputation is almost certainly the reason why he shows comparatively few contributed papers in his curriculum vitae.

Besides his amazing list of research contributions, he was the principal investigator in many highly recognised research projects and grants, of which many stemmed from the National Research Foundation. This amazing list was enriched by further collaborations with important organisations such the U.S. Census Bureau, the Bureau

of Labor Statistics, and the National Agriculture Statistics Services. This again proves how the many theoretical findings of Professor Ghosh serve as an important contribution to applications in many fields of statistics.

Let me focus on some of Professor Ghosh's outstanding contributions, though this is more a personal view. His contributions to small area statistics are surely pioneering.

In 1987, he published, together with Professor Lahiri, a paper on Robust empirical Bayes estimation of variances from stratified samples in the "Journal of the American Statistical Association". The focus of the paper is a simultaneous estimation of means in multiple finite populations. This is essentially useful and applicable for estimating annual incomes or unemployment rates, which are to be estimated in many areas simultaneously. The research has influenced a wide spectrum of research from empirical Bayes prediction to frequentist methodology of small area estimation, which can be drawn from far more than 100 citations in these research areas.

His article on Small area estimation: An appraisal, together with J. N. K. Rao, has been published in "Statistical Science" in 1994 and received major attention being cited over 1,000 times.

Professor Ghosh's research covers so many different areas: he has also provided important contributions to economics. Several papers and presentations focused on specific aspects of income and its parameters. The methods in use cover Bayesian cross-sectional and intertemporal approaches as well as regional aspects using small area techniques. Especially research on regional incomes provides an important topical theme, which plays an important role in applications and even in policy support. Reliable figures more and more play an utmost important role for a modern democracy and Professor Ghosh's research on benchmarking in small area statistics enables the provision of the necessary basis. In light of this economics related research, Professor Ghosh visited Katowice several times.

Another important area of his research focuses on the American census which surely serves as an important source of economic data for the society. I don't need to remind you that his many contributions on a variety of regression methods surely serve as an outstanding basis for modern quantitative research. Finally, Professor Ghosh's research on finite population sampling and Bayesian inference for statistics, in general, and for econometrics provides excellent findings which are necessary input for modern economic research.

Professor Ghosh's eminence as a statistician has already been well recognised internationally. He is a fellow of the American Statistical Association, the Institute of Mathematical Statistics, and the International Society for Bayesian Statisticians. He is an elected member of the International Statistical Institute. He is a holder of the Jerzy Spława-Neyman Medal (2012), the Lifetime Achievement Award from the Indian Statistical Association (2017), and the Small Area Estimation Award (2019).

Without a shadow of doubt, Professor Ghosh is a distinguished authority in statistics. His contributions to statistics have not only been an example for junior researchers but for everyone working in this field. They have shown important methods providing paths for future research. In all ways, Professor Ghosh's a remarkably positive kind manner combined with his incredible productivity and creativity deserves to be awarded the doctorate honoris causa.

Finally, I recommend with strong emphasis this outstanding researcher, Professor Malay Ghosh, for the doctorate honoris causa from the University of Economics in Katowice.

PRESENTATION/LAUDATION

Professor Malay Ghosh the doctor honoris causa
of the University of Economics in Katowice

On May 14, 2021, Professor Malay Ghosh was officially awarded the honorary doctorate of the *University of Economics in Katowice*, Poland. This title was awarded by the Senate of this University to Professor Malay Ghosh in recognition of his outstanding achievements in the field of statistics and its applications, as well as his commitment to promoting science and international cooperation amongst scientific communities.

Professor Malay Ghosh was born on 15 April 1944 in Calcutta, India. He completed his undergraduate and graduate courses in statistics in 1962, at the age of 18, and in 1964, at 20, respectively from Calcutta University. As a student he was awarded several scholarships. He completed his undergraduate, as well as his graduate course as a First Class First and Gold Medallist. This enabled him to enrol on a PhD course at the University of North Carolina at Chapel Hill in the United States, where in 1969 he obtained his doctor's degree written under the supervision of Professor Pranab Sen.

From 1971 to 1978 he worked as an associate professor at the *Indian Statistical Institute*. For eight years, beginning in 1974, he was employed as a professor at the *Department of* Statistics of *Iowa State University*. Since 1982 up to now Professor Malay Ghosh has held a position at the *Department of Statistics of the University of Florida* – first as a professor, and next as a Distinguished Professor. Besides, as a visiting professor, he has lectured at 7 universities outside the US on a variety of topics, including advanced statistical inference, decision theory, Bayesian theory, multidimensional analysis, sequential analysis as well as reliability theory.

Professor Malay Ghosh's research works significantly enrich the theoretical and applied statistics in many fields and are widely cited. Let us list some of them, starting with broad area of sequential estimation, which is directly related to the criterion of maximizing economic efficiency with limited outlays, among other things. Professor Ghosh, together with N. Mukhopadhyay and P.K. Sen, wrote a monograph entitled *Sequential estimation*, which was published by John Wiley & Sons, Inc. in 1997. Small area estimation is another field which has been of special interest to Professor Ghosh

up to the present day. A major work which falls under this category is the monograph published by Chapman & Hall in 1997, entitled *Bayesian Methods for finite population sampling*, which Professor M. Ghosh co-authored with G. Mideeden. The book deals with the latest modes of applying Bayesian inference to the survey sampling method widely used by statistical offices organizations conducting consumer market research and public opinion polls. Professor Ghosh's works on small area estimation use Bayesian estimation methods along with other methods of statistical inference, which are of great significance in economic research. Let us stress the fact that Professor Ghosh's merits were honoured by awarding him the *Small Area Estimation Award*. The actual range of Professor Ghosh's scientific output exceeds these areas of research mentioned here because his articles also concern issues like reliability theory and non-parametric inference, which is a valued tool in the statistical analysis of economic data. Besides, Professor Ghosh devoted many of his works to the problem of estimator admissibility. Professor Ghosh's scientific output, including his monographs, constitutes theoretical and methodological foundations for innovative approaches to statistical analyses in several disciplines - economics, medicine, technology and agriculture. This demonstrates the unusual versatility of his scientific output as an author or co-author of over 310 publications. His works are widely cited, most of them in journals from the prestigious list of the *Journal of the Citation Report*.

Many of Professor Ghosh's publications resulted from his work on at least 30 research grants. Those projects were financed by such prominent institutions as *The National Science Foundation, The United State Army Research Office, The US Census Bureau, The National Institute of Health as well as The United State Center for Disease Control*. Professor Ghosh has participated in 150 prestigious international conferences. He has generously shared his knowledge also through editorial activities - as an editor or associate editor of 20 scientific journals, world-wide, among them as a member of the Editorial Boards of the following journals: *Sequential Analysis, Communications in Statistics, Statistics in Transition*.

Professor Ghosh's merits and academic prestige are recognized on the statisticians' international forum, which is corroborated by his involvement in international scientific bodies. For example, he was an elected fellow of *The American Statistical Association, The International Society for Bayesian Analysis, The International Statistical Institute and The Institute of Mathematical Statistics*. The statisticians' community expressed their admiration for Professor Ghosh's scientific achievements organizing in his honour a conference at the *University of Maryland*. It is worth emphasizing that Professor Ghosh's extraordinary personality, his knowledge and generosity, as well as his scientific ideas along with creative problem-solving approach, have inspired over sixty people to write their PhD dissertations under his supervision.

It is also noteworthy that in his scientific itinerary across the world Professor Ghosh have not forgotten about his native Indian roots. The *International Indian Statistical Association* honoured his merits in this field by granting him the Lifetime Achievement Award in 2017.

The aforementioned examples of Professor Ghosh's scientific activity give us a sense of pride in the fact that it was him that on many occasions participated in scientific conferences organized in Poland. Four times he was an invited lecturer at the conference on "Survey sampling in economic and social research" organized by the *Department of Statistics, Econometrics and Mathematics of the University of Economics in Katowice, in collaboration with the Polish Statistical Association*. In 2012 he was hosted as a special invited speaker at the First *Congress of Polish Statistics* held on the occasion of the one hundredth anniversary of the foundation of the *Polish Statistical Association*. At the same congress, Professor Ghosh was honoured with the Jerzy Neyman Medal. In 2014 as an invited speaker, Professor Gosh read a paper at the conference *Small Area Sampling* organized by *University of Economics in Poznań and International Association for Survey Statisticians*. During that Conference, he was invited to join the *Editorial Board of Statistics in Transition*.

Professor Ghosh's academic profile needs to be supplemented with his uncommon personality traits. He has always been an extremely outgoing, generously knowledge-sharing person and kind person, ready to offer good advice, and happy about an opportunity to do it. His commitment to science is simply part and parcel of his life. And can serve as a role model for us all, not only for statisticians. Professor Malay Ghosh is a prominent member of the world league of scientists. In view of this, the Senate of *University of Economics in Katowice* was fully justified in initiating the proceedings to award the honorary doctor's degree of this *University* to Professor Malay Ghosh in recognition of his outstanding achievements in the field of statistics and its applications, as well as his commitment to promoting science and international scientific cooperation.

Prof. Janusz Leszek Wywiał, Ph.D.

INVITED SPEAKERS – CONGRATULATIONS

I want to congratulate Professor Malay Ghosh, also known as my dad, on his receiving an honorary doctorate from the University of Economics in Katowice. He and my mom have visited Katowice on numerous occasions and have very much enjoyed the hospitality of the university as well as the people of the town.

In addition to being his son, I am also his colleague in the statistical community. During his 50+ years as a statistician, there are five things I have found remarkable about my father professionally. First, his research has adapted and continuously evolved over time. From nonparametric methods in sequential analysis to Bayes/Empirical Bayes methods in small-area estimation to his current interest in sparsity in high-dimensional Bayesian inference, his mind has never stopped moving onto the next topic. Second, his writing ability remains unparalleled. A senior colleague once told me that my father is only person he knew who could write a perfect first draft of a research paper. Third, his passion about statistics has remained sky-high over more than five decades. He has the same hunger for research now that I see in many of my junior colleagues who are just beginning their careers. Fourth, his willingness to take time to help any of his younger colleagues is marvelous. The one thing I have observed in my own career is the most valuable asset available to an academic is time, and my father have always sought to help those who ask for it. Fifth, his focus on a single research/academic topic at any point in time remains unparalleled.

Dad, congratulations on this supremely well-deserved honor.

Mr. D. Ghosh, Jr

Ministry of Economic Development,
Labour and Technology
———
Robert Tomanek
Undersecretary of State

May 14, 2021

Professor Malay Ghosh
Doctor Honoris Causa
University of Economics in Katowice

Dear Professor Ghosh,

I am both honored and privileged to extend to you my most sincere congratulations on being awarded the honorary doctorate degree from the University of Economics in Katowice.

Awarding the highest academic degree, the university community has expressed its deepest appreciation of your achievements as a distinguished scholar.

Honored to join all those who are extending to you their warm thoughts on this solemn occasion, I wish you many years of good health and continued success in your career and personal life.

I remain sincerely yours,

Robert Tomanek, PhD

Magnificence Rector, Honorable Members of the Senate,
Dear professor Malay Ghosh,
Members of the academic community,
Distinguished colleagues and friends, from other universities,
Ladies and Gentlemen.

Dear Professor Malay Ghosh. Congratulations!

It is a great privilege and honor to congratulate and thank you, even in just a few words.

You are widely known as an exceptional scientist and statistician, particularly in Bayesian inference and small area estimation. Your extensive academic and professional career has so many outstanding aspects that, summarizing it goes beyond our capacity, especially the capacity of such a short intervention. However, I want to stress one thing: it is essential to keep in mind that pushing the boundaries of knowledge requires outstanding effort and a clear vocation of service. Today, celebrating your achievements, we celebrate this unique and precious composition of virtues that you represent.

We have experienced your uniqueness and numerous invaluable achievements for many years, and by "we," I mean both the Polish and international community of academic and official statisticians. The work of statistical offices has been influenced extensively by your research. I am convinced that this is a source of satisfaction for you, and it should be. Official statistics serve people, societies, so it can be said that a great deal of the progress enjoyed by modern society is achieved thanks to your work.

I want to mention the second Congress of Polish Statistics in 2012, during which you were awarded the *Jerzy Spława-Neyman Medal*. Taking this opportunity, on behalf of the entire community of Polish statisticians, I would like to invite you, Professor, to participate and deliver a keynote speech at the following third congress of Polish statistics. The congress, organized on the 110th anniversary of the Polish Statistical Society, will be held in April 2022 in Krakow. This occasion will also be unique because the International Association for Official Statistics conference will be held simultaneously in the same conference center on the same days.

Dear Professor, congratulations again!

Thank you very much for your attention.

Dominik A. Rozkrut,
Statistics Poland, President, Chief Statistician

Madam Rector, honourable doctor honoris causa, ladies and gentlemen, with great pleasure I accepted the information about the initiative of the University of Economics in Katowice about honouring Professor Malay Ghosh with the title of doctor honoris causa of the University of Economics in Katowice.

Honourable Professor Malay Ghosh!

This day is a huge celebration of Polish Statistics. You are an outstanding statistician, whose works are a determinant for conducting statistical research.

Especially valid are lectures on various topics, including advanced statistical inference, decision theory, Bayesian theory, multidimensional analysis, sequential analysis, and reliability theory. The achievements are impressive to over 310 scientific publications, including articles, monographs and chapters in books on statistical methodology and its applications, and his works are widely cited in scientific circles. The relationship between honourable Professor Malay Ghosh and Polish statistics from 2008, likewise the participation in many conferences in Poland, deserve special recognition. At the First Congress of Polish Statistics held on the occasion of the one-hundredth anniversary of the foundation of the Polish Statistical Association, Professor Malay Ghosh was honoured with the Jerzy Spława-Neyman Medal.

Your recognition for international statistics can be best described by the words of John Paul II: „Man is great not by what he has, but by what he is; not by what he has, but by what he shares with others".

As the head of the Polish Statistical Society, I would like to congratulate one more time on the granted award to Professor Malay Ghosh. Also, I would like to thank the University of Economics in Katowice for carrying out this event.

Waldemar Tarczyński
Rector, University of Szczecin

Dear Magnificence Rector!
Dear Members of Senate of University of Economics in Katowice!
Dear Professor Ghosh!
Distinguished Guests, Ladies and Gentlemen!

It is my great pleasure and honour to present this address on behalf of The Committee on Statistics and Econometrics of the Polish Academy of Sciences.

Dear Professor Ghosh!

Please accept my congratulations on the occasion of awarding you doctor honoris causa, which in Polish academic tradition is regarded as the highest academic honour. University of Economics in Katowice pays tribute to your great achievements in the area of statistics.

Polish statisticians regard your scientific achievements as very valuable for both theoreticians and practitioners. Your theoretical work in the area of sequential estimation and nonparametric inference created new directions in mainstream statistical research. We value very highly your scientific contribution in the area of competing risks, which is a great example of development arising from reliability theory. Finally, I mention your papers on small area estimation, which are of particular interest for practitioners representing public statistics. Your scientific contribution is great added value to the theory and practice of statistics.

Dear Professor Ghosh, I wish you good health and continuation of your scientific achievements.

Dear Professors of University of Economics in Katowice!

I congratulate you to have Prof. Ghosh as a member of your scientific community. I wish all the best in extending the scientific cooperation with your Honorary Doctor.

Krzysztof Jajuga

Magnificence Mrs. Rector, High Senate, Dear Professor, Ladies and Gentlemen

Dear Malay, **Congratulations! And thanks.**

Along with congratulations on recognizing your extraordinary achievements and professional status world-wide – as a Master and teacher of teachers of statistics, and a person whose dedication to science made him uniquely deserving of the great honour you have received today – I would like to take this opportunity to thank you on behalf of the Statistics in Transition, an international journal of the Polish Statistical Association and Statistics Poland: For all the help and contribution you have made to our journal – I say ours, because you are part of it also as a long-term member of the Editorial Board and as an author and reviewer, of what we are especially proud and grateful.

This is, of course, only a part of your multi-threaded contribution to Polish statistics, and statistics in general, that was already recognized with the Jerzy Neyman medal awarded to you during the Polish Statistical Association 100th anniversary Congress in Poznań, 2012. Recent example is Your spectacular contribution to the Special Issue of SiT – organized by Partha Lahiri, published in August 2020, and commented by such leading experts in the field as Jon Rao, Danny Pfeffermann and others, which confirms the constancy of your presence in Polish statistics.

To conclude this statement, I would like to express my appreciation to the University of Economics in Katowice and its Senate, with which I turn to Her Magnificence Rector, for arranging this ceremony to honour our mutual friend, Distinguished Professor Malay Ghosh.

Thank you Malay in advance for your continued collaboration with us – we wish you further successes in excellent conditions and health.

Thank you all for your attention.

Włodzimierz Okrasa
Editor-in-Chief, Statistics in Transition new series
Cardinal Stefan Wyszynski University in Warsaw,
Statistics Poland

For May 2021, University of Economics, Katowice, Poland

Professor Malay Ghosh was born in Calcutta, today in the state of West Bengal in India, and studied at the University of Calcutta for an undergraduate and an MA degree, before continuing his studies in the United States at the University of North Carolina. His PhD. advisor was another outstanding statistician who also hails from Calcutta, Pranab Kumar Sen. After graduating, Professor Ghosh held appointments at the Indian Statistical Institute and then back in the U.S.A. at Iowa State University, before joining the Department of Statistics at the University of Florida in 1982.

Within the alotted time I could not possibly go through all the honours that he has received during his distinguished career at the University of Florida, nor list all the fields to which he has made profound contributions. So, here is a selection. Professor Ghosh had a conference held in his honour at College Park, Maryland, in 2014; he received from the American Statistical Association the Samuel Wilks Memorial Award in 2020; he is a coauthor of two monographs; he supervised over 40 PhD students; and has been a sought-out consultant by various organisations, including the U.S. Decennial Census and the Office for National Statistics in the United Kingdom. You will come across his name if you study or work in nonparametric statistics, Bayesian modelling, sequential analysis, small-area estimation, sampling methods, and a clutch of other important topics.

It is my great honour to congratulate Professor Ghosh on the award of an honorary doctorate from the University of Economics in Katowice. I would also like to thank the University for such an appropriate choice for an honorand.

Nicholas T. Longford
Imperial College London, UK

CONGRATULATORY LETTERS

April 20, 2021

Graham Kalton
SILVER Spring ,
MD 20 906
gkalton@gmail.com

Dear Malay,

I was delighted to learn that you are being awarded the degree of doctor honoris causa by the University of Economics in Katowice. I believe that your very extensive contributions to statistical theory and methods, together with your contributions to Polish statistics, fully justify this great honor.

My wholehearted congratulations to you. With warmest regards.

Sincerely,

Graham Kalton

<div align="right">Ystad, Sweden, 2021-04-23</div>

Carl-Erik Särndal
Professor Emeritus,

Professor Malay Ghosh,

Dear professor Ghosh,

 My heartfelt congratulations to you, at the occasion of the honor bestowed on you by the University of Economics in Katowice.
With warm regards,

Carl-Erik Särndal
Statistics Sweden

Professor Malay Ghosh
Department of Statistics
University of Florida

Dear Professor Ghosh,

Please accept our sincerest congratulation on the occasion of your being awarded the title of Doctor Honoris Causa from University of Economics in Katowice. Yours accomplishments in the field of sequential analysis, Bayesian statistics and small-area estimation have made a significance contribution to the development of statistics.

Although your participation in the growth of the Polish statistical thought has already been awarded in the form of the Jerzy Spława - Neyman Medal during the Congress of Polish Statistics in 2012, it still merit our warmest gratitude.

We wish you every success in your professional activities which, we believe, will result in many memorable achievements. May you enjoy happiness and contentment in your personal life.

Yours very sincerely
Professor Czesław Domański

Poznań, 13 May 2021

prof. Malay Ghosh, PhD
Distinguished Professor
University of Florida

Dear Professor Malay Ghosh,

I am deeply honoured to congratulate you on behalf of the Senate of the Poznań University of Economics and Business and the whole academic community for the highest academic degree, the title of doctor honoris causa which you were awarded by the **University of Economics in Katowice**.

Let me express my greatest respect for you as the most prominent scientist in the field of statistical theory, small area estimation, and its applications in, inter alia, economics and social sciences, eminent mentor and teacher, and a very caring and insightful man.

We all appreciate your contributions to highlighting and solving important theoretical and practical research problems. Thank you for the opportunity to learn from your scientific achievements, for the knowledge provided and for openness to the problems of science, people and the world. Thank you for your friendship and kindness.

Wishing you happiness in your private life and hoping you will continue to make further contributions to statistical knowledge

I remain yours sincerely,

prof. dr hab. Elżbieta Gołata
Vice-Rector for Research and International Relations

Katowice, May 5, 2021
RIG/2/05/2021/KO

Mr. Professor Malay Ghosh
Malay Ghosh

Dear Mr. Professor,

On behalf of the representatives of the Chamber of Commerce and Industry in Katowice and my own, I would like to congratulate you receiving the honorable title of Doctor Honoris Causa of the University of Economics in Katowice.

The dignity of Doctor Honoris Causa, awarded by the University of Economics in Katowice, is a unique and extremely important distinction, including emphasizing the rich scientific achievements of the Professor and recognition of the authority that was born out of real merits, from knowledge and values brought to the world of science, and thus also to the space of public life.

The Professor's huge knowledge and experience, which is a consequence of many years of work and scientific activity, contributed to the creation of interesting publications and many valuable achievements in the field of the theory of statistics and its applications, both in economic sciences and social, which inspires our admiration and respect.

We wish that the honorable title of Doctor Honoris Causa of the University of Economics in Katowice will be for you not only a token of appreciation, but also that it will be a great tribute to your years of work and great achievements, that in the future it will forge the next steps of cooperation with the University which result with a many interesting projects, bringing contentment and satisfaction.

We also sincerely congratulate the University of Economics in Katowice that such a wonderful and prominent scientist has become the Ambassador of University in the world, expressing the conviction that the results of joint activities will certainly contribute to further multidimensional development.

With kind regards,

Tomasz Zjawiony
President of the Chamber of Commerce and Industry in Katowice

Dear Professor Ghosh,

On behalf of all academic community of Gdynia Maritime University, please accept our sincerest congratulations on the occasion of awarding you the Doctor Honoris Causa of the University of Economics in Katowice.

Your accomplishments in the field of statistical theory have made a significant contribution to the development of mathematics and economic sciences.

Dear Professor Ghosh, you are a person with outstanding academic and scientific achievements and there is no doubt that you are a world-leading researcher in statistics.

I wish you a lot of success in your professional career and also happiness and satisfaction in your private life.

With kind regards,

Prof. Adam Weintrit
Rector
Gdynia Maritime University

About the Authors

Ashutosh is an Assistant Professor (Guest) at the Department of Statistics, Faculty of Science & Technology, Mahatma Gandhi Kashi Vidyapith, Varanasi, India. He defended his PhD thesis "Estimation of Small Area Parameter under Different Situations" at Banaras Hindu University. He has published 10 manuscripts in reputed international journals and 2 book chapters based on small area estimation (SAE) using auxiliary variable(s). Recently, he has been doing calibration estimation of domain parameter using the distribution functions such as exponential, power, gamma vie simulation. Currently, he is a reviewer of Statistica Journal.

Audu Ahmed is a Senior Lecturer at the Department of Mathematics, Faculty of Science, Usmanu Danfodiyo University Sokoto, Nigeria. His research interests are sample survey, multivariate statistical analysis, statistical inference and data analysis in particular. He is an active member of many scientific professional bodies like Professional Statisticians Society of Nigeria (PSSN), Nigeria Statistical Association (NSA), Mathematical Association of Nigeria (MAN) and Science Association of Nigeria (SAN).

Ayhan, H. Öztaş is a Professor Emeritus at the Department of Statistics in Middle East Technical University, Ankara, Turkey. His research interests are survey sampling techniques, survey methodology research and web survey methodology. Professor Ayhan has published many research papers in international/national journals and conferences. He has also published several books and monographs. Professor Ayhan is an elected member of the International Statistical Institute, International Association of Survey Statisticians, and International Association for Official Statistics.

Batmaz, İnci is a Professor at the Department of Statistics in Middle East Technical University, Ankara, Turkey. Her research interests include statistical design and analysis of experiments, response surface methodology, computational statistics, statistical modelling and simulation, machine and statistical learning, data science and analytics, and big data mining. She has published various research papers on these topics in inter/national journals and conferences. She is a member of International Association for Statistical Computing (IASC) of International Statistics Institute, ERCIM Working Group of Computational and Methodological Statistics and Continuous Optimization Working Group of EURO.

Bąk Tomasz is an Associate of the Department of Statistics, Econometrics and Mathematics, Faculty of Management, University of Economics in Katowice. He defended his doctorate 'Spatial Sampling Methods in Economic Studies' in 2018. For the last few years he has been running a company delivering statistical models and analysis. His main areas of interest include: spatial statistics, adaptive sampling and statistical methods in security. He is an author of several research papers in international journals.

Beręsewicz Maciej is an Assistant Professor at the Department of Statistics, Poznań University of Economics and Business. Simultaneously he holds the position of a Senior Specialist at Centre for Small Area Estimation at the regional Statistical Office in Poznań. His main area of interest is survey methodology, in particular non-probability samples, missing data problems and small area estimation.

Dehnel Grażyna is an Associate Professor at the Department of Statistics, Poznań University of Economics and Business. Her main research domain is small area estimation, survey sampling, short-term and structural business statistics. She is also interested in outlier robust regression applied on business data, business demography and data integration.

Denkowska Anna is an Assistant Professor in the Mathematics Department of the Cracow University of Economics (Poland). MSc in mathematics (Jagiellonian University), PhD in mathematics (Jagiellonian University). Her research interests include functional analysis, statistics, risk modelling and network theory.

Fahmi, Fidan Mahmut is a Graduate Student of the Middle East Technical University, Ankara, Turkey. Fidan Fahmi is a statistical consultant with PhD in survey statistics, specializing in survey response error estimation. Dr. Fahmi is also working on alternative interviewer allocation designs for survey interviewing. She has worked on statistical analysis of climate change signals. She has recently worked on the use of center-based clustering methods for determining the climate zones of Turkey.

Gołata Elżbieta is a Full Professor of Economics at the Poznan University of Economics and Business. Her main areas of interest include: demographics, labour market, social statistics, statistics for small areas, multivariate statistical analysis and quality of statistical data. Professor Gołata has published over 100 research papers and three monographs. Professor Gołata is an active member of many scientific professional bodies. Currently, she holds the position of President of the Committee on Demographic Studies of the Polish Academy of Sciences.

Kedem Benjamin is the STAT Director and Professor in the Department of Mathematics, University of Maryland, College Park. His research interests include time

series analysis, spatial statistics, semiparametrics, data fusion, and generalized linear models. At present he is working on applications of augmented reality in the estimation of very small tail probabilities.

Khare Brij Behari is a Professor at the Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India. He has supervised 10 PhD students and 2 are pursuing. He has published over 110 manuscripts, 14 book chapters in the reputed national and international journals. His main contribution is the estimation of population parameter(s) using auxiliary variable(s) in the presence of non-response. He has been the Fellow of M. N.A.S.I., India. He has given mostly contribution in the interest of survey sampling, non-response, measurement error, class of estimators, small area estimation, etc. He is a reviewer of national and international publishers, including Taylor & Francis, Springer, CESER.

Rai Piyush Kant is a Full Professor of Statistics at Department of Statistics, Banaras Hindu University, Varanasi (B.H.U.), India. His research interests include: sample survey, small area estimation (SAE), demography, Bayesian modelling, biostatistics and bioinformatics. Prof. Piyush has published substantial quality research articles in many national and international journals of repute and is also an active reviewer of some of them. He has also published different books and book chapters through renowned national and international publishers, including Handbook of Statistics, Elsevier, Springer and IGI, Global. He is an active associate and executive member of many national and international academic statistical societies and professional bodies such as ISBA, Indian Bayesian Society, Society of Statistics and Computer Applications, Rajasthan Statistical Society

Sen Rijji currently works as an Assistant Professor at Behala College, Calcutta University, after having completed her PhD in Statistics from Institute of Science, Banaras Hindu University. Her research interests mainly include Bayesian inference, Bayesian computation, econometrics, reliability analysis and survival models.

Singh Rajesh is a Professor of Statistics at the Department of Statistics, Institute of Science, Banaras Hindu University. His area of research interest includes sampling theory and estimation techniques. He has published around 130 research papers in journals of national and international repute. He is also a referee and a member of various reputed national and international journals.

Tripathi Praveen Kumar currently works as an Assistant Professor at Banasthali Vidyapith, Rajasthan, India. He has awarded his PhD degree (2017) in Statistics from Banaras Hindu University. His research contribution is mainly in the area of Bayesian inference, time series analysis, econometrics and demographic data modelling.

Upadhyay S. K. is a Professor and former Head of the Department of Statistics, Banaras Hindu University. He is also the former coordinator of DST Centre for Interdisciplinary Mathematical Sciences at Banaras Hindu University, where he is involved as a core group researcher. He obtained his MSc (1980) and PhD (1984) degrees from Banaras Hindu University, and started his teaching career as lecturer in statistics in 1981. He received a number of distinctions in his career, and worked as EEC research bursary at Imperial College, London, during 1992-93. His research interests include preliminary test procedures, life testing and reliability, and Bayes inference. He has published several research papers, books, and other academic materials in his career.

Wanat Stanisław is an Associate Professor in the Mathematics Department at the Cracow University of Economics (Poland), MSc in mathematics (Jagiellonian University), PhD in economics (Cracow University of Economics), habilitation in economics (Cracow University of Economics). His research interests include statistics, actuarial methods, risk modelling and forecasting.

Wawrowski Łukasz is a Data Analyst at the Research Network Łukasiewicz - Institute of Innovative Technologies EMAG, former employee of the Department of Statistics, Poznań University of Economics and Business. His main areas of interest is small area estimation, social statistics and machine learning.

Xuze Zhang is a PhD candidate in Mathematical Statistics at Department of Mathematics, University of Maryland, College Park. He is under the guidance of Professor Benjamin Kedem and his research interest is semiparametric inference in Data Fusion.

Žmuk Berislav graduated with the major Accounting, post-graduated Statistical Methods for Economic Analysis and Forecasting and gained his PhD degree in Business Economics at Faculty of Economics and Business, University of Zagreb. Currently, he is an Assistant Professor at the Department of Statistics, Faculty of Economics and Business, University of Zagreb, where he teaches the following subjects: Statistics, Business Statistics, Business Forecasting and Introduction to Economic Statistics. His main research fields include applications of statistics in business and economy, survey methodology and statistical quality control. He is an active member of scientific societies. Currently, he is the President of the Croatian Statistical Association and the editor-in-chief at the Croatian Review of Economic, Business and Social Statistics journal.

GUIDELINES FOR AUTHORS

We will consider only original work for publication in the Journal, i.e. a submitted paper must not have been published before or be under consideration for publication elsewhere. Authors should consistently follow all specifications below when preparing their manuscripts.

Manuscript preparation and formatting

The Authors are asked to use *A Simple Manuscript Template (Word or LaTeX) for the Statistics in Transition Journal (published on our web page:* http://stat.gov.pl/en/sit-en/editorial-sit/).

- *Title and Author(s)*. The title should appear at the beginning of the paper, followed by each author's name, institutional affiliation and email address. Centre the title in **BOLD CAPITALS**. Centre the author(s)'s name(s). The authors' affiliation(s) and email address(es) should be given in a footnote.

- *Abstract.* After the authors' details, leave a blank line and centre the word **Abstract** (in bold), leave a blank line and include an abstract (i.e. a summary of the paper) of no more than 1,600 characters (including spaces). It is advisable to make the abstract informative, accurate, non-evaluative, and coherent, as most researchers read the abstract either in their search for the main result or as a basis for deciding whether or not to read the paper itself. The abstract should be self-contained, i.e. bibliographic citations and mathematical expressions should be avoided.

- *Key words*. After the abstract, Key words (in bold) should be followed by three to four key words or brief phrases, preferably other than used in the title of the paper**.**

- *Sectioning*. The paper should be divided into sections, and into subsections and smaller divisions as needed. Section titles should be in bold and left-justified, and numbered with **1.**, **2.**, **3.**, etc.

- *Figures and tables*. In general, use only tables or figures (charts, graphs) that are essential. Tables and figures should be included within the body of the paper, not at the end. Among other things, this style dictates that the title for a table is placed above the table, while the title for a figure is placed below the graph or chart. If you do use tables, charts or graphs, choose a format that is economical in space. If needed, modify charts and graphs so that they use colours and patterns that are contrasting or distinct enough to be discernible in shades of grey when printed without colour.

- *References.* Each listed reference item should be cited in the text, and each text citation should be listed in the References**.** Referencing should be formatted after the Harvard Chicago System – see http://www.libweb.anglia.ac.uk/referencing/harvard.htm. When creating the list of bibliographic items, list all items in alphabetical order. References in the text should be cited with authors' name and the year of publication. If part of a reference is cited, indicate this after the reference, e.g. (Novak, 2003, p.125).