// STATISTICS
IN
TRANSITION

new series

An International Journal of the Polish Statistical Association

CONTENTS

Volume 19, Number 3, September 2018

ISSN 1234-7655

Scientific Secretary
Marek Cierpiał-Wolan, e-mail: m.wolan@stat.gov.pl
Secretary
Patryk Barszcz, e-mail: P.Barszcz@stat.gov.pl, phone number 00 48 22 — 608 33 66
Technical Assistant
Rajmund Litkowiec, e-mail: r.litkowiec@stat.gov.pl

Address for correspondence

Statistics Poland, al. Niepodległości 208, 00-925 Warsaw, POLAND, Tel./fax:00 48 22 — 825 03 95

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 387−389

FROM THE EDITOR

The *sampling methods and estimation* section, which traditionally opens each of the journal's issue, contains just one article: ***Developing single-acceptance sampling plans based on a truncated lifetime test for an Ishita distribution*** by **Amjad D. Al-Nasser, Amer I. Al-Omari, Ahmed Bani-Mustafa, Khalifa Jaber**. The acceptance sampling plans encompass statistical procedures that are used for quality control and improvement in situation when it is not possible to test every item in a lot of materials. In this type of procedures, an important characteristic of the materials is their lifetime sampling distribution, which can vary from sample to sample. Authors propose to use a new lifetime distribution, known as an Ishita distribution, for developing a new single-acceptance sampling plan, which is developed under a condition wherein the mean lifetime test is truncated within a pre-specified time period. The parameters of the acceptance sampling plan including the operating characteristics function and the minimum sample size are obtained. The producer's risk in relation to the entire lot of materials is derived and illustrated by numerical examples. The consumers and producers are advised to adopt this plan in order to save time and minimize the production process cost.

A set of four papers in *research articles* starts with **Laxmi Kant Dwivedi's** article ***The role of breastfeeding vis-à-vis contraceptive use on birth spacing in India: a regional analysis***. Since birth spacing is one of the important aspects of reproductive health, it is studied from time to time in view of the epidemiological transition taking place worldwide. Using the third round of National Family Health Survey-3 data, the central hypothesis of this paper stresses the relative advantages of breastfeeding over other methods of contraception among non-sterilized women. A simulative approach of the Cox regression analysis for India and its regions has been used to verify such expectations. The results show that if women were not having amenorrhea period and had a high level of breastfeeding, the chance of not having next live birth was only two percent lower than those of women who were using spacing methods in India. An effort has also been made to apprise the policymakers of the interrelation between breastfeeding, postpartum amenorrhea, contraceptive use and birth spacing. Nonetheless, policymakers should promote programs that encourage both breastfeeding and contraceptive use. Breastfeeding has direct benefits for infant health in addition to its role in lengthening birth intervals beyond postpartum amenorrhea.

In the paper ***Dealing with heteroskedasticity within the modelling of the quality of life of older people***, **Katarzyna Jabłońska** demonstrates that using the estimation method of ordinary least squares leads to unreliable results in the case of heteroskedastic linear regression model. Alternative methods – including weighted least squares, division of the sample and heteroskedasticity-consistent covariance matrix estimators – can give estimators with better properties than ordinary least squares. The data come from the first wave of the COURAGE –

Poland study. The comparison of estimators and their practical application may provide interesting example of searching for the most appropriate estimation tool after detection of heteroscedasticity. Due to growing availability of techniques and access to more precise tools, it seems advisable to use HC-estimators in the occurrence of heteroskedasticity to verify the significance of model parameters, and to analyse them in depth.

Anna Majdzińska's paper, ***Spatial measures of development in evaluating the demographic potential of Polish counties*** presents a demographic potential-based typology of Polish counties. The typology was created using the spatial measures of development (available in the literature and proposed by the author) applied to Statistics Poland's data, The major information related to such features as age structure and changes in the natural movement of the population, and migrations in counties in the years 2005 and 2016.

The next paper, ***Another look at the stationarity of inflation rates in OECD countries: application of structural break-GARCH-based unit root test*** by **OlaOluwa Simon Yaya** discusses the unit root hypotheses of inflation rates in 21 OECD countries using the newly proposed GARCH-based unit root tests with structural break and trend specifications. The results show that classical tests over-accept unit roots in inflation rates, whereas these tests are not robust to heteroscedasticity. As it is observed from the pre-tests, those tests with structural break reject more null hypotheses of unit roots of most inflation series than those without structural breaks. By applying variants of GARCH-based unit root tests, which include those with structural breaks and time trend regression specifications, it was found that unit root tests without time trend give most rejections of the conventional unit roots. Batteries of unit root tests for discriminating between stationarity and nonstationarity of inflation rates are recommended, since in the case of over-differenced series, wrong policy decision will be made, particularly when inflation series is considered in a cointegrating relationship with other variables.

The *other articles* section includes a paper based on the presentation at the Multivariate Statistical Analysis Conference in Łódź (2016), ***Discriminant coordinates analysis in the case of multivariate repeated data*** by **Mirosław Krzyśko, Wojciech Łukaszonek,** and **Waldemar Wołyński.** The paper presents an innovative approach to analysis of multivariate repeated measures data using the classical discriminant coordinates. The proposed solution is based on the relationship between the discriminant coordinates and canonical variables. The quality of these new discriminant coordinates is examined on real data.

The next section, *research communicates and letters*, contains three papers. It starts with **Faizan Danish's** paper ***A mathematical programming approach for obtaining optimum strata boundaries using two auxiliary variables under proportional allocation.*** The paper commences with an observation that optimum stratification – the method of choosing the best boundaries that make the strata internally homogenous – is often attempted when a study variable is itself a stratification variable. However, in many practical situations fetching information regarding the study variable is either difficult or sometimes not available. Using auxiliary information many authors are redefining the problem as the problem of optimum strata width, and developed a solution procedure using dynamic programming technique. In this paper, under proportional allocation optimum stratification boundaries are determined for the study variable using two auxiliary variables as the basis of stratification with uniform, right-triangular, exponential and lognormal frequency distribution by formulating the problems which are executed by using dynamic programming. Empirical studies are presented to illustrate the computation details of the solution procedure and its comparison with the existing literature. The empirical studies suggest that the proposed method is more preferable than the existing methods.

In the next paper, ***Comparison of diabetic nephropathy onset time of two groups with left truncated and right censored data***, **Alka Sabharwal** and **Gurprit Grover** present a comparison of the nephropathy onset time of type-2 diabetic patients, grouped on the basis of gender and age at the time of diabetes diagnosis. Diabetic Nephropathy (DN) onset time is assumed to follow Weibull distribution with fixed left truncation. The likelihood ratio test is applied on uncensored cases and Thoman and Bain two sample tests are applied with generated left truncated Weibull distributions. To avoid the model validity issues for left truncated and right censored data (LTRC), the nonparametric approach, suggested by Kaplan and Meier, is used to compare the survival function of two groups over different time periods. Another method based on median survival time of the pooled group is applied to compare the survival function of two groups with LTRC data. The major advantage of developing methods for comparing the nephropathy onset times of DM patients is that the expected DN onset time of new DM patients can be predicted depending on the patient group.

Milena Bieniek's paper, ***Channel performance under Vendor Managed Consignment Inventory Contract with additive stochastic demand*** is devoted to the so-called consignment as the shifting of the inventory ownership to a supplier in virtual market. In this form of business arrangement the supplier places goods at a retailer's location without receiving payment, until the goods are sold. The author considers a single period supply chain model, where the supplier contracts with the retailer with some probability of return. Market demand is additive, linearly price–dependent and uncertain. Focus is on vendor managed consignment inventory (VMCI) channel, in which the supplier decides the consignment price and his service level and the retailer chooses the retail price. Also, a channel performance under VMCI setting is studied by analysing how the model parameters influence decision quantities, channel profit and risk function. The obtained results are illustrated by a numerical example.

Włodzimierz Okrasa
Editor

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 391

SUBMISSION INFORMATION FOR AUTHORS

***Statistics in Transition new series (SiT*)** is an international journal published jointly by the Polish Statistical Association (PTS) and the Central Statistical Office of 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:
P.Barszcz@stat.gov.pl.,
GUS / Central Statistical Office
Al. Niepodległości 208, R. 287, 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.

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 393–406, DOI 10.21307/stattrans-2018-022

DEVELOPING SINGLE-ACCEPTANCE SAMPLING PLANS BASED ON A TRUNCATED LIFETIME TEST FOR AN ISHITA DISTRIBUTION

Amjad D. Al-Nasser[1], **Amer I. Al-Omari**[2],
Ahmed Bani-Mustafa[3], **Khalifa Jaber**[4]

ABSTRACT

Acceptance sampling plans are statistical procedures that are used for quality control and improvement in cases where it is not possible to test every item in a lot of materials. The outcome of this test determines whether the entire lot is accepted or rejected based on a random sample. In this procedure, an important characteristic of the materials is their lifetime sampling distribution, and this can vary from sample to sample. In this article, a new lifetime distribution, known as an Ishita distribution, is considered for developing a new single-acceptance sampling plan. The new acceptance sampling plan is developed under a condition wherein the mean lifetime test is truncated within a pre-specified time period. Based on this condition, the parameters of the acceptance sampling plan including the operating characteristics function and the minimum sample size are obtained. The producer's risk in relation to the entire lot of materials is derived and illustrated by numerical examples.

Key words: acceptance sampling plans, Ishita distribution, producer's risk, consumer's risk, operating characteristics, truncated lifetime-distribution test.

1. Introduction

Single-acceptance sampling plans (SASP) are usually used to determine whether to accept or reject a lot of materials, while taking into consideration that the important characteristic of the materials is its potential lifetime distribution. Moreover, the most important stage is deciding which unit in a given lot should be selected for testing procedure. From a statistician's point of view, this decision involves two types of errors: a Type I error (α), which reflects the probability of rejecting a good lot of a materials (producer's risk); and a Type II error (β^{*}), which occurs when there is a probability of accepting a bad lot (consumer's risk). Several studies consider applying different procedures to SASPs (Al-Omari

[1] Department of Statistics, Science Faculty, Yarmouk University, Irbid, Jordan. E-mail: amjadyu@yahoo.com.
[2] Al al-Bayt University, Mafraq, Jordan. E-mail: alomari_amer@yahoo.com.
[3] College of Engineering, Australian College of Kuwait (ACK), Kuwait. E-mail: a.mustafa@ack.edu.kw.
[4] Al Falah University, Dubai, UAE. E-mail: COBA@afu.ac.ae.

(2015); Al-Omari et al., 2016a, 2016b; Al-Nasser and Al-Omari, 2013 and Al-Omari, 2014). In some plans, a product's quality characteristics are measured on a numerical scale (variable SASP); in other plans, they are measured based on the number of failures (attribute SASP). Kao (1966) provides a detailed comparison between the attribute and variable SASP.

In this article, we aim to develop a new SASP to test a product's attributes under the condition that the experiment is terminated at a pre-assigned time (say *t*). Here, a binomial experiment is carried out in which a random sample size (say *n*) is drawn from a large lot and tested for its potential lifetime distribution. In this type of test, if less than a specified accepted number of units (say *c*) fail, then the consumer accepts this lot; otherwise, they reject the lot. In previous studies the test performed by assuming the lifetime distribution is either gamma, Generalized Exponential, Log Normal, Weibull, or Pareto, etc. (see, for example, Al-Nasser and Gogah, 2017; Gogah and Al-Nasser, 2018; Balakrishnan et al., 2007; and Aslam et al., 2011.) In this article, we propose considering a new lifetime distribution known as the Ishita distribution (Shanker and Shukla, 2017; Shukla and Shanker, 2018).

This paper is organized as follows. Section 2 introduces the Ishita distribution and discusses some of its properties. Section 3 includes the suggested acceptance sampling plans for the truncated Ishita distribution; here, the necessary tables are also provided. The conclusions are presented in section 4.

2. Ishita distribution

In many quality control applications, lifetime distributions, such as those mentioned in the previous section, and many more, play primary roles in analysing and understanding data. These distributions are generally used to obtain the minimum sample size in an acceptance sampling context and other sampling plan parameters, such as the average run length (ARL). Shanker and Shukla (2017) suggest a new lifetime distribution known as the "Ishita distribution", which could also be used as an alternative to the classical lifetime distribution, when obtaining the parameters of any acceptance sampling plan. The one-parameter Ishita probability density function (PDF) is defined as follows:

$$f(x)=\frac{\beta^3}{\beta^3+2}\left(\beta+x^2\right)e^{-\beta x};,\ x>0,\ \beta>0, \tag{1}$$

Figure 1 shows the plot of the pdf with the different parameter values.

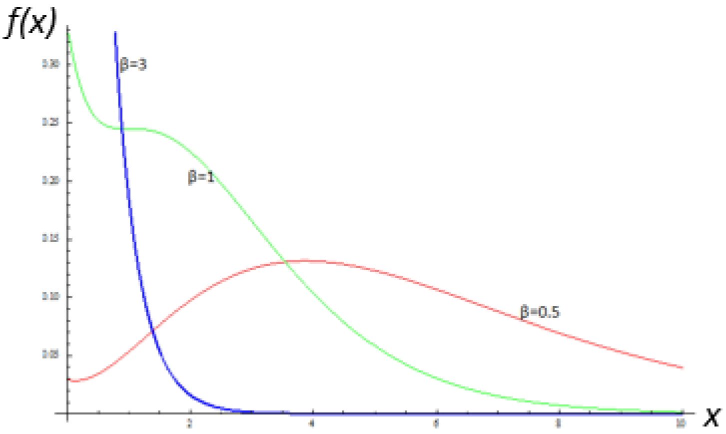


Figure 1. Ishita probability density function with β = 0.5,1,3

The cumulative distribution function (CDF) is defined as follows:

$$F(x) = 1 - \left(1 + \frac{\beta x(\beta x + 2)}{\beta^3 + 2}\right) e^{-\beta x}; \quad x > 0, \beta > 0. \tag{2}$$

Figure 2 provides an illustration of the Ishita CDF.

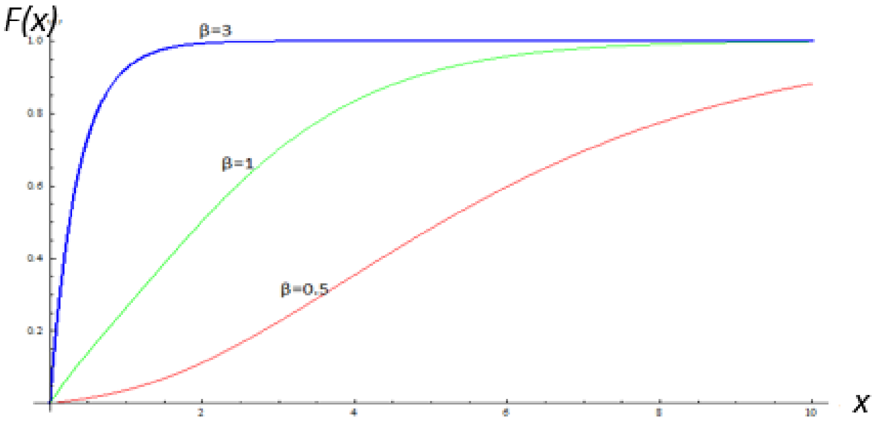


Figure 2. Ishita cumulative distribution function when β =0.5,1,3

The *r*th moment of the origin of the Ishita distribution is as follows:

$$E\left(X^r\right)=\frac{r!\left[\beta^3+(r+1)(r+2)\right]}{\beta^r(\beta^3+2)};\ r=1,2,3,....$$

For our acceptance sampling plan, based on this distribution, we consider the distribution when parameter ***β = 3***. Therefore, the moment behaviour that is based on this value can be visualized in Figure 3.

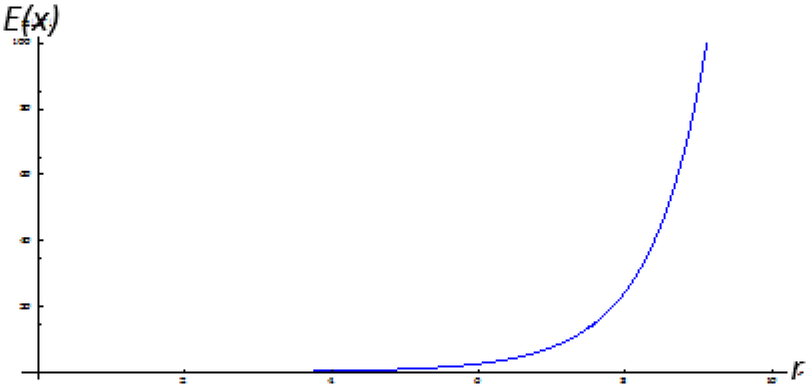


Figure 3. The rth moment value of the Ishita distribution when β = 3

In general, it can be noted that the distribution mean is

$$E\left(X\right)=\frac{\beta^3+6}{\beta\left(\beta^3+2\right)},$$

Then, the maximum likelihood estimator (MLE) $\hat{\beta}$ of β is the solution of the following equation:

$$\frac{6n}{\beta\left(\beta^3+2\right)}+\sum_{i=1}^{n}\frac{1+x_i}{\beta+\beta x_i^2}-n\bar{x}=0,$$

while the moment estimator of the parameter, β, is the solution of the following equation:

$$\bar{x}\beta^4-\beta^3+2\beta\bar{x}=0,$$

where $\overline{x}$ is the sample mean and *n* is the sample size. See Shanker and Shukla (2017) for more details on the Ishita distribution and its properties, and Shanker and Shukla (2018) for applications of the Ishita distribution in the context of lifetime-distribution data.

3. The suggested single-acceptance sampling plan

In this study, we suggest developing an SASP that uses the Ishita distribution to model the lifetime distribution of a product. To begin, suppose a producer claims that the specified mean lifetime distribution of the units is μ_0. Here, the problem is to find the minimum sample size necessary to ensure a certain average life (μ), when the lifetime-distribution test is terminated at a pre-assigned time (t) and when the number of failures does not exceed a given acceptance number, c. To perform the test according to this plan, a random sample of m units is selected from a lot. If μ_o can be obtained with a pre-assigned probability, P^*, as specified by the consumer, then the lot is accepted. If not, then it is rejected. Figure 4 provides a visual illustration of the SASP.

Assume that a product's lifetime distribution follows an Ishita distribution. An acceptance sampling plan that is based on truncated lifetime distributions tests consists of the following:

(1) The number of units, m, in the test.

(2) The acceptance number, c.

(3) The maximum test duration time, t.

(4) The ratio t / μ_0, where μ_0 is the specified average lifetime distribution.

In this case, the usual practice is to take a random sample from a product lot and then perform a truncated lifetime-distribution test. Accordingly, the lot under investigation will be accepted if there is enough evidence that $\mu \geq \mu_o$, at a certain level of producer risk α and consumer risk β^*. Now, assume that the lot is large enough to apply the binomial theory.

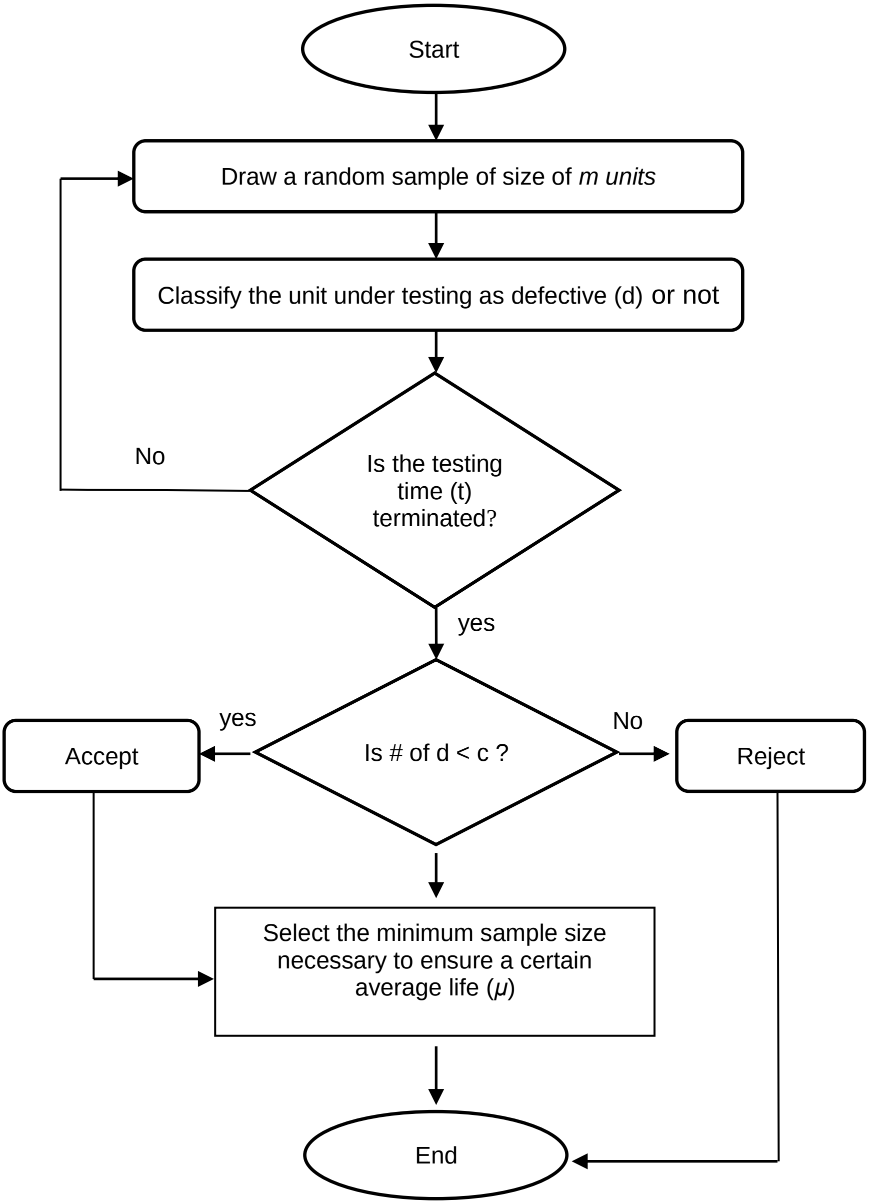


Figure 4. SASP algorithm

3.1. Minimum sample size

Given P^* and assuming that the lot size is large enough to be considered infinite, the probability of accepting a lot can be obtained based on the cumulative binomial distribution function up to the acceptance number, *c,* and the smallest sample size, *m,* to ensure that $\mu > \mu_0$ satisfies the following inequality:

$$\sum_{i=0}^{c} \binom{m}{i} p^i (1-p)^{m-i} \leq 1-P^*, \tag{3}$$

where $P^* \in (0,1)$, and $p = F(t;\mu_0)$ is a monotonically increasing function of t/μ_o and is known as the probability of a failure being observed during time *t*. If the number of observed failures within time *t* is *c*, at most, then, from inequality (3), we can confirm, with probability P^*, that $F(t;\mu) \leq F(t;\mu_0)$, which implies $\mu_0 \leq \mu$.

The minimum sample sizes that satisfy the above inequality for $t/\mu_0 = 0.628$, 0.942, 1.257, 1.571, 2.356, 3.141, 3.927, 4.712, with $P^* = 0.75$, 0.9, 0.95, 0.99 and $c = 0,1,2,...,10$. The values of t/μ_0 and P^* , presented in this work, are the same as the corresponding values found in Baklizi and El Masri (2004), for the Birnbaum Saunders model; in Kantam et al. (2001), for the log-logistic model; and in Gupta and Groll (1961), for the gamma distribution. Table 1 gives the minimum sample sizes for the new acceptance sampling plan for $\beta = 3$ in the Ishita distribution.

3.2. Operating characteristic function

For the sampling plan $(m, c, t/\mu_0)$, the operating characteristic (OC) function gives the probability of accepting the lot as the following:

$$OC(p) = P_r(Accepting\ a\ lot) = \sum_{i=0}^{c} \binom{m}{i} p^i (1-p)^{m-i}$$

where $p = F(t;\mu)$ is a function of μ (the parameter of the lot quality). It is of interest to say that, for a fixed *t*, the average lifetime of the products is increasing $\mu \geq \mu_0$ implying that the failure probability $p = F(t;\mu)$ is a monotonically decreasing function of $\mu \geq \mu_0$, and therefore the operating characteristic function *OC(p)* is increasing in the average lifetime; which means that as the ratio value $\frac{\mu}{\mu_0}$ is increasing say from "2 to 12", then the operating characteristic function *OC(p)* is increasing. Table 2 presents the *OC* function values as a function of $\mu \geq \mu_0$ for the sampling plan $\left(m,\, c=2,\, t/\mu_0\right)$

3.3. Producer's risk

The producer's risk is defined as the probability of the consumer rejecting the lot when $\mu > \mu_0$. This is given by the following:

$$P_r(p) = P_r(Rejecting\ a\ lot) = \sum_{i=c+1}^{m} \binom{m}{i} p^i (1-p)^{m-i}$$

For a given value of the producer's risk, say ϕ, under a given sampling plan, one may be interested to know the smallest value of μ/μ_0 that will ensure that the producer's risk is ϕ, at most. The value of μ/μ_0 is the minimum positive number for which $p = F\left(\frac{t}{\mu_0}\frac{\mu_0}{\mu}\right)$ satisfies the inequality presented in Inequality 4

$$P_r(p) = \sum_{i=c+1}^{m} \binom{m}{i} p^i (1-p)^{m-i} \leq \phi \qquad (4)$$

Table 3 summarizes the smallest values of the ratio μ/μ_0 that satisfy Inequality (4) for $\beta = 3$ in an Ishita distribution with a given acceptance sampling plan $\left(m,\, c,\, t/\mu_0\right)$ at the confidence level P^*. As the tables show, we assume that the lifetime distribution follows the Ishita distribution where $\beta = 3$. Table 1 shows the smallest sample sizes necessary to assert that the product's mean lifetime exceeds μ_0 with a probability of at least P^* and an associated acceptance number *c.* For example, assume that the researcher aims to ensure that the product's mean lifetime distribution is at least 1000 hours, with probability $P^* = 0.90$ with and the acceptance number of $c = 2$, such that the lifetime-distribution test is at least $t = 2356$ hours; that is, $t/\mu_0 = 2.356$. Then, from Table 1, the sample size is $m = 4$ units, which is the number of units that should be tested. Further, if 2 units, at most, out of the 4, fail before the specified time, *t,* within 1000 hours, then the product lot is accepted with a confidence level of 0.90. Hence, the process is to truncate the time of the test to a time of 2.356 of the specified mean lifetime distribution.

Table 2 summarizes the values of the OC function for the new acceptance sampling plans that were adopted, based on Table 1, for various values of t/μ_0

and P^* with $c=2$. Table 3 shows that the OC values for the sampling plan $\left(m=4,\ c=2,\ t/\mu_0=2.356\right)$ are as follows:

$\frac{\mu}{\mu_0}$	2	4	6	8	10	12
OC	0.352196	0.749243	0.88624	0.940036	0.964799	0.977655

Therefore, if the true mean lifetime distribution is six times the specified mean $(\mu/\mu_0=6)$, then the producer's risk is about 0.11376, and the producer's risk is about 0.647804, 0.250757, 0.059964, 0.035201, 0.022345, respectively, for $\mu/\mu_0=2,4,8,10,12$, and approaches zero for large values of μ/μ_0.

Table 3 provides the minimum ratios of the true mean lifetime distribution to the specified mean lifetime distribution (μ/μ_0) for the acceptance of a product's lot with the producer's risk $\phi=0.05$, with different values of t/μ_0.and the acceptance number c. As an illustration, the smallest value of μ/μ_0 is 8.645 for $P^*=0.90$, $t/\mu_0=2.356$, and $c=2$. That is, the product must have a mean lifetime distribution that is greater than 8645 hours, in order for the lot to be accepted. Hence, Table 3 shows the true mean lifetime distribution that is necessary for accepting 95% of the product's lots.

4. Concluding Remarks

In this paper, we developed a single-acceptance sampling plan for an Ishita distribution (a new lifetime distribution) under the assumption that the mean lifetime is a preassigned quality parameter. The proposed new SASP parameters are obtained to simultaneously measure the consumer and producer's risk in a product as measured by the lifetime distribution of a representative lot of the product. Several computations and numerical examples and their outcomes were presented and discussed.

The operating characteristics of the single-acceptance sampling plan is provided numerically. Based on the results, the consumers and producers are advised to adopt this plan in order to save time and minimize production process cost. Future research could be conducted on single-acceptance sampling plans by assuming the products' has different lifetime distributions.

Acknowledgments

We thank the anonymous referees and the journal editors for their useful suggestions.

REFERENCES

AL-NASSER, A. D., AL-OMARI, A. I., (2013). Acceptance sampling plan based on truncated life tests for exponentiated Frechet distribution. Journal of Statistics and Management Systems, 16 (1), pp. 13–24.

AL-NASSER, A. D., GOGAH, F. S., (2017). On Using The Median Ranked Set Sampling for Developing Reliability Test Plans Under Generalized Exponential Distribution. Pakistan Journal of Statistics and Operation Research, 13 (4), pp. 757–774.

AL-OMARI, AI, AL-NASSER A. D., GOGAH, F. S., (2016a). Double Acceptance Sampling Plan for Time-Truncated Life Tests Based on Half Normal Distribution, Economic Quality Control, 31 (2), pp. 93–99.

AL-OMARI, A. I., AL-NASSER, A. D. , GOGAH, F. S., (2016b). Double acceptance sampling plan for time truncated life tests based on transmuted new Weibull-Pareto distribution. Electronic Journal of Applied Statistical Analysis, 9 (3), pp. 520–529

AL-OMARI, A. I., (2015). Time truncated acceptance sampling plans for generalized inverted exponential distribution. Electronic Journal of Applied Statistical Analysis, 8 (1), pp. 1–12.

ASLAM, M., JUN, C. H., AHMED, M., (2011). New acceptance sampling plans based on life tests for Birnbaum-Saunders distributions. Journal of Statistical Computation and Simulation, 81 (4), pp. 461–470.

BAKLIZI, A., E. L. MASRI, A., (2004). Acceptance sampling based on truncated life tests in the Birnbaum Saunders model. Risk Analysis, 24 (6), pp. 1453–1457.

BALAKRISHNAN, N., LEIVA, L., LOPEZ, J., (2007). Acceptance sampling plans from truncated life tests based on the generalized Birnbaum-Saunders distribution. Communications in Statistics – Simulation and Computation, 36, pp. 643–656.

GOGAH, F. S, AL-NASSER, A. D., (2018). Median Ranked Acceptance Sampling Plans for Exponential Distribution. Afrika Matematika, 29 (3-4), pp. 477-497. https://doi.org/10.1007/s13370-018-0555-7.

GUPTA, S. S., GROLL, P. A., (1961). Gamma distribution in acceptance sampling based on life tests. Journal of American Statistical Association, 56, pp. 942–970.

KAO, J. K., (1966) “Single sample attri-variate plans for item variability in percent defective,” Annual Technical Conference Transactions of the American Society for Quality Control, New York, NY, USA, pp. 743–758.

KANTAM, R. R. L., ROSAIAH, K., RAO, G. S., (2001). Acceptance sampling based on life tests: log-logistic model. Journal of Applied Statistics, 28, pp 121–128.

SHANKER, R., SHUKLA, K. K., (2017). Ishita distribution and its applications. Biometrics & Biostatistics International Journal, 4 (7), pp. 1–9.

SHUKLA, K. K., SHANKER, R., (2018). Power Ishita distribution and its application to model lifetime data. To appear in Statistics in Transition new series, 19 (1).

APPENDIX

Table 1. Minimum values of *m* necessary to assert the mean life exceeds μ_0, with various combinations of P^*, *c* and t/μ_0 with $\beta = 3$.

		t/μ_0							
P^*	*c*	0.628	0.942	1.257	1.571	2.356	3.141	3.927	4.712
0.75	0	3	2	2	1	1	1	1	1
	1	5	4	3	3	2	2	2	2
	2	8	6	5	4	4	3	3	3
	3	10	8	6	6	5	4	4	4
	4	12	9	8	7	6	5	5	5
	5	15	11	9	8	7	7	6	6
	6	17	13	11	10	8	8	7	7
	7	19	15	12	11	9	9	8	8
	8	22	16	14	12	11	10	9	9
	9	24	18	15	14	12	11	10	10
	10	26	20	17	15	13	12	11	11
0.90	0	4	3	2	2	1	1	1	1
	1	7	5	4	4	3	2	2	2
	2	10	7	6	5	4	4	3	3
	3	12	9	8	7	5	5	4	4
	4	15	11	9	8	7	6	6	5
	5	18	13	11	9	8	7	7	6
	6	20	15	12	11	9	8	8	7
	7	22	17	14	12	10	9	9	8
	8	25	19	15	14	11	10	10	9
	9	27	20	17	15	13	11	11	11
	10	30	22	19	16	14	13	12	12
0.95	0	5	4	3	2	2	1	1	1
	1	8	6	5	4	3	3	2	2
	2	11	8	7	6	5	4	4	3
	3	14	10	8	7	6	5	5	4
	4	17	12	10	9	7	6	6	6
	5	19	14	12	10	8	7	7	7
	6	22	16	13	12	10	9	8	8
	7	25	18	15	13	11	10	9	9
	8	27	20	17	15	12	11	10	10
	9	30	22	18	16	13	12	11	11
	10	32	24	20	17	14	13	12	12
0.99	0	8	5	4	3	2	2	2	2
	1	11	8	6	5	4	3	3	3
	2	15	10	8	7	5	5	4	4
	3	18	13	10	9	7	6	5	5
	4	21	15	12	10	8	7	6	6
	5	23	17	14	12	9	8	8	7
	6	26	19	15	13	11	9	9	8
	7	29	21	17	15	12	10	10	9
	8	32	23	19	16	13	12	11	10
	9	34	25	20	18	14	13	12	11
	10	37	27	22	19	16	14	13	12

Table 2. OC values for the sampling plan from Table (1) for P^* at $c = 2$ for $\beta = 3$.

			μ / μ_0					
P*	m	t / μ_0	2	4	6	8	10	12
0.75	8	0.628	0.602233	0.890665	0.957192	0.97918	0.98838	0.992875
	6	0.942	0.570169	0.876254	0.950618	0.975735	0.98637	0.991606
	5	1.257	0.540865	0.861888	0.94385	0.972129	0.984245	0.990255
	4	1.571	0.601762	0.886188	0.954551	0.977643	0.98743	0.992253
	4	2.356	0.352196	0.749243	0.88624	0.940036	0.964799	0.977655
	3	3.141	0.500641	0.829677	0.926184	0.961977	0.977985	0.986153
	3	3.927	0.368648	0.745812	0.881115	0.936076	0.96196	0.975614
	3	4.712	0.266775	0.660235	0.829642	0.904632	0.941746	0.961967
0.90	10	0.628	0.437279	0.813833	0.921665	0.960433	0.977397	0.985921
	7	0.942	0.447328	0.817579	0.923199	0.961184	0.977816	0.986176
	6	1.257	0.385506	0.779748	0.903955	0.950524	0.971384	0.982023
	5	1.571	0.39875	0.785633	0.906547	0.951846	0.97214	0.982493
	4	2.356	0.352196	0.749243	0.88624	0.940036	0.964799	0.977655
	4	3.141	0.19108	0.601946	0.79763	0.886265	0.930449	0.954586
	3	3.927	0.368648	0.745812	0.881115	0.936076	0.96196	0.975614
	3	4.712	0.266775	0.660235	0.829642	0.904632	0.941746	0.961967
0.95	11	0.628	0.365942	0.771385	0.900386	0.948743	0.970382	0.981407
	8	0.942	0.341996	0.753433	0.890665	0.943192	0.966973	0.97918
	7	1.257	0.263901	0.691085	0.855943	0.92306	0.954488	0.970967
	6	1.571	0.247989	0.67441	0.845759	0.916873	0.950542	0.968323
	5	2.356	0.166013	0.580453	0.785718	0.879448	0.926275	0.95187
	4	3.141	0.19108	0.601946	0.79763	0.886265	0.930449	0.954586
	4	3.927	0.099455	0.4663	0.699831	0.820982	0.886219	0.923671
	3	4.712	0.266775	0.660235	0.829642	0.904632	0.941746	0.961967
0.99	15	0.628	0.164133	0.593595	0.797939	0.888306	0.93251	0.956321
	10	0.942	0.188243	0.619592	0.813833	0.897949	0.938652	0.960433
	8	1.257	0.175063	0.601759	0.802013	0.890475	0.93377	0.957109
	7	1.571	0.147227	0.563691	0.776478	0.874204	0.923088	0.949808
	5	2.356	0.166013	0.580453	0.785718	0.879448	0.926275	0.95187
	5	3.141	0.063146	0.398954	0.648376	0.78576	0.862024	0.906613
	4	3.927	0.099455	0.4663	0.699831	0.820982	0.886219	0.923671
	4	4.712	0.050625	0.352196	0.601885	0.749243	0.83473	0.88624

Table 3. Minimum ratio of true mean life over μ_0 at the producer's risk of 0.05 under the $\beta = 3$.

		t/μ_0							
*P**	*c*	0.628	0.942	1.257	1.571	2.356	3.141	3.927	4.712
0.75	0	38.909	38.897	51.903	32.403	48.594	64.785	80.997	97.188
	1	8.347	9.685	9.112	11.389	9.775	13.031	16.292	19.549
	2	5.627	5.969	6.302	5.765	8.645	7.123	8.906	10.686
	3	4.071	4.622	4.16	5.199	5.862	5.079	6.35	7.619
	4	3.309	3.412	3.854	3.928	4.52	4.058	5.074	6.088
	5	3.119	3.119	3.101	3.196	3.736	4.981	4.31	5.171
	6	2.771	2.915	3.046	3.271	3.224	4.298	3.799	4.558
	7	2.526	2.763	2.639	2.851	2.863	3.817	3.432	4.118
	8	2.489	2.426	2.643	2.542	3.219	3.458	3.155	3.786
	9	2.329	2.361	2.378	2.642	2.937	3.18	2.938	3.525
	10	2.202	2.307	2.4	2.418	2.713	2.958	2.762	3.314
0.9	0	51.887	58.363	51.903	64.869	48.594	64.785	80.997	97.188
	1	12.111	12.52	12.923	16.151	17.079	13.031	16.292	19.549
	2	7.266	7.207	7.965	7.876	8.645	11.525	8.906	10.686
	3	5.054	5.366	6.168	6.461	5.862	7.815	6.35	7.619
	4	4.333	4.449	4.553	4.816	5.891	6.026	7.533	6.088
	5	3.891	3.901	4.162	3.875	4.793	4.981	6.228	5.171
	6	3.386	3.538	3.47	3.807	4.082	4.298	5.374	4.558
	7	3.035	3.278	3.341	3.299	3.585	3.817	4.772	4.118
	8	2.92	3.082	2.941	3.304	3.219	3.458	4.323	3.786
	9	2.702	2.741	2.895	2.972	3.458	3.18	3.976	4.771
	10	2.64	2.641	2.854	2.711	3.178	3.617	3.699	4.438
0.95	0	64.865	77.83	77.88	64.869	97.282	64.785	80.997	97.188
	1	13.99	15.346	16.706	16.151	17.079	22.769	16.292	19.549
	2	8.084	8.44	9.617	9.955	11.811	11.525	14.409	10.686
	3	6.034	6.106	6.168	6.461	7.797	7.815	9.771	7.619
	4	5.014	4.963	5.247	5.691	5.891	6.026	7.533	9.039
	5	4.148	4.29	4.685	4.542	4.793	4.981	6.228	7.472
	6	3.795	3.847	3.889	4.336	4.905	5.442	5.374	6.448
	7	3.541	3.534	3.687	3.739	4.275	4.78	4.772	5.725
	8	3.206	3.3	3.53	3.676	3.812	4.291	4.323	5.187
	9	3.074	3.118	3.151	3.297	3.458	3.915	3.976	4.771
	10	2.859	2.973	3.078	2.999	3.178	3.617	3.699	4.438
0.99	0	103.799	97.297	103.856	97.334	97.282	129.695	162.15	194.564
	1	19.62	20.984	20.477	20.879	24.221	22.769	28.467	34.157
	2	11.349	10.899	11.263	12.02	11.811	15.746	14.409	17.289
	3	7.991	8.316	8.147	8.948	9.689	10.395	9.771	11.724
	4	6.372	6.5	6.623	6.557	7.222	7.854	7.533	9.039
	5	5.173	5.451	5.725	5.856	5.812	6.389	7.988	7.472
	6	4.611	4.772	4.721	4.86	5.709	5.442	6.804	6.448
	7	4.215	4.298	4.374	4.608	4.947	4.78	5.976	5.725
	8	3.92	3.948	4.113	4.045	4.388	5.082	5.365	5.187
	9	3.569	3.68	3.658	3.938	3.962	4.61	4.895	4.771
	10	3.403	3.468	3.524	3.567	4.066	4.236	4.522	4.438

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 407–431, DOI 10.21307/stattrans-2018-023

THE ROLE OF BREASTFEEDING VIS-À-VIS CONTRACEPTIVE USE ON BIRTH SPACING IN INDIA: A REGIONAL ANALYSIS

Laxmi Kant Dwivedi[1]

ABSTRACT

Birth spacing is one of the important aspects of reproductive health. Therefore, it is felt by demographers that birth spacing needs to be studied from time to time in view of the epidemiological transition taking place worldwide. Using the third round of National Family Health Survey-3 data, the central hypothesis of this paper is to find out the relative advantages of breastfeeding over other methods of contraception among non-sterilized women by using simulative approach of the Cox regression analysis in India and its regions. The results show that if women were not having amenorrhea period and had a high level of breastfeeding, the chance of not having next live birth was only two percent lower than those women who were using spacing methods in India. This pattern was found to be almost similar in all the regions of India except central and southern regions. There is no significant gain in postponing the next live birth has been observed in using the contraceptives than breastfeeding. An effort has also been made to apprise the policymakers of the interrelation between breastfeeding, postpartum amenorrhea, contraceptive use and birth spacing. Nonetheless, policymakers should promote programs that encourage both breastfeeding and contraceptive use. Breastfeeding has direct benefits for infant health in addition to its role in lengthening birth intervals beyond postpartum amenorrhea.

Key words: breastfeeding, birth spacing, contraception, Cox regression, simulation analysis.

1. Introduction

The numerous advantages of breastfeeding have been accepted by the health and family planning policy makers and various initiatives have been taken to promote breastfeeding. In addition to the benefits to child from many illnesses, it protects mother against another pregnancy. Further, it increases child survival and suppresses ovulation during which the chances of conception are virtually nil. These important roles of breastfeeding have been well documented in the literature of demography/and public health. Therefore, it is felt by

[1] Assistant Professor, Department of Mathematical Demography & Statistics International Institute for Population Sciences Govandi Station Road, Deonar, Mumbai 400 088, India. E-mail: laxmikdwivedi@gmail.com, laxmikant@iips.net.

demographers/epidemiologists working in the areas of public health that the issues related to the advantages of breastfeeding need to be studied from time to time in view of the epidemiological transition that has been taking place worldwide.

Previous investigations of postpartum amenorrhea (PPA) in developing countries suggest that the distribution of amenorrhea is bimodal composed of a "normal" duration subgroup and a short duration subgroup that resumes menses within 3 or 4 months (Henry, 1961; Saxena and Pathak, 1977; Holman *et al.,* 2006). The duration as well as the nature of breastfeeding are the major determinants of prolonged PPA and are well documented in both aggregate and individual level analyses. This phenomenon has been verified by eminent researchers (Saxena, 1977; Howie and McNeilly, 1982; Bongaarts, 1983; Srinivasan *et al.,* 1989; Nath *et al.*, 1994; Singh *et al*., 1999; Arokiasamy, 2002). Further, prolonged PPA works as a catalyst in increasing the birth interval. Many studies found the contraceptive effect of breastfeeding, especially regarding the circumstances when it becomes more effective and safe. But, contraceptive role of breastfeeding is not fully established. In a consensus statement, (Family Health International, 1988) a group of international societies put forward the view that when mothers breastfeed exclusively or near to that, there is a higher chance that a woman remains under amenorrhea. Under such conditions, almost 95 percent women are protected against pregnancy.

Birth spacing is defined as an interval between termination of one completed pregnancy and the termination of the next (Last, 1988). The study on birth spacing pattern not only determines the pace of childbearing but also reflects the likelihood of progressing to a higher parity, which further determines the completed family size. The high levels of fertility, especially in the central region of India, are the major concern to the planners and policy makers. Therefore, analysis of birth spacing is of interest in this context since it can provide further insight into the mechanism underlying fertility change (Potter, 1963; Sheps, 1964; Pathak, 1966; Sehgal, 1971; Srinivasan, 1980; Njogu & Martin, 1991). Studies also revealed that birth spacing is preferred over other conventional measures of fertility because of its sensitiveness to small and short term changes in the reproduction rate (Singh, 1964; Sheps & Menken, 1972; Namboodiri, 1974; Namboodiri, 1983). Further, birth spacing provides the mechanism of reproductive process and, therefore, it can be considered as a major determinant of population change (Mturi, 1997). A study of birth interval length with various socio-economic and demographic variables helps in finding out the relative importance of factors that contribute to fertility decline. Further, it may also help in identifying the factors that create obstacles to further reduction in the fertility.

There are many hypotheses that have been tested in the areas of public health/demography that are solely based on analytical research on birth spacing. Hypotheses related to birth spacing have generated various important clues towards public health programs and have important implications for a number of reasons. For example, "to what extent does the length of the preceding birth interval affects the risks of infant and child mortality?" has been examined with the help of birth spacing data (DaVanzo et al., 2004). A study between birth spacing and breastfeeding may also help in deciding the need for an individual woman to initiate contraception at proper time (Anderson, 1986). Further, a hypothesis may be framed in relation to the benefits of breastfeeding over other methods of contraception in the context of extending the subsequent birth spacing.

Using birth spacing data of three Southeast Asian countries, the researcher concluded that the length of previous birth interval is an important covariate in explaining the risk of pregnancy leading to a live birth after controlling the breastfeeding behaviour and the use of contraception (Trussell, 1985). Others also derived the same findings that duration of breastfeeding has a significant effect on the likelihood for a woman to go on to have a second or third birth in Vietnam (Swenson & Thang, 1993). The study compared results of identical structural models for nine countries and also found that the woman's education and the length of the previous birth interval had a substantial effect on birth interval (Rodriguez et al., 1983). They also concluded that parity is a relatively unimportant covariate. Finally, one of their general conclusions (Rodriguez et al., 1983) is that "It seems likely that many of the differences are the consequence of differing patterns of breastfeeding and contraceptive use." However, they have not included these variables in their analyses.

Using Malaysian Family Life Survey-1979-77 data, authors have investigated the contribution of different factors in elucidation the short birth interval (less than 15 months) in Peninsular Malaysia. Further, they have also explored how factors relate to breastfeeding and the use of contraceptive affect birth spacing (Da Vanzo and Starbird, 1991). They came out with the findings that breastfeeding had a considerably greater aggregate protective effect against early subsequent conceptions as compared to the use of contraceptives because more women breastfed than use contraceptives. They have also found that breastfeeding and contraceptive use are negatively related. Analysing the data from National Family Health Survey for Uttar Pradesh and Tamil Nadu, India, authors developed the hazards life table models for parity specific live birth intervals (Dwivedi and Singh, 2003). They came out with the findings that in each state, breastfeeding emerged as an important protective covariate that extended the birth spacing, irrespective of parity. However, another study found that breastfeeding is a statistically significant covariate in determining the length of birth interval (Ojha, 1998).

2. Objectives

Previous research on breastfeeding and contraceptive use has shown that the main determinants of breastfeeding and contraceptive use often act in opposite directions. For example, variables associated with modernization have a negative impact on breastfeeding but positive effects on contraceptive use (Potter, 1987b; Potter et al., 1987a; DaVanzo and Habicht, 1986; Butz, and Da Vanzo, 1981). Therefore, there is a need to investigate the impact of extended breastfeeding beyond PPA has an advantage in extending the birth spacing over the use of other methods of contraception. The main objective of the paper is to determine the impact of breastfeeding on birth spacing in India and its regions, among non-sterilized women who gave birth(s) during the last five years from the date of survey. An attempt has also been made to find out the relative advantages of breastfeeding over other methods of contraception in relation to birth spacing among amenorrheic and non-amenorrheic women in India and its regions. An appropriate simulation analysis will be carried out to explore important clues for the policy planners involved in population control/public health programs.

3. Methods

3.1. Data

To accomplish the objective, the relevant data have been taken from National Family Health Survey (NFHS), conducted in 2005-06. The analysis was carried out for India and its six regions: the northern region, which includes Delhi, Haryana, Himachal Pradesh, Jammu and Kashmir, Punjab and Rajasthan; the central region, which consists of Chhattisgarh, Madhya Pradesh, Uttaranchal and Uttar Pradesh; the eastern region, which comprises Bihar, Jharkhand, Orissa and West Bengal; the north-eastern region, which consists of Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura; the western region, which includes Goa, Gujarat and Maharashtra; and the southern region, which comprises Andhra Pradesh, Karnataka, Kerala and Tamil Nadu.

The dependent variable is birth spacing, that is interval (in months) between one live birth and next live birth. The absence of later birth, the birth spacing is considered to be the censored observation. The information on breastfeeding is available only for those women who gave birth(s) during last five years prior to the survey. Therefore, women who are not sterilized and gave birth(s) during last five years prior to the survey are included in the analysis. There were 40,905 women who had live births during last five years prior to the survey; of these 73 percent were censored observations at India level. The censored observation varies from 68 percent in the central India to 81 percent in the southern India.

There are also women in the sample who had more than one child during the last five years prior to the date of survey. In this paper, wherever women or mothers have been mentioned, these women/mothers actually refer to mothers of the index child.

The combined variable of women currently breastfeeding and women in amenorrhea is the independent variable. The categories of this variable are as follows:

(i) currently breastfeeding and amenorrheic;

(ii) currently not breastfeeding and amenorrheic;

(iii) never breastfed and amenorrheic;

(iv) currently breastfeeding and non-amenorrheic;

(v) currently not breastfeeding and non-amenorrheic; and

(vi) never breastfed and non-amenorrheic.

3.2.1. Kaplan-Meier (K-M) methods

For bivariate analysis, Kaplan-Meier (K-M) survival analysis has been used to determine the mean duration of birth spacing. The differences in the mean duration of birth spacing among different combined categories of breastfeeding and amenorrhea status (as stated above) have been examined by log-rank test.

3.2.2. Cox proportional hazards model and its Simulation Analysis

The Cox proportional hazards model was used to examine the adjusted impact of combined variables of breastfeeding and amenorrhea on birth spacing. Further, to find out the relative advantages of breastfeeding over other methods of contraception among amenorrheic and non-amenorrheic women, the simulation approach has been adopted.

There are several important reasons why the Cox model is more widely adopted in the field of demography/public health. The exponential part of the Cox model is appealing because it ensures the validity of the definition of hazard function, that is, the estimated hazard will be always non-negative. Another appealing property of the Cox model is that the unknown coefficient in the exponential part of the model can be estimated without specifying the baseline hazard. In the absence of the specific baseline hazards function, the hazards function and its corresponding survival curve can also be estimated for the Cox model. Thus, the primary information desired for a survival analysis, namely, a hazard ratio and a survival curve, may be obtained using a minimum of assumptions. This model is also preferred over the logistic model when survival time information is available and there is censoring (Kleinbaum, 1996a; Klienbaum, 1996b). The Cox model uses more information-the survival times-than the logistic model, which considers a (0, 1) outcome and ignores survival times and censoring. Therefore, analysis of survival time, the time to next live birth for a non-sterilized currently married woman has been carried out through the use of the Cox hazards model (Cox, 1972).

The simulation exercise of the Cox hazards model has been done in the four steps.

Step 1: The exponential expression of the Cox model, also known as 'Risk score' and generally denoted by R, may be defined as follows:

$$R = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p \quad (1)$$

where X_1 , X_2 , ..., X_p are the defined levels of p predictor variables and β_1 ,β_2 , ….β_p are respective unknown regression coefficients.

Risk score R has been calculated for every woman included in the data set after substituting the observed values of the covariates for each individuals using maximum likelihood estimates of regression coefficients. Further, the average risk score (R_1), which is a constant for a given data set, has been computed.

Step 2: The next step is to calculate the average value of risk score R_2 after changing the levels of included variables in Step 1 by using the equation (1). Risk score will be obtained after substituting the changed levels of the selected variable. However, the changed level will remain the same for every woman. The value of risk score may be varied from woman to woman as a result of variation in the levels of the selected variable.

Step 3: The baseline survival probabilities ($S_0(t)$) at different time points for a woman with average risk score R_1 may be obtained using the Kaplan-Meier method.

Step 4: The gain in survival probability may be worked out by

$$S(t) = S_0(t)^{\exp(R_2 - R_1)} \qquad (2)$$

The survival probabilities in relation to R_1 are listed in the first row of the concerned table, whereas the survival probabilities due to change in the level (R_2) of the selected variable are listed in successive rows. The difference between the two probabilities provides gain or loss as a result of proposed change in the levels of the selected variable or set of variables.

4. Results

4.1. Kaplan-Meier/Survival Analysis

Table 1 shows the mean duration of next birth intervals (in month) with 95% confidence interval (CI) estimates for India and its regions. The mean duration of next birth interval for India (all regions combined) was 42 months. The central region had the lowest mean duration of next birth interval (39.3 months). However, the highest figure was observed for the south region (Mean: 45.8, C.I. 45.0 – 46.5) followed by the western region (Mean: 44.3, C.I. 43.6 – 45.0). The other three regions were all between 42 and 43 months (see Table 1). Table 2 shows the mean duration of next birth intervals (in month) with their 95% confidence interval (CI) estimates for India and its regions found by the current states of breastfeeding and amenorrhea of the women as classified in different states (i) to (vi).

The mean duration of next birth interval appears to increase when mothers were not having amenorrhea period in comparison with those mothers who were amenorrheic, irrespective of the breastfeeding status as found for women in India and its regions. For example, mothers who were still breastfeeding and were not amenorrheic had a significantly longer mean duration of next birth interval in comparison with those mothers who were also still in state (i) for breastfeeding and amenorrhea.

Similarly, the impact of breastfeeding in postponing the next live birth can be seen for women in India and its regions, irrespective of whether women were having or not having postpartum amenorrhea. For example, women who were still breastfeeding and were in amenorrhea period had a significantly longer duration of birth interval than those women who were currently not breastfeeding at the time of survey but were amenorrhic. These findings have been found to be consistent for women in India and its regions.

Taking India as a whole, 45 percent of women were not having next live birth at least by 48 months - the same was true for women of the eastern region, whereas 44 percent and 37 percent of mothers belonging to the northern and central regions respectively, did not give birth up to 48 months. In the case of 48 percent of women of the north-eastern, 51 percent of western and 57 percent of southern regions did not give birth up to 48 months from previous birth. These rates at 48 months were found to be highest among those mothers of index children who were still breastfeeding but not found to be amenorrheic at the time of the survey.

4.2.1. Multivariate Analysis

To identify important factors affecting birth spacing among women in India and its regions, a host of possible covariates were considered in the Cox proportional hazards model. The choice governing the selection of explanatory variables of the Cox proportional hazards model was that each variable considered should show at least moderate association ($p < 0.05$) in bivariate analyses for at least one out of six regions considered. However, in some cases, more importance was given to theoretical rather than purely statistical considerations. The analysis has been carried out separately for India and its six regions and the results are presented in terms of rate ratio/relative risks (exponential of regression coefficients) and their 95% confidence intervals in Table 3 and 4.

Table related to India clearly shows that mothers of index children from the northern and central regions had a significantly higher likelihood to experience the next live birth in comparison with the southern region. Mothers of index children other than the northern and central regions had not statistically significant association with birth spacing as opposed to the southern region.

With regard to breastfeeding and postpartum amenorrhea, mothers who were still breastfeeding and were having amenorrhea period were not found statistically significant association with birth spacing (RR: 1.05, C.I. 0.89 – 1.23). However, women who were amenorrheic but either never breastfed or were currently not breastfeeding had significant risk factor against the next live birth. It is also true for those women who never breastfed and were not having amenorrhea period (RR: 1.31, C.I. 1.20 – 1.43). Surprisingly, women who were still breastfeeding and were not having amenorrhea period had a lower chance of having next live birth as compared to those women who were currently not breastfeeding and were not having amenorrhea period (RR: 0.16, C.I. 0.14 – 0.18). A regional analysis showed that mothers who were still breastfeeding and were amenorrheic (state-i) at the time of survey had not statistically significant higher chance as compared to those mothers who were not breastfeeding and were not found to be amenorrheic (state-v) at the time of survey except for women of the central region where probability of having next live birth was found to be less than one (RR: 0.79, C.I. 0.58 – 1.09). Further, the table shows that the adjusted chance of having next live birth was found to be significantly high among those mothers who were having amenorrhea period at the time of survey and had never breastfed or were not breastfeeding at the time of survey. The association was found to be poor in the southern region ($p>0.05$). However, in the western region, the risk was not statistically significant for those mothers who were amenorrheic but never breastfed. Mothers who were still breastfeeding and had resumed menstruation cycle had significantly less chance of having next birth in comparison with those women who were not breastfeeding and were not in amenorrhea period in all the regions of India. The lowest value of relative risk was found to be in the central region (RR: 0.09, C.I. 0.06 – 0.12). Women who had never breastfed and were not found in amenorrheic state had a significantly greater likelihood to experience the next live birth as compared to the women who were currently not breastfeeding and were not in amenorrhea period. However, the association was found to be poor in the central and south regions ($p>0.05$).

4.2.2 Simulation Analysis

The results of the Cox proportional hazards model discussed in the earlier section can be found useful for policy planners who are primarily responsible for public health management. The utilities of the Cox proportional hazards model to the policy planners may be easily shown by calculating the predicted survival probabilities at a considered level of a variable by holding all other variables at their average level in the model. In a similar way, the survival probabilities at considered levels of various variables at a time can also be calculated by keeping all other variables at their prevailing average level. It may, however, be noted that the probability of survival in the present context means the probability of not attaining the next live birth. The probabilities of not having next live birth may help in finding the expected gain by changing the level of the selected predictor and keeping other factors on their average value. This additional gain can easily be obtained by subtracting the probabilities found for a given level of a variable from that reported for the average level.

The selected predictors and their combinations considered in the present prediction analysis are: (i) primary education of woman; (ii) secondary education of woman; (iii) survival of index child; (iv) current use of contraceptive; (v) breastfeeding and postpartum amenorrhea; (vi) breastfeeding and postpartum amenorrhea with survival of index child; (vii) breastfeeding and postpartum amenorrhea with current use of contraceptive.

The categories of breastfeeding are: (i) never breastfed (ii) currently not breastfeeding and (iii) still breastfeeding. The present states and the changed states of breastfeeding for the given level of breastfeeding are defined as:

Level of breastfeeding	**Present status of breastfeeding**	**Changed status of breastfeeding**
Lowest	Never breastfed	Currently not breastfeeding
Low	Never breastfed	Still breastfeeding
Moderate	Currently not breastfeeding	Still breastfeeding
High	Never breastfed and currently not breastfeeding	Still breastfeeding

In this section, the lowest level of breastfeeding is defined for those mothers who never breastfed and are considered in the category of currently not breastfeeding. The low level stands for mothers who never breastfed and are considered in the category of still breastfeeding, whereas the moderate level of breastfeeding refers to those mothers who were currently not breastfeeding and are included in the category of still breastfeeding by keeping the original value of the third left out category of the three categories defined as above for different levels of breastfeeding. However, the high level of breastfeeding is defined for those mothers who never breastfed and were currently not breastfeeding. They were included in the category of still breastfeeding.

The probabilities of not having next live birth by selected variables as well as their combinations were worked out for various durations of birth spacing (considered months were 12, 18, 24, 30, 36, 42, and 48) for India and its regions. It may, however, be noted that the results related to the above mentioned variables and their combinations are possible only if these variables are present in the finally considered model. The prediction results in relation to the considered variables and their combinations are presented in Table 5 and Figures 1, 2 and 3 for India and all the six regions. These results give the probabilities of not having the next live birth for women by their selected characteristics or their combinations keeping the remaining variables as constant at their average value.

If it is assumed that all mothers of index children had a high level of breastfeeding and were not amenorrheic, the maximum benefit of birth spacing can be derived only after the period of 24 months. The gain in the birth spacing at the 48 months was found to be highest in the central and southern regions of India. The least benefit at the 48 months has been observed in the northern part of India. However, if mothers of index children had a moderate level of breastfeeding, the probability of not having next live birth at 48 months was higher than the average value in India for all the regions of India. The gain from the average value was found to be highest in the north, eastern and western regions (23 percent) followed by the north-eastern (21 percent), central (20 percent) and southern regions (12 percent). The benefit reaches its minimum value if the simulation of data is done only on those mothers of index children who have never breastfed and are treated as still breastfeeding (low level of breastfeeding).

If it is considered that all mothers of index children were amenorrheic, a large benefit was not found compared to the average value according to the different levels of breastfeeding over different time periods of birth spacing for mothers in India and its regions. A comparison can also be made between mothers who were amenorrheic and those who were not by different levels of breastfeeding with survival status of the index child. The results clearly exhibit the importance of breastfeeding when women were found to be non-amenorrheic as opposed to their counterparts who were in the postpartum amenorrhea period, irrespective of survival status of the index child. These findings were found to be consistent in all the regions of India, but the magnitude varied from one region to another. For example, in the central region, mothers who were not amenorrheic but having a high level of breastfeeding had 28 percent higher chance of not having next live birth at 48 months than those mothers who were amenorrheic with a high level of breastfeeding. The value for the northern, eastern and western regions was 24 percent whereas for the north-eastern and southern regions, the values were 22 percent and 12 percent, respectively. It clearly reflects the importance of breastfeeding beyond the amenorrhea period. These results are also found to be consistent, irrespective of the current use of contraceptives.

However, the impact of different levels of breastfeeding in postponing the next live birth has not been observed among those mothers who were amenorrheic, irrespective of the survival of index child or current use of contraceptives. But still, the probability of not having next live birth according to the different levels of breastfeeding among those mothers who were having amenorrhea over a different time period was slightly higher than the average value in all the regions of India.

All those women who were not sterilized and were protected by using any birth spacing method (traditional and/or modern), were amenorrheic and had a high level of breastfeeding, had only two percent higher chance of not having next live birth as compared to its average value. The figure for this group of women in the central region was four percent higher than its average value and was found to be highest. It is interesting to note that this value was lower than the average value in the southern region. It clearly suggests the argument that overlapping of different methods of contraception does not help in further postponing the next live birth.

If women were not amenorrheic and had a high level of breastfeeding, the chance of not having next live birth was only two percent lower than those mothers who were using spacing methods in case of India. This pattern was found to be almost similar in all the regions of India except the central and southern regions. Further, the value for those regions where the pattern was similar to India was one percent lower for the northern and northeastern regions and two percent lower for the eastern and western regions. No difference could be found between these two different groups of women in the central region. However, the value was three percent higher among those mothers of index children who were not amenorrheic and had a high level of breastfeeding as compared to those mothers who were using spacing methods in the case of mothers of the southern region. It can be reasonably argued that breastfeeding is one of the important factors in postponing the next live birth.

5. Discussion and Conclusions

The relationship between breastfeeding and birth spacing has important implications for public health programs for a number of reasons. These include the direct health benefits of breastfeeding for the child, the maternal and child health benefits of birth spacing, the aggregate impact of breastfeeding on the birth rate and the need for individual women to initiate contraception at the proper time. Birth spacing, that is the timing between two live births, is also a major determinant of population growth. It is observed that the shorter the spacing between successive live births, the narrower the interval between generations and the faster would be the rate of population growth. In areas where the practice of contraception is not widespread, the spacing of births may be the primarily responsible factor for population growth. Being an important component of reproductive process, birth spacing is sensitive to small and short term changes in the reproduction. Therefore, analytical research on birth spacing may help in testing several hypotheses relating to population issues and in generating various important clues towards public health programs. The determinants of changes in the pace of child-bearing can also provide more immediate feedback for administrators dealing with population problems. Similar arguments have been made in relation to important public health indicator-child nutritional status.

In an attempt to demonstrate the utility of the proposed models for policy planners, the expected survival probabilities for a subject (woman/child) with some selected characteristics were worked out. At the outset, using a particular final model, the average survival probability has been worked out by keeping all the variables in the model at their average levels. Further, these probabilities for a subject with a particular level of a variable (or a particular combination of variables) may be calculated by holding all remaining variables in the models constant, that is, at their average levels. Thus, difference in one of these probabilities with average survival probability will provide expected gain/loss from the proposed changes.

The findings clearly reveal the benefits of breastfeeding over other methods of contraception in postponing the next live birth beyond the PPA period. The explanation is that many more women breastfed for longer duration than using contraception. If it is assumed that all women in the sample were breastfeeding beyond the PPA period, around 11 percent additional women compared to the average value were able to postpone their next birth until 24 months. If all the women had breastfed, the gain in extending the period due to the use of contraceptives would not have been noticed. A study also came out with the results that breastfeeding had a considerably greater effect on preventing short birth intervals than did contraceptive use (Da Vanzo, and Starbird, 1991). Some researchers also found that breastfeeding beyond the PPA period had a positive correlation with the waiting time to conception (Nath et al., 1994).

If a woman is still breastfeeding and continues to do so after menstruation has resumed, it appears to decrease her chances of conceiving. One possible explanation of this finding is that breastfeeding may make women less available for sexual relations than would otherwise be the case (Anderson et al., 1986). The inverse relationship could be observed between breastfeeding and the use of contraceptive in all the six regions of India. Further, simulation results suggest that women who both breastfed and practiced contraception did neither as effectively as women who did only one.

After controlling the important socio-economic and demographic confounders, the effect of region was found to be significant. It clearly indicates that there were some unobserved factors that might be contributing to lengthening/shortening of the birth spacing. For example, the desire for another child may be varied from one to another regions and it may affect the women's behaviour regarding the use of contraception and prolonged breastfeeding.

It appears that place of residence, mother's education and standard of living were not found significantly associated with birth spacing. Studies clearly showed that place of residence, education and standard of living are significant predictors of breastfeeding behaviour and contraceptive adoption (Sahu, 1998; Shekhar, 2004; Dwivedi et al., 2007). They affect these two proximate determinants of interval length in opposite directions and essentially offset each other in their effects on the likelihood of having next live birth. However, in the case of central, north-eastern and western regions, the standard of living had positive impact on extending the birth spacing, therefore, the same results were also found at India level. The possible reason might be that as standard of living increases, the use of contraceptive also increases significantly in these regions.

Table 1. Summary of significant findings

Variables	India	North	Central	East	Northeast	West	South
Place of residence							
Urban®							
Rural	0	0	0	0	-	0	0
Religion							
Hindu®							
Muslim	0	0	0	+	0	0	+
Others	+	0	0	0	+	0	0
Mother's education							
Higher®							
Illiterate	0	+	0	0	0	0	0
Primary	0	0	0	0	-	0	0
Secondary	0	0	0	0	-	0	0
Standard of living							
Richest®							
Poorest	+	0	+	0	+	+	0
Poorer	+	0	+	0	+	0	0
Middle	+	0	+	0	+	+	0
Richer	+	0	0	0	+	+	0
Sex of index child							
Male®							
Female	+	+	0	+	0	+	0
Survival status of index child							
Alive®							
Dead	+	+	+	+	+	+	+
Mother's age at birth of index child (yrs.)							
20-24®							
< 20	+	0	+	+	+	0	0
25-29	-	-	-	-	-	-	-
30-34	-	-	-	-	-	-	-
35+	-	-	-	-	-	-	-
Current use of contraceptive							
Yes®							
No	+	0	0	0	0	0	0
Previous birth interval							
24-36 months®							
First birth	+	+	+	+	+	+	+
<24 months	+	0	+	0	+	0	+
>36 months	-	0	0	0	-	0	-

Table 1. Summary of significant findings (cont.)

Variables	India	North	Central	East	Northeast	West	South
Number of surviving children							
Two®							
None	-	-	-	-	-	-	-
One	-	-	-	-	-	-	-
Three	+	+	+	+	+	+	+
Four and above	+	+	+	+	+	+	+
Maternal BMI							
>=18.5Kg/m^2®							
<18.5Kg/m^2	+	0	0	+	0	0	0
Missing	0	0	0	0	0	0	0
Breastfeeding (BF) and postpartum amenorrhea(PPA)							
Currently not BF and not in PPA®							
Currently BF and not in PPA	-	-	-	-	-	-	-
Never breastfed and not in PPA	+	+	0	+	+	+	0
Currently BF and in PPA	0	0	0	0	0	0	0
Never BF and in PPA	+	+	+	+	+	0	0
Currently not BF and in PPA	+	+	+	+	+	+	0

Note: 0= not statistically significant at the 0.05 level. - = negative influence, significant at 0.05 level (2-tailed test). + = positive influence, significant at 0.05 level (2-tailed test).

It may be noted that women who were found to be sterilized at the time of survey were excluded from the analysis. However, if these women were included in the analysis, their inclusion would have resulted in an increase in the proportion of birth intervals that are very long, including those which remained open. This possibility could not be ignored that the proportion of long birth intervals would be reduced because Indian women has tendency to go for sterilization just after birth.

Muslim women from the eastern and southern regions; whereas other than Hindus and Muslims from the north-eastern region were more likely to experience the next live birth. Hindu-Muslim differentials in fertility, to an extent, may be due to duration of sexual abstinence after child birth (Bhat and Zavier, 2005). However, among all the predictors of Hindu-Muslim growth differentials, less use of contraceptives among eligible Muslim women has repeatedly been cited as the pivotal factor (Ram et al., 2007). Further, an increase in the fertility in the north-eastern states was also observed (Marbaniang, 2003). One of the important reasons was the increase in the level of wanted fertility (Shekhar et al., 2006).

In few cases, the relationships of a particular variable with breastfeeding and with contraceptive use reinforce one another in affecting the likelihood that the birth interval is either short or long. A child death reduces both the probability that the child is breastfed and the likelihood that contraception is practiced. Both the reasons are associated with a shorter birth interval. Female index child reduces the likelihood of breastfeeding and of contraceptive use, and is associated with a greater incidence of subsequent short intervals. The result was not found to be significant in the case of the central, north-eastern and southern regions; the possibility that in the southern and north-eastern regions the preference for sons over daughters is virtually nil. However, in the central region where preference for sons exists, chances were high that women might not be aware of contraceptive effect of breastfeeding; therefore, they preferred to go for longer duration of breastfeeding, irrespective of sex of the child. Further, the use of spacing methods was relatively low as resulted into no effect on birth spacing.

The mother's age at birth of the index child has a direct effect on birth interval in addition to its influence on breastfeeding or contraceptive use. For example, although women aged 35 and over are less likely to use spacing methods than younger women (Dwivedi et al., 2007), they are also significantly less likely to have next live birth. The possible reason may be due to their lower fecundity or frequency of intercourse. Women who had three or more children had a greater likelihood to have next live birth as compared to those women who had two children. The possible reasons for shorter birth spacing may be due to the preference for particular sex combinations of child or desire for a higher number of children by women. Chances are also high that a particular group of women might be living in the high fertility states. Women who had no living child or had one child were less likely to move to the next birth. These women may be experiencing primary or secondary sterility. The possibility is also that highly educated women have longer birth spacing after first birth. There could also be possibility that they desire no more children and such birth intervals remain open.

Results of 'previous birth interval' show that women who had shorter preceding birth interval, had a greater chance to have next child, whereas women who had longer preceding birth interval had their succeeding birth interval to be longer. This could be because the women would like to have next child due to loss of the index child or may desire to have a higher number of children. The nutritional status of mother was poorly associated with birth spacing in all the regions except in the eastern region. However, undernourished women were more likely to have longer duration of PPA after certain duration of breastfeeding but the effect on birth spacing was not observed.

The findings of the paper underscore how important it is for policymakers concerned with fertility reduction, health promotion of both mother and the child to monitor the level and trends in breastfeeding. An effort has been made to apprise the policymakers of the interrelation between breastfeeding, postpartum amenorrhea, contraceptive use and birth spacing. Nonetheless, policymakers should promote programs that encourage both breastfeeding and contraceptive use. Breastfeeding has direct benefits for infant health in addition to its role in lengthening birth intervals beyond PPA. Contraceptive use permits women to recuperate from the depleting effects of both pregnancy and breastfeeding (Merchant and Martorell, 1998), and also helps in increasing the birth interval.

In summary, in terms of policy implications, this study has revealed the importance of region specific epidemiological understanding of public health issues like postpartum amenorrhea and birth spacing. Within a region, findings may be helpful in planning that showed more region specific strategies. Child loss extended the birth spacing in each region significantly. Modifications in the behaviour of extended breastfeeding may also improve child survival leading to extended birth spacing. The results have reaffirmed the interrelationship between breastfeeding and birth spacing. In other words, extended birth spacing is necessary to ensure the survival of a child. In the same way, the survival of a child is necessary for extending the birth spacing.

REFERENCES

ANDERSON, J. E., BECKER, S., GUINENA, A. H., MCCARTHY, B. J., (1986). Breastfeeding Effects on Birth Interval Components. A Prospective Child Health Study in Gaza. Studies in Family Planning, Vol.17, 3, pp. 153–160.

AROKIASAMY, P (2002): Breastfeeding and Its Contraceptive Role on Postpartum Amenorrhoea and the Waiting Time to Conception in India. GENUS, 58 (1), pp. 121–158.

BHAT, P. N. M., ZAVIER, A. J. F., (2005). Role of Religion in Fertility Decline: The Case of Indian Muslims. Economic and Political Weekly, XL (5), pp. 385–402.

BONGAARTS, J, POTTER, R. G., (1983): Fertility, Biology and Behaviour: An Analysis of the Proximate Determinants, New York: Academic Press.

BUTZ, W. P., DA VANZO, J., (1981). Determinants of Breastfeeding and Weaning Patterns in Malaysia, Paper Presented at the Annual Meeting of the Population Association of America. Washington, DC, pp. 26–28 March.

COX, D. R., (1972). Regression Models and Life Tables, Journal of the Royal Statistical Society. Series B, 34, pp. 187–220.

DA VANZO, J., RAZZAQUE, A., RAHMAN, M., HALE, L., AHMED, K., KHAN, M., A., MUSTAFA, G., GAUZIA, K., (2004). The Effects of Birth Spacing on Infant and Child Mortality, Pregnancy Outcomes, And Maternal Morbidity and Mortality in Matlab, Bangladesh. Rand Corporation: Rand Labor and Population Working Paper Series, WR-198 (October).

DA VANZO, J., STARBIRD, E. H., (1991). Correlates of Short Inter-birth Intervals in Peninsular Malaysia: Their Pathways of Influence through Breastfeeding and Contraceptive Use. Studies in Family Planning, Vol. 22, 4, pp. 241–54.

DAVANZO, J., HABICHT, J. P., (1986). Infant Mortality Decline in Malaysia, 1946-1975: The Roles of Changes in Variables and Changes in the Structure of Relationship, Demography, Vol. 23, 2, pp. 143–160.

DWIVEDI, L. K., RAM, F., RESHMI, R. S., (2007). An Approach to Understanding Change in Contraceptive Behaviour in India. GENUS, LXIII (3-4), pp. 19–54.

DWIVEDI, S. N., SINGH, R., (2003). On Assessing the Child Spacing Effect of Breastfeeding Using Cox Proportional Hazard Model with NFHS Data, Demography India, Vol. 32, 2, pp. 215–224.

FAMILY HEALTH INTERNATIONAL, (1988). Consensus Statement: Breastfeeding as a Family Planning Method. The Lancet, 8621, pp. 1204–1205.

HENRY, L (1961). Some Data on Natural Fertility. Eugenics Quarterly, 8 (2), pp. 81–91.

HOLMAN, D. J, GRIMES, M. A., ACHTERBERG, J. T., BRINDLE, E., O'CONNOR, K. A. (2006). Distribution of Postpartum Amenorrhea in Rural Bangladeshi Women. American Journal of Physical Anthropology, 129 (4), pp. 609–619.

HOWIE, P. W., MCNEILLY, A. S., (1982). Effect of Breastfeeding Patterns on Human Birth Intervals. Journal of Reproduction and Fertility, 65, pp. 545–557.

KLEINBAUM, D. G., (1996a). Survival Analysis. Statistics in Health Sciences. Springer-Verlag, New York, Inc.

KLIENBAUM, D. G., (1996b). Statistics in the Health Sciences. Springer, New York, Inc. USA.

LAST, J. M., (1988). A Dictionary of Epidemiology (2nd edition). New York Oxford, Toronto, Oxford University Press.

MARBANIANG, S. L., (2003). Health and Family Welfare Programme and Fertility Change in Meghalya. Unpublished PhD Thesis, International Institute for Population Sciences, Mumbai, India.

MERCHANT, K., MARTORELL, R., (1998). Frequent Reproductive Cycling: Does It Lead to Nutritional Depletion of Mothers? Progress in Food and Nutrition Science, Vol. 12, 4, pp. 339–369.

MTURI, A. J., (1997). The Determinants of Birth Intervals among Non-Contraception Tanzanian Women. African Population Studies, Vol. 12, 2 Accessed at http://bioline.utsc.utoronto.ca/archive/00000518/ on September 26, 2012.

NAMBOODIRI, N. K., (1974). Which Couples at Given Parties Expect to have Additional Births? An Exercise in Discriminant Analysis. Demography, Vol. 11, 1, pp. 45–56.

NAMBOODIRI, N. K., (1983). Sequential Fertility Decision Making and the Life Course. In: Bulatao, RA and Lee, RD (Eds.), Determinants of Fertility in Developing Countries, Fertility Regulation and Institutional Influences, New York, Academic Press, 2, Studies in Population, pp. 444–472.

NATH, D. C., LAND, K. C., SINGH, K. K., (1994). The Role of Breastfeeding beyond Postpartum Amenorrhoea on the Return of Fertility in India: A Life Table and Hazards Model Analysis. Journal of Biosocial Sciences, Vol. 26, 2, pp. 191–206.

NJOGU, W., MARTIN, T., C., (1991). Fertility Decline In Kenya: The Role of Training and Spacing of Births in IRD/Macro International, Proceedings of the Demographic and Health Surveys World Conference, Vol. III, Washington, D.C., Columbia, Maryland.

OJHA, A., (1998). The Effect of Sex Preference on Fertility in Selected States of India. The Journal of Family Welfare, Vol. 44, 1, pp. 42–48.

PATHAK, K. B., (1966). A Probability Distribution for the Number of Conceptions. Sankhya. Series B, Vol. 28 (3/4), pp. 213–218.

POTTER, J., E., (1987b). The Influence of Maternal Health Care on the Prevalence and Duration of Breastfeeding in Rural Mexico, Studies in Family Planning, Vol.18, 6, pp. 309–319.

POTTER, J. E., MOJARRO, O., NUNEZ, L., (1987a). The Influence of Health Care on Contraceptive Acceptance in Rural Mexico. Studies in Family Planning, Vol.18, 3, pp. 144–156.

POTTER, R. G., (1963). Birth Intervals: Structure and Change, Population Studies, Vol. 17, 2, pp. 155–166.

RAM, U., DWIVEDI, L. K., GOSWAMI, B., (2007). Understanding Contraception Use among Muslims of Indian, Pakistan and Bangladesh, Journal of Population and Social Studies, Vol.15, 2, pp. 101–130.

RODRIGUEZ, G, HOBCRAFT, J., MCDONALD, J., MENKEN, J., TRUSSELL, J., (1983). A Comparative Analysis of the Determinants of Birth Intervals. Comparative Studies, Number 30, London: World Fertility Survey.

SAHU, D., (1998). Breastfeeding Practices, Postpartum Amenorrhoea and Fertility Transition in Orissa. 1982-93: A Study Based on Two Large Scale Sample Surveys, International Institute for Population Sciences, Mumbai, India.

SAXENA, P. C., (1977). Breast-feeding: It's Effects on Post-partum Amenorrhea, Social Biology, 24 (1), pp. 45–51.

SAXENA, P. C., PATHAK, K. B., (1977). On a Distribution of Post-partum Amenorrhea Period Following a Live Birth, Demography India, 6 (1&2), pp. 174–181.

SEHGAL, J. M., (1971). Indices of Fertility Derived from Data on the Length of Birth Intervals. Using Different Ascertainment Plans. Unpublished PhD Thesis, School of Public Health, University of North Carolina, Chapel Hill, USA.

SHEKHAR, C., DWIVEDI, L., K., RAM, F., (2006). Fertility Goals and Regulation Dynamics in Northeastern States of India: An Application of Supply-Demand Framework. Journal of Empirical Research in Social Science, Vol. 1, 2, pp. 12–33.

SHEKHAR, C., (2004). Understanding Reproductive Changes and Proximate Determinants in India. Unpublished PhD Thesis, International Institute for Population Sciences, Mumbai, India.

SHEPS, M. C., (1964). On the Time Required for Conception. Population Studies, Vol. 18, pp. 85–97.

SHEPS, M. C., MENKEN, J. A., (1972). Distribution of Birth Intervals According to the Sampling Frame. Theoretical Population Biology, Vol. 3, 1, pp. 1–26.

SINGH, S. N., (1964). On the Time of First Birth. Sankhya. Series B, Vol. 26 (1/2), pp. 95–102.

SINGH, K. K., SUCHINDRAN, C. M., SINGH, K., (1999). Breast-feeding and Postpartum Amenorrhoea: An Indian Experience. Demography India, 28 (1), pp. 1–12

SRINIVASAN, K., (1980). Birth Interval Analysis in Fertility Surveys. WFS Scientific Reports, 7.

SRINIVASAN, K, PATHAK, K. B., PANDEY, A., (1989). Determinants of Breastfeeding and Postpartum Amenorrhoea in Orissa. Journal of Biosocial Science, 21 (3), pp. 365–371.

SWENSON, I., THANG, N. M., (1993). Determinants of Birth Intervals in Vietnam: A Hazard Model Analysis. Journal of Tropical Pediatrics, Vol. 39, 3, pp. 163–167.

TRUSSELL, J., MARTIN, L. G., FELDMAN, R., PALMORE, J. A., CONCEPCION, M., DATIN NOOR LAILY B. T., DATO' ABU BAKAR, (1985). Determinants of Birth-Interval Length in the Philippines. Malaysia and Indonesia: A Hazard Model Analysis. Demography, Vol. 22, 2, pp. 145–168.

APPENDIX

Table 1. Mean duration of succeeding birth interval (in months) and its 95% confidence interval (CI) estimates for India and its regions -2005-06.

Country/Regions	Mean	95% CI	
		Lower	Upper
India	42.0	41.8	42.2
North	41.5	40.9	42.1
Central	39.3	38.9	39.8
East	42.6	42.1	43.2
Northeast	42.7	42.2	43.1
West	44.3	43.6	45.0
South	45.8	45.0	46.5

Table 2. Mean duration (Me) of succeeding birth interval (in months) and its 95% confidence interval (CI) estimates with respect to combination of breastfeeding and postpartum amenorrhea (PPA) for India and its regions -2005-06.

Country/ Regions	PPASBF-State (i)			PPACNBF-State (ii)			PPANEBF-State (iii)		
	Me	95% CI		Me	95% CI		Me	95% CI	
		Lower	Upper		Lower	Upper		Lower	Upper
India	36.1[a]	34.3	38.0	29.8[e]	29.5	30.1	26.3[d]	25.2	27.4
North	32.9[a]	29.3	36.5	29.1[e]	28.2	30.0	25.7[d]	23.2	28.2
Central	40.7[d]	36.5	45.0	29.7[f]	29.1	30.3	24.6[e]	22.7	26.6
East	33.5[a]	30.7	36.3	31.1[e]	30.3	31.9	26.5[d]	24.3	28.7
Northeast	33.7[a]	30.7	36.7	29.6[d]	28.8	30.3	26.0[c]	23.6	28.3
West	35.0[a]‡	28.9	41.2	29.4[d]ns	28.2	30.6	31.0[a]ns	25.8	36.2
South	38.1[a]‡	31.8	44.4	29.7[c]	28.2	31.2	32.0[c]‡	26.7	37.2
India	57.1[b]	56.8	57.4	43.1[a]	42.9	43.4	36.2[c]	35.1	37.4
North	55.4[b]	54.0	56.7	42.8[a]	42.1	43.5	36.0[c]	33.3	38.8
Central	57.6[a]	57.1	58.1	40.0[c]	39.4	40.6	35.9[b]	33.7	38.1
East	57.5[b]	57.0	58.0	42.9[a]	42.1	43.6	34.1[c]	31.5	36.8
Northeast	57.0[b]	56.5	57.5	43.3[a]†	42.6	43.9	37.0[a]†	34.2	39.7
West	57.0[b]	55.7	58.3	46.9[c]‡	46.0	47.7	34.8[a]	30.7	38.9
South	55.0[b]	53.6	56.4	48.0[a]†	47.1	48.8	39.2[a]†	35.7	42.8

Note: Means without common superscript letters are significantly different within row, P<=0.05

(Log-rank test) except few cases then it is denoted by †, ‡ and ns.

† Significantly different from other value, P<=0.05.

‡ Significantly not different from other value, P<=0.05.

ns Significantly not different from other value, P<=0.05.

State (i): PPASBF denotes mothers of those index children who were currently in PPA and breastfeeding.

State (ii): PPACNBF denotes mothers of those index children who were currently in PPA but not breastfeeding.

State (iii): PPANEBF denotes mothers of those index children who were currently in PPA but never

breastfed.

State (iv): NPPASBF denotes mothers of those index children who were currently not in PPA but breastfeeding.

State (v): NPPACNB denotes mothers of those index children who were currently not in PPA and not breastfeeding.

State (vi): NPPANEB denotes mothers of those index children who were currently not in PPA and never breastfed.

Table 3. Cox Hazards model for birth spacing by selected characteristics for India-2005-06

Variables	Exp(β)	95% CI Lower	95% CI Upper
Region of residence			
South	1.00	-	-
North	1.19	1.09	1.30
Central	1.14	1.05	1.24
East	0.99	0.91	1.09
Northeast	1.03	0.94	1.13
West	1.02	0.92	1.12
Place of residence			
Urban	1.00	-	-
Rural	0.97	0.93	1.02
Religion			
Hindu	1.00	-	-
Muslim	1.04	0.99	1.10
Others	1.23	1.16	1.31
Mother's education			
Higher	1.00	-	-
Illiterate	1.01	0.91	1.13
Primary	0.95	0.85	1.07
Secondary	0.97	0.88	1.08
Standard of living			
Richest	1.00	-	-
Poorest	1.25	1.15	1.37
Poorer	1.29	1.19	1.40
Middle	1.29	1.20	1.39
Richer	1.20	1.12	1.29

Table 3. Cox Hazards model for birth spacing by selected characteristics for India-2005-06 (cont.)

Variables	Exp(β)	95% CI	
		Lower	Upper
Sex of index child			
Male	1.00	-	-
Female	1.08	1.04	1.12
Survival status of index child			
Alive	1.00	-	-
Dead	3.97	3.72	4.24
Mother's age at birth of index child (yrs.)			
20-24	1.00	-	-
< 20	1.10	1.04	1.15
25-29	0.63	0.60	0.67
30-34	0.42	0.38	0.45
35+	0.23	0.21	0.27
Current use of contraceptive			
Yes	1.00	-	-
No	1.08	1.03	1.13
Previous birth interval			
24-36 months	1.00	-	-
First birth	6.01	5.60	6.45
<24 months	1.16	1.10	1.23
>36 months	0.88	0.82	0.94
Number of surviving children			
Two	1.00	-	-
None	0.05	0.04	0.07
One	0.11	0.10	0.12
Three	2.90	2.73	3.07
Four and above	4.26	3.97	4.57
Maternal BMI			
>=18.5Kg/m^2	1.00	-	-
<18.5Kg/m^2	1.06	1.02	1.10
Missing	0.92	0.84	1.01
Breastfeeding (BF) and postpartum amenorrhea (PPA)			
Currently not BF and not in PPA	1.00	-	-
Currently BF and not in PPA	0.16	0.14	0.18
Never breastfed and not in PPA	1.31	1.20	1.43
Currently BF and in PPA	1.05	0.89	1.23
Never BF and in PPA	1.51	1.36	1.68
Currently not BF and in PPA	1.31	1.25	1.37

Table 4. Cox Hazards model for birth spacing by selected characteristics in the different regions of India-2005-06

Variable	North	Central	East	Northeast	West	South
Breastfeeding (BF) and postpartum amenorrhea (PPA)						
Currently not BF and not in PPA	1.00	1.00	1.00	1.00	1.00	1.00
Currently BF and not in PPA	0.23 (0.17 0.32)	0.09 (0.06 0.12)	0.13 (0.10 0.18)	0.18 (0.14 0.24)	0.24 (0.14 0.41)	0.38 (0.21 0.66)
Never breastfed and not in PPA	1.27 (1.04 1.55)	1.10 (0.93 1.29)	1.45 (1.18 1.78)	1.58 (1.28 1.94)	1.96 (1.42 2.70)	0.87 (0.61 1.24)
Currently BF and in PPA	1.03 (0.66 1.60)	0.79 (0.58 1.09)	1.23 (0.88 1.72)	1.23 (0.90 1.68)	1.22 (0.60 2.49)	1.12 (0.57 2.21)
Never BF and in PPA	1.40 (1.07 1.84)	1.74 (1.44 2.10)	1.67 (1.32 2.10)	1.49 (1.17 1.89)	1.03 (0.64 1.67)	0.73 (0.45 1.18)
Currently not BF and in PPA	1.23 (1.10 1.38)	1.24 (1.14 1.34)	1.45 (1.30 1.62)	1.41 (1.28 1.56)	1.25 (1.07 1.47)	1.12 (0.93 1.34)

Note: All the covariates mentioned in the Table 4 were controlled.

Table 5. Estimated probabilities of not having next live birth at specific months by selected characteristics for India -2005-06

Characteristics	Probability of not having next live birth at months						
	12	**18**	**24**	**30**	**36**	**42**	**48**
Average	0.98	0.90	0.77	0.64	0.54	0.46	0.40
Primary educated women[a]	0.98	0.90	0.77	0.65	0.55	0.47	0.41
Secondary educated women[b]	0.98	0.90	0.77	0.64	0.54	0.46	0.40
Survival of index child[c]	0.98	0.91	0.78	0.66	0.56	0.49	0.43
Current use of contraceptives[d]	0.98	0.91	0.78	0.65	0.55	0.48	0.41
In postpartum amenorrhea							
+Currently breastfeeding[e]	0.98	0.90	0.77	0.64	0.54	0.46	0.40
+Currently breastfeeding[f]	0.98	0.90	0.77	0.65	0.55	0.47	0.40
+Currently breastfeeding[g]	0.98	0.90	0.77	0.65	0.55	0.47	0.41
+Currently not breastfeeding[h]	0.98	0.90	0.77	0.64	0.54	0.46	0.40
Not in postpartum amenorrhea							
+Currently breastfeeding[e]	0.98	0.91	0.78	0.66	0.56	0.48	0.42
+Currently breastfeeding[f]	0.99	0.95	0.88	0.81	0.74	0.69	0.64
+Currently breastfeeding[g]	0.99	0.95	0.89	0.82	0.76	0.71	0.66
+Currently not breastfeeding[h]	0.98	0.90	0.77	0.64	0.54	0.46	0.40

Table 5. Estimated probabilities of not having next live birth at specific months by selected characteristics for India -2005-06 (cont.)

Characteristics	Probability of not having next live birth at months						
	12	18	24	30	36	42	48
Breastfeeding and postpartum amenorrhea (PPA) with survival of index child							
In PPA							
+Currently breastfeeding[e]	0.98	0.91	0.78	0.66	0.57	0.49	0.43
+Currently breastfeeding[f]	0.98	0.91	0.79	0.67	0.57	0.49	0.43
+Currently breastfeeding[g]	0.98	0.91	0.79	0.67	0.57	0.50	0.43
+Currently not breastfeeding[h]	0.98	0.91	0.78	0.66	0.57	0.49	0.43
Not in PPA							
+Currently breastfeeding[e]	0.98	0.91	0.80	0.68	0.59	0.51	0.45
+Currently breastfeeding[f]	0.99	0.96	0.89	0.82	0.76	0.71	0.67
+Currently breastfeeding[g]	0.99	0.96	0.90	0.83	0.78	0.73	0.68
+Currently not breastfeeding[h]	0.98	0.91	0.79	0.67	0.57	0.49	0.43
Breastfeeding and postpartum amenorrhea (PPA) with current use of contraceptives							
In PPA							
+Currently breastfeeding[e]	0.98	0.91	0.78	0.66	0.56	0.48	0.42
+Currently breastfeeding[f]	0.98	0.91	0.78	0.66	0.56	0.48	0.42
+Currently breastfeeding[g]	0.98	0.91	0.78	0.66	0.56	0.49	0.42
+Currently not breastfeeding[h]	0.98	0.91	0.78	0.66	0.56	0.48	0.42
Not in PPA							
+Currently breastfeeding[e]	0.98	0.91	0.79	0.67	0.58	0.50	0.44
+Currently breastfeeding[f]	0.99	0.95	0.89	0.82	0.76	0.70	0.66
+Currently breastfeeding[g]	0.99	0.96	0.89	0.83	0.77	0.72	0.68
+Currently not breastfeeding[h]	0.98	0.91	0.78	0.66	0.56	0.48	0.42

Note:

[a] Mothers of those index children who were illiterate considered as educated up to primary level.

[b] Mothers of those index children who were illiterate and has primary level education considered as educated up to secondary level.

[c] All dead children were considered as surviving.

[d] Mothers of those index children who were not using spacing method considered as [c] urrently using spacing method.

[e] Mothers of those index children who never breastfed considered as currently breastfeeding.

[f] Mothers of those index children who were currently not breastfeeding considered as currently breastfeeding.

[g] Mothers of those index children who never breastfed and were currently not breastfeeding considered as currently breastfeeding.

[h] Mothers of those index children who never breastfed considered as currently not breastfeeding.

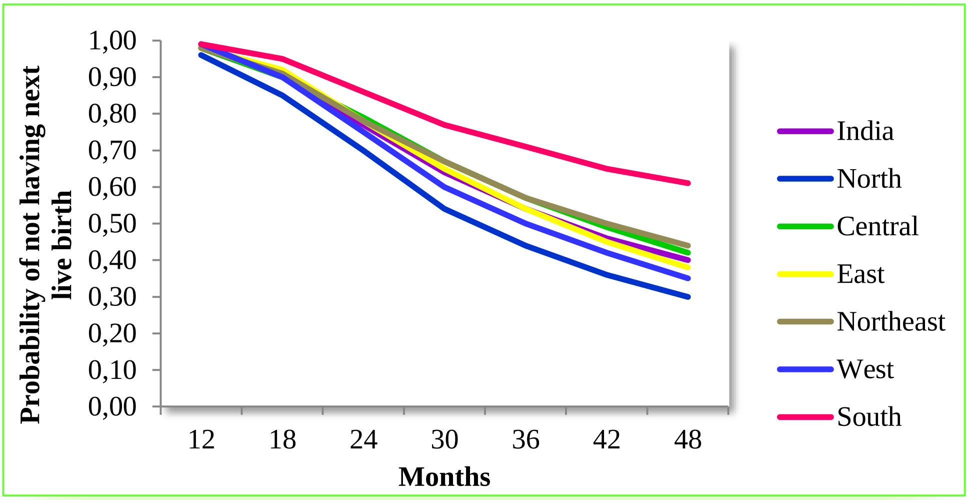


Figure 1. Estimated probabilities of not having next live birth at specific months in India and its regions-2005-06

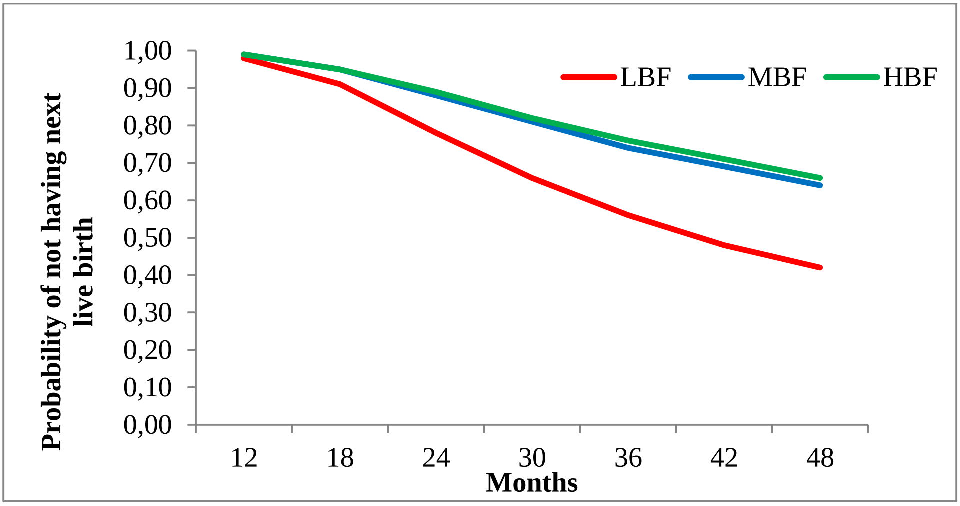


Figure 2. Estimated probabilities of not having next live birth at specific months by those women who were not in PPA by different levels of breastfeeding in India-2005-06

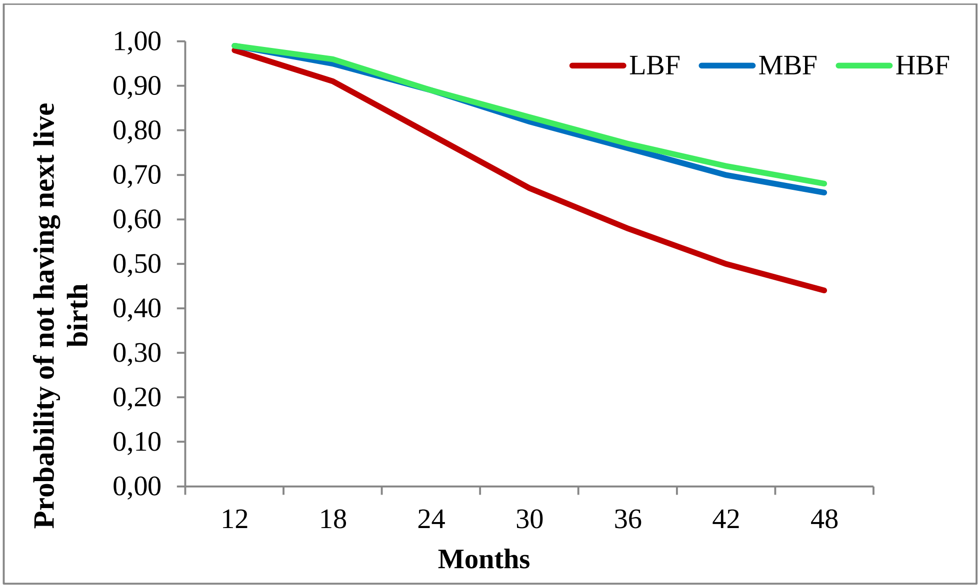


Figure 3. Estimated probabilities of not having next live birth at specific months by those women who were not in PPA and were currently using contraceptives by different levels of breastfeeding in India-2005-06

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 433–452, DOI 10.21307/stattrans-2018-024

DEALING WITH HETEROSKEDASTICITY WITHIN THE MODELING OF THE QUALITY OF LIFE OF OLDER PEOPLE

Katarzyna Jabłońska[1]

ABSTRACT

Using the estimation method of ordinary least squares leads to unreliable results in the case of heteroskedastic linear regression model. Other estimation methods are described, including weighted least squares, division of the sample and heteroskedasticity-consistent covariance matrix estimators, all of which can give estimators with better properties than ordinary least squares. The methods are presented giving the example of modelling quality of life of older people, based on a data set from the first wave of the COURAGE – Poland study. The comparison of estimators and their practical application may teach how to choose methodologically the most appropriate estimation tool after detection of heteroscedasticity.

Key words: heteroskedasticity, linear regression, HC-estimators, quality of life.

1. Introduction

Homogeneity of error variance, called homoskedasticity, is one of the main assumptions of linear regression. Many models, especially based on cross-sectional data, do not satisfy it (Greene, 2012, p.297). Such a situation is called heteroskedasticity. Then, parameters estimation with ordinary least squares method (OLS) does not give optimal results. There are many alternative methods which are either resistant to disturbance of homoskedasticity or they transform a model into a new one, which is henceforth homoskedastic.

Our aim is to discuss methods of parameters estimation in linear regression models in the case of heteroskedasticity and to focus on their strengths, weaknesses and important properties of obtained estimators. While weighted least squares method and the division of the sample are well known, HC-estimators are not commonly used, which may be surprising in light of the fact that new ones are still being created, improving previous ones. A comparison of those methods can be valuable for professional sociologists and practitioners to help them choose an appropriate estimation tool in the occurrence of heteroscedasticity in linear regression model.

[1] Department of Medical Sociology, Chair of Epidemiology and Preventive Medicine, Jagiellonian University Medical College, Kopernika 7a, 31-034 Krakow, Poland. Phone number: (+48) 12 423 1003. E-mail address: k.jablonska@cm-uj.krakow.pl.

Methodological considerations will be presented giving the example of modelling of quality of life of older people depending on psychosocial, demographical and other factors, based on data from the first wave of the COURAGE – Poland population-based study from 2011. Due to the population aging, the group of older people has been recently a subject of great interest and its analysis can be crucial in understanding how quality of life is affected by, in particular, psychosocial factors in an old age. We investigate a group without chronic diseases to detect what accompanies healthy aging. Analyses are done separately for a group of men for whom regression model is heteroskedastic and for a group of women for whom regression model is homoskedastic. The division into gender is justified because there are some significant gender-related differences in effects of psychosocial factors on older people's quality of life, as shown by Tobiasz-Adamczyk (et al., 2017).

In Section 2 properties of estimators from the ordinary least squares method are recalled. Alternative methods of parameters estimation are presented in Sections 3-6. The next section contains an empirical example of older people's quality of life models using previously described methods, separately for men and women. Results for both genders are discussed in Subsection 7.5. Section 8 includes conclusions and indications as to the proper choice of the method of estimation in the case of heteroskedasticity.

2. Linear regression model and the method of least squares estimation

In a classic linear regression model we have $Y = \beta X + \varepsilon$. Given a vector $Y = (y_1, \dots, y_n)$ of n-observations, called dependent variable, and a matrix $X = (x_1, .., x_p)$ with p-independent variables, where $\forall_{j=1,\dots,p}: x_j = (x_{j1}, \dots, x_{jn})$, we want to find the value of an unknown parameter $\beta = (b_0, b_1, .., b_p)$ to be able to predict values of Y with a random error ε, called residual.

Unknown parameters must be estimated, which means approximated in a sufficiently good way. Fortunately, there exist some objective measures of such sufficiency goodness: consistency, unbiasedness and effectiveness (the last is considered in a specified class of estimators). Consistency is a stochastic convergence to the estimated parameter; unbiasedness means that its expected value is equal to the value of the parameter which is estimated; effectiveness characterizes an estimator with the least variance within the specified class of estimators.

A classic way to estimate parameters is to use the method of ordinary least squares (OLS), where the value of the sum of squares of errors ($\sum_{i=1}^{n} \varepsilon_i^2$) is minimized. To make a model and its estimation purposeful, a number of assumptions are required. The first is homogeneity of error variance, called homoskedasticity. If, in addition, a model has a correct linear structure, a matrix X of fixed independent variables has rank p, the size of a sample is greater than the number of all parameters (n>p+1), random errors have mean zero and they are uncorrelated, then, on the basis of Gauss-Markov theorem (Dodge, 2008, p. 217-218), the OLS estimator will be linear, unbiased and effective among all linear and unbiased estimators. It will also be consistent (Verbeek, 2004). This estimator is expressed by the formula $\hat{\beta} = (X^T X)^{-1} X^T Y$. By the same theorem, the estimator of covariance matrix of the examined parameter, expressed by $\hat{V}(\hat{\beta}) = s^2 (X^T X)^{-1}$, where $s^2 = \frac{1}{n-(p+1)} \sum_{i=1}^{n} \varepsilon_i^2$, will also be unbiased. It gives important information about approximated standard errors of components of $\hat{\beta}$. If we additionally assume that errors are normally distributed, then significance tests (F test and Student's t-tests) will be possible to conduct. On this basis, one can determine which elements of vector $\hat{\beta}$, thereby, which independent variables, have significant relationship with variable Y.

If the assumption of homoskedasticity is violated, then we are talking about heteroskedasticity. To detect it, diagnostic tests should be conducted, for example three most popular: Breusch-Pagan test (Breusch and Pagan, 1979), which tests hypothesis that the error variance is linearly dependent with variables from the model; White's test (White, 1980), which finds out whether error variance is constant or Goldfeld-Quandt test (Goldfeld and Quandt, 1965), which checks whether heteroskedasticity is due to the one specified variable. Heteroskedasticity can also be detected with the help of OLS regression plots: errors and squared errors against predicted values, as well as errors against independent variables.

If we conclude that the analysed model is heteroskedastic, then OLS estimator of β is still consistent and unbiased, but no longer effective (Verbeek, 2004). Also, covariance matrix estimator $\hat{V}(\hat{\beta})$ is biased and inconsistent, and there is a problem with conducting statistical tests of the significance of parameters, because test statistics do not have required distributions (Verbeek, 2004). OLS estimation becomes unfounded in the case of heteroskedasticity, because there is a risk of both incorrect parameters approximation and untrustworthy tests results. Therefore, other methods of estimation should be used.

3. Weighted least squares method

A model with heteroskedasticity differs from a classic one in that consecutive observations have distinct values of error variance, that is $Var(\varepsilon) = \sigma^2 \Omega$ with different positive numbers $w_1, \dots, w_n$ (called weights) on the main diagonal of matrix Ω. Then, the sum $\sum_{i=1}^{n} \varepsilon_i^2 w_i$ is to be minimized. If Ω is known, estimator $\hat{\beta} = (X^T \Omega^{-1} X)^{-1} X^T (\Omega)^{-1} Y$ is effective among its unbiased estimators (Verbeek, 2004), covariance matrix for $\hat{\beta}$ equals

$$V(\hat{\beta}) = \sigma^2 (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}$$

and its unbiased estimator is

$$\hat{V}(\hat{\beta}) = s^2 (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}.$$

A method which uses weights is called weighted least squares method (WLS).

Knowledge of a matrix Ω is an unrealistic assumption and its values can only be approximated. We can use the model $\varepsilon^2 = \alpha X + c$ or eventually modify it to receive predicted values of ε^2, which we use as the diagonal of $\hat{\Omega}$. Then we can complete the main model of Y with weights equal to reciprocals of elements from the diagonal of $\widehat{\Omega}$. WLS estimator with the weights estimated before can be asymptotically more effective than classic OLS (Davidian, Carroll, 1987), assuming we used a proper and well-fitted model to predict ε^2. Otherwise, there is a big risk that the new model will still be heteroskedastic. It is one of the biggest WLS disadvantages, but this method has one important strength. It helps to detect presumptive cause of heteroskedasticity in a model. Assuming error squares regression was analysed and a variable significantly dependent on ε^2 was found, we can suspect that we detected the reason of the problem. It leads us to the next method of dealing with lack of homoskedasticity.

4. Division into subsamples

Having a variable significantly dependent on error squares, we can try to divide the sample into subsamples which depend on its values. It is obvious that we look for a division such that error variances among both subsamples are constant. Similar approach was considered by Goldfeld and Quandt (1965), and yield a test of heteroskedasticity based on the assumption that variances heterogeneity is due to the one specified variable.

It is crucial to select the division with a strong theoretical justification. Only then our original aim, which is drawing conclusions about the whole population based on its randomly selected part, will be preserved. It is much easier to isolate subsamples relying on a factorial variable than on a continuous one (in the latter case we have to arbitrary impose cut-off points). Other good idea is to divide the sample based on factorial variables like: gender, marital status, group of age and so on. Such divisions are almost always justified, but we should not forget to check homoskedasticity of new subsample models – the division had no sense without its occurrence.

5. Heteroskedasticity-consistent covariance matrix estimators

There is another estimation method which can be applied without any assumption about the error variance – heteroskedasticity-consistent covariance matrix estimators (HC-estimators). This method uses OLS to estimate β and one of HC-estimators to estimate its covariance matrix (then, standard errors) and to conduct tests. The main purpose of their use is to minimize the violation of inference caused by heteroskedasticity.

The first HC-estimator was HC0 proposed by White (1980), called a Sandwich estimator. Next were: HC1 by Hinkley (1977), HC2 by MacKinnon and White (1985), HC3 by Efron (1982), HC4 by Cribari-Neto (2004), HC5 by Cribari–Neto, Souza and Vasconcellos (2007), HC4m by Cribari-Neto and da Silva (2011) and there are still being created new ones, like the newest HC5m by Li et al. (2017), each improving previous ones. Formulas of all eight are given below.

$$\mathrm{HC0} = (X^T X)^{-1} \begin{pmatrix} \hat{\varepsilon}_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\varepsilon}_n^2 \end{pmatrix} (X^T X)^{-1},$$

$$\text{HC1} = \frac{n}{n-p-1}(X^TX)^{-1}\begin{pmatrix}\hat{\varepsilon}_1^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \hat{\varepsilon}_n^2\end{pmatrix}(X^TX)^{-1},$$

$HC2$ to $HC5m$ are equal to $(X^TX)^{-1}\begin{pmatrix}\frac{\hat{\varepsilon}_1^2}{(1-h_{11})^{\delta_1}} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{\hat{\varepsilon}_n^2}{(1-h_{nn})^{\delta_n}}\end{pmatrix}(X^TX)^{-1}$ with different values of δ_i $(i = 1, \dots, n)$ for each:

- $\text{HC2: } \delta_i = 1,$
- $\text{HC3: } \delta_i = 2,$
- $\text{HC4: } \delta_i = \min\left\{\frac{nh_{ii}}{p+1}, 4\right\},$
- $\text{HC4m: } \delta_i = \min\left\{\frac{nh_{ii}}{p+1}, \gamma_1\right\} + \min\left\{\frac{nh_{ii}}{p+1}, \gamma_2\right\},$
- $\text{HC5: } \delta_i = \min\left\{\frac{nh_{ii}}{p+1}, \max\left\{4, \frac{n\cdot k\cdot h_{max}}{p+1}\right\}\right\},$
- $\text{HC5m: } \delta_i = k_1\min\left\{\frac{h_{ii}}{\bar{h}}, \gamma_1\right\} + k_2\min\left\{\frac{h_{ii}}{\bar{h}}, \gamma_2\right\} + k_3\min\left\{\frac{h_{ii}}{\bar{h}}, \max\{4, \frac{kh_{max}}{\bar{h}}\}\right\},$

where:

$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - x_i\hat{\beta}_i$,

h_{ii} is i-th element of a matrix $H = X(X^TX)^{-1}X^T$ (i-th leverage),

$h_{max} = \max\{h_{ii} \mid i \in \{1, .., n\}\}$,

$\bar{h} = \frac{\sum_{i=1}^n h_{ii}}{n}$,

$k \in [0,1]$ (Cribari-Neto, Souza and Vasconcellos (2007) recommend k=0.7),

$\gamma_1, \gamma_2 > 0$ (Cribari-Neto and da Silva (2011) recommend $\gamma_1 = 1$, $\gamma_2 = 1.5$),

$k_1, k_2, k_3 \geq 0$ (Li (et al., 2017) recommend $k_1 = 1, k_2 = 0, k_3 = 1$).

The fact of the existence of so many HC-estimators inclines to make a comparison of their strengths and weaknesses, which will help to choose best estimators in specific situations.

HC1, in comparison with HC0, is corrected for degrees of freedom, whereas HC2 takes into account values of leverage, which has been improved by HC3 and later. All HC estimators do not demand homoskedasticity and are asymptotically consistent, nonetheless they have their own disadvantages. It is common for HC0 to become severely biased as mentioned by Cribari-Neto, Ferrari and Cordeiro (2000), especially in the case of a small sample (Long and Ervin, 2000) or occurrence of many high-leverage observations (Chesher and Jewitt, 1987).

Moreover, t-tests with HC0-estimator are liberal, which means that it is easy for them to achieve significant results, and similar can be said about HC1 and HC2.

In the case of a small sample, n<250, Long and Ervin (2000) recommend the use of HC3. For large samples HC0, HC1 and HC2 estimators should behave almost the same as HC3, but in the case of occurrence of high-leverage observations they become more biased than HC3. As it was shown by Cribari-Neto and Zarkos (2001), high-leverage observations can have even bigger influence on tests conservation, hence, their reliability, and estimator properties (primarily, their bias) than severity of heteroskedasticity. That is why HC4-estimator was proposed and, as it was shown by Cribari-Neto (2004), it has an advantage over HC3 in the case of many high-leverage observations.

The first HC4 modification, called HC5-estimator, was presented by Cribari–Neto, Souza and Vasconcellos (2007) and their innovation was taking into account the maximal leverage instead of only individual ones. Numerical evaluations showed that HC5-based inferences are much more reliable than HC3 and HC4-based: they are less size-distorted and tests are less liberal. HC5 can be crucial when observations are very strongly leveraged.

Another approach to HC4 modification was HC4m-estimator presented by Cribari-Neto and da Silva (2011). It improves the squared residuals discounting dependently on values of leverage: a heavier discount for low leverage observations and inversely for high leverage ones. HC4m is also a better alternative for HC4 in the case of non-normal errors (Cribari-Neto and da Silva, 2011). HC4m tries to fix HC4 and HC5 weaknesses in the case of a low degree of leverage, but it is worse than them when the high degree of leverage occurs.

The most recently presented HC5m-estimator by Li (et al., 2017), combines strengths of HC4m for low degree of leverages and HC5 for high degree of leverages. Simulations performed by Li (et al., 2017) showed that HC5m-based tests are reliable at points both with low or with high leverages and they have the smallest size distortions among tests based on all of HC3, HC4, HC4m and HC5.

It is worth pointing out that in the case of homoskedasticity HC-estimators can have worse properties than OLS estimator – almost all are biased then (Kauerman and Carroll, 2001). However, only HC2 is an unbiased estimator for homoskedastic data (Hayes and Cai, 2008), which gives it a supremacy in a situation when it is hard to confirm with certainty that there is a lack of variance homogeneity.

The last very important thing is that newer HC-estimators make significance tests less and less liberal. It means that a significant result received with a newer estimator is much more reliable than the one received with the older one, however, it is hard itself to achieve a significant result with the help of newest estimators. When there is no proven need to use a more conservative estimator, researchers can choose a bit less conservative one, provided that the model was analysed in detail and its use is fully justified. The newest does not always mean the best – it depends on model's properties.

6. Other methods

There are many other methods applied for heteroskedastic data. One of the most frequently mentioned in the literature is connected with transforming variables (Carroll and Ruppert, 1984), (Box and Cox, 1964), however, it can be very problematic in regard to interpreting its results (Sakia, 1992).

The other option is the general method of moments – GMM (Cragg, 1983), but as it was shown in (Kiviet and Feng, 2015), it has huge defects, and some modifications are proposed. Other ideas are General Linear Models, Penalized Least Squares Method (Wagener and Dette, 2012) or Residual Maximum Likelihood Estimation (Smyth, 2002).

The latest ideas concern Bayesian regression (Startz, 2017) and Generalized Least Squares based on machine-learning (Miller and Startz, 2017). They still need a deeper exploration, but give hope to accurate estimation without regard to the sample size and the values of leverages.

7. Empirical example of quality of life model

7.1. Statistical analysis

Linear regression will be used to model the assessment of quality of life of people aged 60 or over, who have not been diagnosed with any serious disease, including depression and chronic diseases (angina, arthritis, asthma, POCHP, diabetes and stroke), depending on: age, body mass index (BMI), assessment of activities of daily living (ADL) on Katz's scale described by Wieczorowska-Tobis and Talarska (2010), social network (Zawisza, Gałaś and Tobiasz-Adamczyk, 2014), loneliness (Hughes et al., 2004), social support (Dalgard, 1996) and two types of participation: relations with other people and activity in a local community. Models are additionally adjusted into education level, marital status and having children.

The data come from the first wave of the COURAGE - Poland population-based study from 2011. Values of quality of life are based on the Polish version of WHOQOL-AGE scale (Caballero et al., 2013; Zawisza, Gałaś and Tobiasz-Adamczyk, 2016), ranged from 0 to 100 points, and higher score of WHOQOL-AGE is interpreted as better health-related quality of life. Also loneliness, social support and social network range from 0 to 100 points. Higher score of ADL assessment means that more problems with daily living activities were reported.

We firstly consider a model for men. At the beginning, OLS estimation is conducted, but after detection of heteroskedasticity other methods are used. Finally, the most proper estimator is chosen. The same model for group of women is analysed with OLS and HC2-estimator, because both have good properties in the case of homoskedasticity, which was observed in this model. Then, a comparison of results for women and men is presented.

Analyses are conducted with SAS 9.4. To calculate standard errors and p values with HC4-estimator, we use a macro created by Hayes and Cai (2007) given in (Hayes and Cai, 2007, Appendix), while for HC4m, HC5 and HC5m we created new SAS macros, just like we did to conduct a Goldfeld-Quandt test. The adopted level of significance is 0.05.

7.2. Men's quality of life model

We conducted OLS estimation for the linear regression model of older men's quality of life with the sample size: n=366. Results are given in Table 1. Five psychosocial variables (ADL, social network, loneliness, participation - relations and social support) turned out to be significant with p values below 0.05. The model meets assumptions of errors normality with mean equals 0, significant linear structure of the model, no autocorrelation of errors, no correlation of independent variables, but it has a problem with homoskedasticity. Results of White's test (p=**0.001**) and Breusch-Pagan test (p=**0.009**) show that there is possibly a relationship between error variance and one or more independent variables.

Table 1. Results of OLS estimation of parameters from men's quality of life model, adjusted into: education level, marital status, having children

	OLS *(R^2=0.283, p<0.001)*		
	$\hat{\beta}$	Std. error	P
BMI	-0.095	0.133	.478
Age	-0.057	0.073	.434
ADL	-1.169	0.286	<.001
Social network	0.189	0.054	.001
Loneliness	-0.114	0.032	<.001
Participation - local community	0.580	0.765	.449
Participation - relations	1.654	0.777	.034
Social support	0.150	0.041	<.001

The scatter plot of predicted values of quality of life against error squares with a regression line (Figure 1) confirms that presumably error variance is not constant and seems to decrease with an increase in predicted quality of life. We need the other method of estimation.

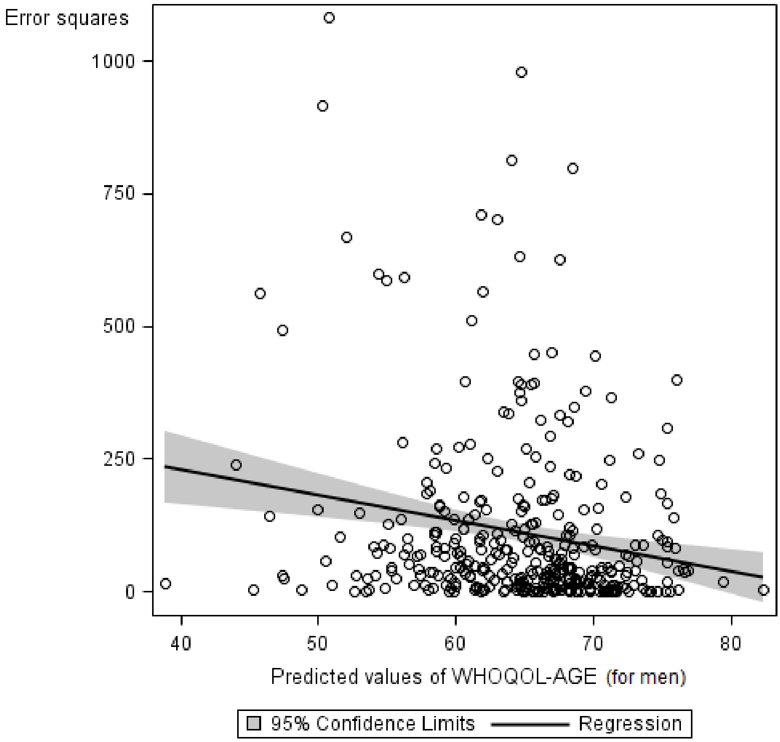


Figure 1. Regression line of error squares against predicted quality of life (WHOQOL-AGE scale) for men

Error squares regression with stepwise selection was conducted with all independent variables used in the quality of life model. It helped to choose their best linear model (Table 2), which turned out to be the one with social support (p=0.001) and loneliness (p=0.067). Their additive effect makes the whole model significant (R^2=0.044; p<0.001), but it has many failures, like non-normally distributed errors with non-zero mean.

Table 2. Results of the OLS estimation from the error squares model

	$\hat{\beta}$	Std. error	P
Social support	-1.950	0.566	.001
Loneliness	0.826	0.449	.067

One unit higher feeling of social support ($\hat{\beta} = -1.950;\ p = 0.001$) is associated with almost two units lower error variance and this variable can be suspected to cause heteroskedasticity in the quality of life model. Indeed, error variance is much higher for men who have the lowest level of social support (below its first quartile) compared to groups with higher levels (Table 3). Unfortunately, after adding weights to the basic quality of life model and conducting WLS estimation, a new model is still heteroskedastic. However, a strong relationship between error squares and social support could be seen and a natural idea is to divide the sample into subsamples relying on values of social support. We propose to create a dichotomous variable with value 1 if social support is below or equal to its first quartile (<=54.55) and value 0 for the other case.

Table 3. Error variances for social support quartile-based levels from quality of life OLS model

Social support	<=54.55 (Q1)	<=63.64 (Q2)	<=72.73 (Q3)	<=100
N	121	91	75	73
Error variance	144.12	102.51	89.65	83.99

To see whether such a division has a chance to be proper, let us conduct a Goldfeld-Quandt test. A test statistics is of the form

$$F[n_2 - (p+1), n_1 - (p+1)] = \frac{\frac{s_2}{n_2-(p+1)}}{\frac{s_1}{n_1-(p+1)}},$$

where p is the number of independent variables, n_i is an i-th subsample size, s_i^2 – i-th subsample sum of squares, where i=1 is for a subsample with relatively small error variance (higher social support in our case) and i=2 is for a subsample with relatively large error variance (lower social support). It has a chi-square distribution when the equality of variance of both subsample hypotheses is fulfilled, as followed by Goldfeld and Quandt (1965). In our case p=**0.010,** which means that error variances are significantly different among both subsamples and the division is reasonable. Quality of life model will be study separately for them.

For men with a low social support, Breusch-Pagan test indicates it still has a problem with heteroskedasticity (White: p=0.294; Breusch-Pagan: p=**0.043**) and the same can be told about White's test for the group with a higher social support (White: p=**0.019**; Breusch-Pagan: p=0.218). The division is not appropriate and we should try another one or consider a different method of estimation.

Results of estimation with HC-estimators for men's quality of life model are given in Table 4. The same five psychosocial variables as with OLS method turned out to be significant after estimations with HC0, HC1, HC2, HC3, HC4 and HC5. Standard errors are usually greater for HC than for OLS and the same can be told about p values. Participation – relations was just below 0.05 with HC3, HC4 and HC5, but crossed this line after HC4m and HC5m estimations. For HC5m also ADL assessment was not significant (p=0.067). HC4 and HC5 gave almost exactly the same results, which clearly indicates that the degree of leverage must not be very high in our model.

Table 4. Results of HC estimation of standard errors with p values from men's quality of life models, adjusted into: education level, marital status, having children

	HC0		*HC1*		*HC2*		*HC3*		*HC4*		*HC5*		*HC4m*		*HC5m*	
	Std. error	P	Std. error	P	Std. error	P	Std. error	P	Std. error	P	Std. error	P	Std. error	P	Std. error	P
BMI	0.133	.478	0.135	.485	0.137	.490	0.141	.503	0.144	.511	0.144	.511	0.143	.507	0.162	.560
Age	0.075	.448	0.076	.455	0.077	.458	0.078	.468	0.079	.469	0.079	.469	0.079	.471	0.085	.503
ADL	0.361	.001	0.367	.002	0.384	.003	0.409	.005	0.461	.012	0.461	.012	0.422	.006	0.636	.067
Social network	0.062	.003	0.063	.003	0.064	.004	0.066	.005	0.068	.006	0.068	.006	0.067	.005	0.076	.014
Loneliness	0.033	.001	0.034	.001	0.034	.001	0.035	.001	0.035	.001	0.035	.001	0.035	.001	0.039	.003
Participation-local community	0.682	.396	0.694	.404	0.697	.406	0.712	.415	0.708	.413	0.708	.413	0.714	.417	0.741	.434
Participation-relations	0.800	.039	0.813	.043	0.818	.044	0.837	.049	0.834	.048	0.834	.048	0.842	.050	0.877	.060
Social support	0.043	.001	0.044	.001	0.044	.001	0.045	.001	0.045	.001	0.044	.001	0.045	.001	0.046	.001

An observation is called an outlier if the absolute value of its studentized residual is greater than 2, whereas it is called a high-leveraged point if its leverage is greater than $\frac{2(p+1)}{n} \approx 0.0655$. Outliers and high leveraged observations are shown in Figure 2. There are 16 outliers (4%), 20 high-leverage observations (5%) and 3 points with both these features (0.8%). In our case, $h_{max} = 0.162$ and $\frac{n \cdot 0.7 \cdot h_{max}}{p+1} < 4$, which means that the degree of maximum leverage is moderate and not very high.

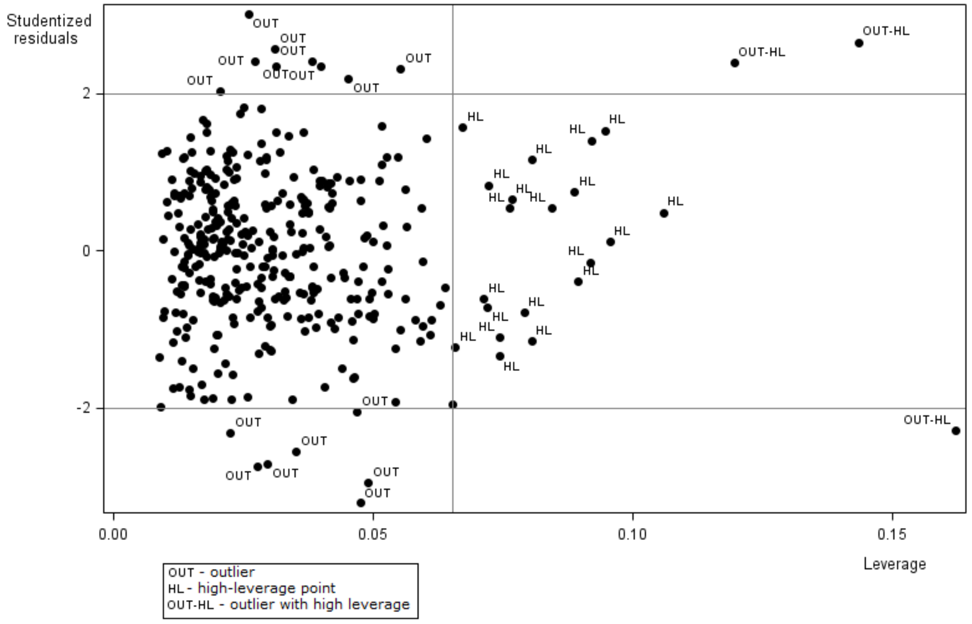


Figure 2. Scatter plot of leverages towards studentized residuals

The questions is which of HC-estimators is the most suitable in our case. We have a strong evidence of heteroskedasticity, the sample size is not very small and the maximum degree of leverage is moderate. HC4m, HC5m or even HC3 are better adapted for the situation of low or moderate leverages than HC4 and HC5, which are better in the case of very high leverages, so we can restrict to those 3 estimators. As it was outlined earlier in Section 5, the conservativeness of tests increases with every consecutive estimator. HC5m makes them severely conservative, but there is no need to use it for our model. In our opinion, the best option is to choose HC4m, which is still much more conservative than somewhat liberal HC3. As a result of use of HC4m-estimator, four variables (ADL assessment, social network, loneliness and social support) are considered significant in the context of men's quality of life.

7.3. Women's quality of life model

We will now consider the same model of quality of life for the group of women aged 60 or over (sample size: n=519), which meets all linear regression assumptions including homoskedasticity (White: p=**0.452**; Breusch-Pagan: p=**0.590**). The scatter plot of predicted values of quality of life against error squares with a regression line (Figure 3) confirms that error variance is rather constant.

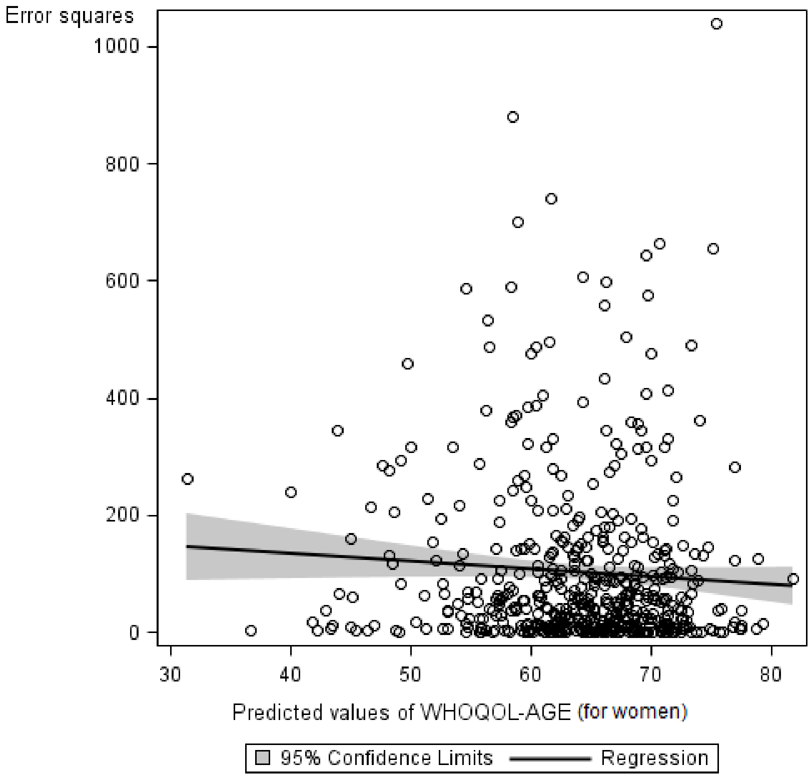


Figure 3. Regression line of error squares against predicted quality of life (WHOQOL-AGE scale) for women

Table 5 presents the results of OLS estimation of women's quality of life. There are four significant variables: ADL, social network, loneliness and social support. In the light of facts previously presented in Section 5, it might be valuable to compare OLS results with an unbiased, but a slightly more conservative HC2-estimator. Despite both OLS and HC2 variance estimators are unbiased, values of standard errors estimators and results of significance tests are not equally same, but they are very similar. This is mainly due to the fact that in both cases the variances are estimated and square roots are computed to finally get standard errors. While p value for participation–local community is equal to the level of 0.05 for HC2-estimator, it is much higher for OLS (p=0.071) and we consider this variable insignificant. Despite HC-estimators almost always increases the conservativeness of tests, it is not a set rule for all variables in a given model.

Table 5. Results of estimation with OLS and HC2-estimator of women's quality of life, adjusted into: education level, marital status and having children

	OLS *(R^2=0.343, p<0.001)*			*HC2*	
	$\hat{\beta}$	Std. error	P	Std. error	P
BMI	-0.020	0.086	.819	0.084	.813
Age	-0.114	0.063	.071	0.064	.078
ADL	-1.121	0.157	<.001	0.174	<.001
Social network	0.128	0.040	.002	0.040	.001
Loneliness	-0.076	0.023	.001	0.025	.003
Participation – local community	1.274	0.704	.071	0.648	.050
Participation - relations	0.887	0.586	.131	0.596	.137
Social support	0.112	0.031	<.001	0.032	.001

7.4. Discussion on the results from quality of life models

Men's quality of life model turned out to have a problem with heteroskedasticity. OLS estimation method was unreliable, therefore we tried methods of WLS and division into subsamples, both of which did not result in homoskedasticity. Then, HC-estimators were used and HC4m was found the most appropriate. Thanks to the estimation with HC4m four variables are considered to have a significant relationship with quality of life among older men: social network, loneliness, social support and ADL assessment.

Men who feel more lonely have worse quality of life assessment ($\hat{\beta} = -0.114;\ p = 0.001$), which means that one unit increase in the feeling of loneliness results in 0.114 unit decrease of quality of life. The more problems connected with daily living activities were reported, the significantly lower quality of life was detected ($\hat{\beta} = -1.169;\ p = 0.006$). In turn, the better the assessment of social network, the higher quality of life was reported ($\hat{\beta} = 0.189;\ p = 0.005$), as well as for social support ($\hat{\beta} = 0.150;\ p = 0.001$).

For women, the applicability of the OLS method was fully justified. As a result, the same four variables turned out to have a significant relationship with quality of life: ADL ($\hat{\beta} = -1.121;\ p < 0.001$), social network ($\hat{\beta} = 0.128;\ p = 0.002$), social support ($\hat{\beta} = 0.112;\ p < 0.001$) and loneliness ($\hat{\beta} = -0.076;\ p = 0.001$). Results of estimation with unbiased HC2-estimator were very similar to OLS results.

As it can be seen, the same four variables turned out to be significant in the context of quality of life both for men and women and trends for all four are very similar. Nonetheless, absolute values of all estimated parameters are higher in the case of men. Especially feeling of loneliness has 1.5 times stronger effect on quality of life among men than among women. While it has been shown before that QOL can be significantly reduced by loneliness among older people (Musich et al., 2015), there are very few studies on gender differences in older people quality of life affected by feeling of loneliness (Tobiasz-Adamczyk et al., 2017). Despite women seem to suffer from loneliness more frequently than men (Beal, 2006; I. Thomopoulou, D. Thomopoulou and Koutsouki, 2010), men's reaction to a higher feeling of loneliness, resulting in poorer quality of life assessment, can be stronger than women's. Deeper sociological inferences were not the main scope of this paper, but we hope that our results will encourage sociologists to make further analysis.

8. Conclusion

Giving the example of men's quality of life model, we could observe what are possible consequences of ignoring heteroskedasticity. We could also investigate how to choose the best alternative method of estimation.

Following Hayes and Cai (2007), if there is rationale for stating that our model does not meet the homoskedasticity assumption, it is recommended to have a very critical view on the results obtained by the OLS and to use other estimation methods. However, not all of them work equally well. WLS requires the form of heteroskedasticity to be known, which is usually difficult to meet and we are not assured that a new model will be free of lack of homoskedasticity. The latter can also be said about the division of a sample into subsamples.

It seems that the best idea is to use heteroskedasticity-consistent variance matrix estimators. Some HC-estimators are offered by statistical programs, for example SAS (HC0-HC3) and R (it offers also HC4m and HC5 in the *Sandwich* package presented in 2017).

Despite HC-estimators asymptotically have desirable properties, some problems with their credibility can occur. Generally, when it comes to choose the best HC-estimator, we recommend the following criteria:

- HC2 is the best option when heteroskedasticity is not clear or it is relatively low.
- If the sample size is small, HC3 and later estimators are preferred. The degree of leverages should be taken into account to choose the best among them.
- In the case of the occurrence of some high leveraged observations, HC3, HC4, HC5 or HC5m are recommended. If the degree of leverage is very high, HC5 and HC5m prevail over others.
- If the degree of leverage is low or moderate $\left(\frac{0.7 \cdot n \cdot h_{max}}{p+1} < 4\right)$, the best choices are HC4m or HC5m.
- In the occurrence of non-normal errors, use of HC4m is preferable.

Since we now have the knowledge and such an easy access to more precise tools, we should use HC-estimators in the occurrence of heteroskedasticity to verify the significance of model parameters and to analyse them in depth.

Acknowledgments

This work was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant number 223071 (COURAGE in Europe) and Polish Ministry for Science and Higher Education under Grant number 1277/7PR/ UE/2009/7, 2009-2012.

REFERENCES

BEAL, C., (2006). Loneliness in older women: a review of the literature. Issues in Mental Health Nursing, 27 (7), pp. 795–813.

BOX, G.E.P., COX, D. R., (1964). An Analysis of Transformations. Journal of the Royal Statistical Society, Series B (26), pp. 211–252.

BREUSCH, T. S., PAGAN, A. R., (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation, Econometrica, 47 (5), pp. 1287–1294.

CABALLERO, F. F., MIRET, M., POWER, M., CHATTERJI, S., TOBIASZ-ADAMCZYK, B., KOSKINEN, S., AYUSO-MATEOS, J. L., (2013). Validation of an instrument to evaluate quality of life in the aging population: WHOQOL-AGE. Health Qual Life Outcomes, 11 (177).

CARROLL, R. J., RUPPERT, D., (1984). Power Transformations When Fitting Theoretical Models to Data. Journal of the American Statistical Association, 79 (386).

CHESHER, A., JEWITT, I., (1987). The bias of a heteroskedasticity consistent covariance matrix estimator. Econometrica, 55 (5), pp. 1217–1222.

CRAGG, J. G., (1983). More efficient estimation in the presence of heteroscedasticity of unknown form. Econometrica, 51 (3), pp. 751–763.

CRIBARI-NETO, F., (2004). Asymptotic inference under heteroskedasticity of unknown form. Computational Statistics & Data Analysis, 45, pp. 215–233.

CRIBARI-NETO, F., DA SILVA, W. B., (2011). A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model. Advances in Statistical Analysis, 95, pp. 129–146.

CRIBARI-NETO, F., FERRARI, S. L. P., CORDEIRO, H. M., (2000). Improved heteroscedasticity-consistent covariance matrix estimators. Biometrika, 87 (4), pp. 907–918.

CRIBARI–NETO, F., SOUZA, T. C., VASCONCELLOS, K. L. P., (2007). Inference Under Heteroskedasticity and Leveraged Data. Communications in Statistics—Theory and Methods, 36, pp. 1877–1888.

CRIBARI-NETO, F., ZARKOS, S. P., (2001). Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. Journal of Statistical Computation and Simulation, 68, pp. 391–411.

DALGARD, O. S., (1996). Community health profile: a tool for psychiatric prevention, in Promotion of Mental Health. D. R. Trent, C.A. Reed (eds). Aldershot: Avebury Press, pp. 681–695.

DAVIDIAN, M., CARROLL, R. J., (1987). Variance Function Estimation. Journal of the American Statistical Association. 82 (400), pp. 1079-1091.

DODGE, Y., (2008). Gauss-Markov Theorem, in The Concise Encyclopedia of Statistics. Springer New York, pp. 217–218.

EFRON, B., (1982). The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia.

GOLDFELD, S. M., QUANDT, R. E., (1965). Some Tests for Homoscedasticity. Journal of the American Statistical Association, 6 (310), pp. 539–547.

GREENE, W. H., (2012). Econometric Analysis (7th Edition). Pearson Education Limited.

HAYES, A. F. CAI, L., (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. Behavior Research Methods, 39 (4), pp. 709–722.

HAYES, A. F., CAI, L., (2008). A New Test of Linear Hypotheses in OLS Regression Under Heteroscedasticity of Unknown Form. Journal of Educational and Behavioral Statistics, 33 (1), pp. 21–40.

HINKLEY, D. V., (1977). Jackknifing In Unbalanced Situations. Technometrics, 19 (3).

HUGHES, M. E., WAITE, L. J., HAWKLEY, L. C., CACIOPPO, J. T., (2004). A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies. Research on aging, 26 (6), pp. 655–672.

KAUERMAN, G., CARROLL, R. J., (2001). A note on the efficiency of sandwich covariance matrix estimation. Journal of the American Statistical Association, 96, pp. 1387–1396.

KOJIMA, M., KOJIMA, T., ISHIGURO, N., OGUCHI, T., OBA, M., TSUCHIYA, H., TOKUDOME, S., (2009). Psychosocial factors, disease status, and quality of life in patients with rheumatoid arthritis. Journal of Psychosomatic Research, 67, pp. 425–431.

KIVIET, J.F., FENG, Q., (2015). Efficiency Gains by Modifying GMM Estimation in Linear Models under Heteroskedasticity. Econometrics Working Papers, 14 (06).

LI, S., ZHANG, N., ZHANG, X., WANG, G., (2017). A new heteroskedasticity-consistent covariance matrix estimator and inference under heteroscedasticity. Journal Of Statistical Computation And Simulation, 87 (1) pp. 198–210.

LONG, J. S., ERVIN, L.H., (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. The American Statistician, 54, pp. 217–224.

MACKINNON, J. G., WHITE, H., (1985). Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. Queen’s Economics Department Working Paper, 537.

MILLER, S., STARTZ, R., (2017). Feasible Generalized Least Squares Using Machine Learning. Social Science Research Network [online] (6th February 2018). Available at: <https://ssrn.com/abstract=2966194> [Accessed on 27 March 2018].

MUSICH, S., WANG, S. S., HAWKINS, K., YEH, C. S., (2015). The Impact of Loneliness on Quality of Life and Patient Satisfaction Among Older. Sicker Adults. Gerontology & Geriatric Medicine, 1.

SAKIA, R. M., (1992). The Box-Cox transformation technique: a review. The Statistician, 41, pp. 169–178.

SMYTH, G. K., (2002). An Efficient Algorithm for REML in Heteroscedastic Regression. Journal of Graphical and Computational Statistics, 11, pp. 836–847.

STARTZ, R., (2017). Bayesian Heteroskedasticity-Robust Regression, 2015. The Economics Department of UCSB, [online] (Revised February 2015). Available

at: <http://econ.ucsb.edu/~startz/Bayesian%20Heteroskedasticity-Robust%20Regression.pdf> [Accessed on 27 March 2018].

THOMOPOULOU, I., THOMOPOULOU, D., KOUTSOUKI, D., (2010). The differences at quality of life and loneliness between elderly people. Biology of Exercise, 6 (2), pp. 13–28.

TOBIASZ-ADAMCZYK, B., GALAS, A., ZAWISZA, K., CHATTERJI, S., HARO, J. M., AYUSO-MATEOS, J. L., LEONARDI, M., (2017). Gender-related differences in the multi-pathway effect of social determinants on quality of life in older age—the COURAGE in Europe project. Quality of Life Research, 26 (7), pp. 1865–1878.

VERBEEK, M., (2004). A Guide to Modern Econometrics (2nd edition), John Wiley & Sons Ltd.

VILHENA, E., PAIS-RIBEIRO, J., SILVA, I., PEDRO, L., MENESES, R., CARDOSO, H., MENDONÇA, D., (2014). Psychosocial factors as predictors of quality of life in chronic Portuguese patients. Health and Quality of Life Outcomes, 12 (3).

WAGENER, J., DETTE, H., (2012). Bridge Estimators and the Adaptive Lasso under Heteroscedasticity. Mathematical Methods of Statistics, 21 (2), pp. 109–126.

WHITE, H., (1980). A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica, 48 (4), pp. 817–838.

WIECZOROWSKA-TOBIS, K., TALARSKA, D., (2010). Geriatria i pielęgniarstwo geriatryczne [Geriatrics and geriatric nursing]. Wydawnictwo Lekarskie PZWL, Warszawa.

ZAWISZA, K., GAŁAŚ, A., TOBIASZ-ADAMCZYK, B., (2014). Polska wersja Courage Social Network Index — skali do oceny poziomu sieci społecznych [Polish version of The Courage Social Network Index — the scale for the assessment of social networks]. Gerontologia Polska, 22 (1), pp. 31–41.

ZAWISZA, K., GAŁAŚ, A., TOBIASZ-ADAMCZYK, B., (2016). Walidacja polskiej wersji skali oceny jakości życia WHOQOL-AGE w populacji osób starszych [Validation of the Polish version of the WHOQOL-AGE scale in older population]. Gerontologia Polska, 1, pp. 7–16.

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 453–475, DOI 10.21307/stattrans-2018-025

SPATIAL MEASURES OF DEVELOPMENT IN EVALUATING THE DEMOGRAPHIC POTENTIAL OF POLISH COUNTIES

Anna Majdzińska[1]

ABSTRACT

The study presents a demographic potential-based typology of Polish counties created using the spatial measures of development (available in the literature and proposed by the author). The typology reveals the location of Polish areas (regions) with the highest and lowest demographic potential.The analysis was performed on data sourced from the publications of the Central Statistical Office (GUS) on the age structure and selected developments in the natural movement and migrations in counties in the years 2005 and 2016.

Key words: synthetic measures of development, spatial measures of development, spatial autocorrelation, demographic potential.

1. Introduction

This article deals with the concept of spatial measures of development. Following an overview of the different approaches to their construction discussed in the literature, the measures' modifications proposed by the author are presented. To explain their practical use, a demographic potential-based typology of Polish counties is created.

The notion of demographic potential still awaits an unambiguous definition. Among the existing ones, one describes it as the "current weight of the population, its potential ability to grow under appropriate external conditions" (Ediev, 2001, p. 291). Demographic potential is generally understood in terms of the population's size, structure and events driving demographic processes. In this article, it is meant as the demographic equivalent of human capital[2] the qualitative and quantitative interpretation of which is determined by the size and age structure of a county's population, and the structure's changes caused by fertility and mortality variations and migrations. The knowledge of the demographic potential of an area (a region or a country) has practical importance as it provides

---

[1] University of Lodz, Unit of Demography and Social Gerontology. E-mail: anna.majdzinska@uni.lodz.pl.

[2] Human capital has many definitions (see, for instance, CIPD, 2017, pp. 5-6; Dawid, 2001, p. 22; Goldin 2016, p. 56; Kotarski, 2013, pp. 10-16; Roszkowska, 2013, pp. 11-16), most of which refer to demographic characteristics (age structure, fertility, mortality rates, etc.) and social characteristics (level of education, occupational structure, stock of knowledge, etc.) (Giza-Poleszczuk and Marody, 2000). One attempt at quantifying the global stock of human capital is *The Human Capital Report 2015*.

a point of reference for adjustments to ongoing population policies, helps design new ones, prompts the intensity and direction of the necessary actions, and helps quantify the likely impacts of changes in population size and structure.

The literature offers a range of methods for evaluating and measuring demographic potential, such as a descriptive approach using several demographic variables to compare the demographic potential of territorial units (see, for instance, Pastuszka, 2017; Obrębalski, 2017), or a synthetic approach involving the construction of measures and models (e.g. Ediev, 2001; Majdzińska, 2016; Scarpaci, 1984).

It is interesting to observe that many studies analyse variables characterising the demographic potential of an area (population size, age structure, reproduction rate, etc.) without referring to “demographic potential”. Most of them compare the level of some phenomenon (e.g. a demographic situation) with areas using single variables such as the mortality rate or total fertility rate.

The data on the populations’ age structure and selected developments related to natural movement and migrations in the years 2005 and 2016 were sourced from the publications of the Central Statistical Office (GUS)[3].

2. Methodology

The study presents three approaches to constructing a spatial measure of development. The first of them refers to the concept proposed by E. Antczak (2013). The other two approaches have been derived by the author from the first one by modifying the construction of the matrix of spatial weights.

The spatial measures of development derive from the synthetic measures that take account of spatial associations between objects and the existence of spatial autocorrelation defined as “the concentration of similar values, correlations or interactions between variables relating to the geographical location of objects” (Suchecki and Olejnik 2010, pp. 102–105; see also Griffith, 2003, pp.3).

2.1. Spatial autocorrelation

Spatial autocorrelation is therefore defined as the degree of “correlation of the values of the observed variable between two areas”, meaning that the values “determine and are determined by [the variable’s] realizations in other areas” (Suchecki and Olejnik, 2010, pp. 102–105)[4].

Positive spatial autocorrelations are represented by “the contiguity of high or low values of the observed variables” and negative spatial autocorrelations by „low values […] being contiguous to high values and vice versa”.

Spatial autocorrelation can be considered in global or local terms (Suchecki and Olejnik, 2010, pp. 107–109, 112–113), which respectively denote “a spatial correlation of the analysed variable across an area” or “the dependency of the analysed variable […] on its surroundings, i.e. on its values in the contiguous areas”.

[3] The numbers of counties in 2005 and 2016 are different, because in the years 2003-2012 the town of Wałbrzych was part of Wałbrzyski powiat.

[4] See also: Getis (2010, p. 256) and Griffith (2003, p. 3).

A widely used measure of the global autocorrelation is Moran's *I* statistics, which allows testing a null hypothesis that „a spatial autocorrelation is not statistically significant" (meaning that the values of the analysed variable are randomly distributed in space) against an alternative hypothesis stating otherwise. Moran's *I* statistics for a row-standardised weight matrix is written as:

$$I = \frac{\sum_{i=1}^{m}\sum_{k=1}^{m} w_{ik}(x_i-\bar{x})(x_k-\bar{x})}{\sum_{i=1}^{m}(x_i-\bar{x})^2} = \frac{z^T W z}{z^T z} \tag{1}$$

where: x_i , x_k - values of variable *X* in the *i*-th and *k*-th object;

w_{ik} - the spatial weight between objects *i* and *k*,

$\boldsymbol{z} = [z_1, z_2, \dots, z_m]^T$ with $z_i = x_i - \bar{x}$,

$\boldsymbol{W} = [w_{ik}]$ - the $m \times m$ matrix of spatial weights w_{ik}.

Moran's *I* statistics is interpreted similarly to the Pearson's linear correlation coefficient (Pearson's product moment – see Johnson, 1984, pp. 90-91).

A useful measure of local autocorrelation is local statistics (*LISA*) identifying "the statistically significant clusters of similar values in contiguous areas" (Suchecki and Olejnik, 2010, pp. 120–125). The most popular method within *LISA* is local Moran's I_i, i.e. an *i*-th component of Moran's *I* statistics (1), given by the following formula:

$$I_i = \frac{(x_i-\bar{x})\sum_{k=1}^{m} w_{ik}(x_k-\bar{x})}{\sum_{i=1}^{m}(x_i-\bar{x})^2} = \frac{z_i\sum_{k=1}^{m} w_{ik} z_k}{\sum_{i=1}^{m} z_i^2} \tag{2}$$

2.2.Synthetic measures of development

The measure proposed by Sokołowski and Zając (Gazińska, 2003, pp. 203-204; Sokołowski and Zając, 1987, pp. 41-42) is constructed by first transforming variables into stimulants (or destimulants) and then by normalising them (a standardisation formula is usually employed to this end). Thereafter, the negative values of the normalised variables are converted to non-negative ones using the formula:

$$u_{ij} = x_{ij} - \min_i\{x_{ij}\} \tag{3}$$

where: x_{ij} - the value of the *j*-th characteristic of the *i*-th object.

In the next step, after the maximum values of transformed variables u_{ij} (characterising the reference model) have been determined, the measure of development is calculated as:

$$z_i = \frac{\sum_{j=1}^{p} u_{ij}}{\sum_{j=1}^{p} \max_j\{u_{ij}\}} \tag{4}$$

The measure takes values in the interval [0; 1]. Its highest value is attributed to the reference model (i.e. the unit where the level of the investigated phenomenon is the most favourable).

In the non-reference method of sums, the measure of development (Panek, 2009, p. 67 as quoted in Malina and Wanat) is constructed by transforming variables into stimulants and calculating their arithmetic average u_i^* for each object. The negative values are removed using the following formula:

$$u_i^{**} = u_i^* - \min_i \{u_i^*\} \tag{5}$$

To make sure that the measure only takes values in the interval [0, 1], it is normalised using the unitarisation formula:

$$z_i = u_i^{**} - \frac{u_i^{**}}{\max_i \{u_i^{**}\}} \tag{6}$$

To construct Hellwig's synthetic measure of development (see Hellwig, 1968; Nermend, 2009, pp. 37-44), first variables are transformed into stimulants (or destimulants) and then they are normalised (in most cases a standardisation formula is used to this end). The measure utilizes the so-called reference unit z_o, for which all stimulant variables from all investigated objects have maximum values (or minimum values when destimulant variables are used), i.e. (see Nowak, 1990, p. 88):

$$z_{oj} = \max_i \{z_{ij}\} \quad \text{(or } z_{oj} = \min_i \{z_{ij}\}\text{)} \qquad j = 1, ..., p \tag{7}$$

Subsequently, the distance between each object and the reference unit (z_{oj}) is determined using the following formulas:

$$d_i = \sqrt{\sum_{j=1}^{p} (z_{ij} - z_{oj})^2}\,, \qquad i = 1, 2, \dots, m \tag{8}$$

Further, to make sure that the relative measure takes values in the interval [0,1] it must be transformed according to the following formula (see Nowak, 1990, pp. 88-89):

$$z_i = 1 - \frac{d_i}{d_0} \quad \text{for} \quad i = 1, 2, \dots, m \tag{9}$$

where: $$d_0 = \bar{d} + 2s_d \tag{10}$$

d_0 - the basis of normalisation

$\bar{d}$ - an average calculated from d_i of all objects

s_d - standard deviation derived from d_i of all objects

The measure tends to take values in the interval [0; 1], but if the level of development or the situation of a specific object is by far worse than noted for other objects its measure may take a negative value. In order to eliminate negative values of measure z_i, formula (10) is modified by adopting three standard deviations (see Nowak, 1990, p. 89). The highest values of z_i are assigned to the reference model (i.e. the unit where the level of the investigated phenomenon is the most favourable).

2.3. Spatial transformation of Indicators

The spatial measure of development as proposed in this study draws on the E. Antczak's concept (2013, pp. 39–43) referring to Hellwig's measure of development (see Hellwig, 1968)[5] and spatial autocorrelation. It is similar to it in that it also allows for autocorrelation, but differs in the underlying synthetic measure.

The methodological difference between the spatial measures of development and the synthetic (non-spatial) measures lies in the construction of the input matrix of variables $\boldsymbol{X}$. Namely, to be accepted as diagnostic variables, the values of territorial units' characteristics must vary considerably (a coefficient of variance > 0.1), be relatively weakly correlated, and show statistically significant global spatial autocorrelation. A matrix of such variables is multiplied by a matrix of spatial weights $\boldsymbol{W}$ (based on a contiguity matrix)[6] according to the following formula proposed by E. Antczak (Antczak, 2013, pp. 39-43):

$$\boldsymbol{X}^*_{m\times p} = \boldsymbol{W}_{m\times m}\boldsymbol{X}_{m\times p}, \tag{11}$$

where:

$\boldsymbol{W} = [w_{ik}]$ – the $m \times m$ matrix of spatial weights,

$\boldsymbol{X} = [x_{ij}]$ – the $m \times p$ data matrix,

m – the number of territorial units,

p – the number of diagnostic variables,

$$w_{ik} = \begin{cases} \frac{1}{m_i} & \text{when unit } k \text{ is contiguoust to unit } i, \quad i \neq k \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

m_i – the number of territorial units constituting the i-th region, i.e. units contiguous to an i-th unit.

The inclusion of spatial interactions causes that different results are obtained than a classical measure would produce because the transformation of (11) shifts the focus of analysis from the specific situation of a given unit to the impacts of its interactions with the contiguous units. This effectively means that we study a situation in which the values of diagnostic variables in one unit depend on their values in the contiguous units.

[5] A description of Hellwig's method in the English language can be found, *inter alia*, in the monograph by Nermend**,** 2009 (pp. 37-44).

[6] For a discussion of different approaches to constructing a matrix of spatial weights see, for instance, Suchecki at al. (2010, pp. 26-34), Suchecki and Olejnik (2010, pp. 105-107).

In this study, the author puts forward her own modification of the concept proposed by E. Antczak, which consists in assigning to the diagonal of the weight matrix ***W*** values different from 0 for first-order contiguity[7]. Matrix *W* in formula (11) is now written as:

$$w_{ik} = \begin{cases} \frac{1}{m_i} & \text{when unit } k \text{ is contiguoust to unit } i \text{ or } i = k, \\ 0 & otherwise, \end{cases} \quad (13)$$

m_i – the number of territorial units constituting the *i*-th region, i.e. units contiguous to unit *i* including that unit.

With formulas (11) and (13), we can obtain the value of the *j*-th diagnostic variable, x_{ij}^{*}, for the *i*-th unit, which represents the average level of the investigated phenomenon for a region consisting of the *i*-th unit and its contiguous units[8].

The above approach utilizes a first-order contiguity matrix (i.e. a matrix of contiguous units), but higher-order matrices can be used as well. A measure obtained from such obtained matrix of variables and any synthetic measure of development will be henceforth referred to as a spatial measure of development with equally weighted units (SMD-EW).

SMD-EW assumes that all territorial units contiguous to the *i*-th unit have equal weights, which hardly ever occurs in real life. The usual case is that larger units, e.g. cities, have stronger influence on the course of some phenomenon in their regions than smaller units, e.g. rural areas. To account for this disproportion, SMD-EW is modified further into a spatial measure of development with intra-regionally weighted units (SMD-IRW) by weight matrix (13) in formula (11) with a matrix $\mathbf{S}$ of intra-regional weights allowing the 'significance' of the unit in the region to be included[9]. In this paper, intra-regional weights are represented by the unit's population in relation to the total population in the region, but weights can be selected according to the purpose of research. Mathematically, the data matrix transformation can be written as:

$$\boldsymbol{X}^{**}_{m\times p} = \boldsymbol{S}_{m\times m}\boldsymbol{X}_{m\times p}, \quad (14)$$

where:

$\boldsymbol{S} = [s_{ik}]$ – the $m\times m$ matrix of intra-regional weights,

---

[7] A description of this modification can also be found in the monograph by A. Majdzińska (2016, pp. 67, 173-174).

[8] The definition of a region as used in this paper is different from that typically used in the geographical or economic literature (for a review of the definitions see, for instance, the monograph by Majdzińska, 2016, pp. 11-13 or by Montello, 2008, p. 305). For the purposes of this research, a region is understood as a single entity representing the level of the analysed phenomenon in a territorial unit (a county) and the units adjacent to it. In the Polish circumstances, there can be as many regions as counties (i.e. 380). This approach excludes the similarity of the analysed phenomenon as a criterion for grouping units and requires the transformation of a contiguity matrix to account for the units' own weights (formulas 11, 13-15). It also involves a different interpretation of the results.

[9] The modification devised by the author has not been published before.

$\boldsymbol{X} = [x_{ij}]$ – the $m \times p$ matrix of diagnostic variables

$$s_{ik} = \begin{cases} \frac{L_k}{\sum_{r=1}^{m_i} L_r} & \text{when unit } k \text{ is contiguoust to unit } i \text{ or } i = k, \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

L_k – the size of the *k*-th territorial unit.

L_r – the size of the *r*-th region.

It is worth noting that the established weights are the same for all diagnostic variables[10].

The spatial measures of development proposed below are designed using similar aggregation methods as proposed by Sokołowski, Zając or Hellwig[11] in the non-spatial framework. They refer to the variables transformed according to formulas (11), (13) and (14), (15) and will be termed "spatial measures of development with equally weighted units" in the case of variable transformations (11), (13) or spatial measures of development with intra-regionally weighted units" in the case of transformations (14), (15). Both approaches utilize the concept of the so-called reference unit. For comparison, a spatial measure without employing a reference unit is also constructed.

3. Application of the spatial measures of development

In this section, the spatial measures of development are used to rank Polish counties according to their demographic potential. To prepare the rankings, four diagnostic variables were selected from a set of variables characterising the population age structure in the counties and developments relating to natural movement and migrations, namely[12]:

x_1 – the share of population aged 15-44 years,

x_2 – a total fertility rate,

x_3 – a standardised death rate[13],

x_4 – the ratio of in-migrants to out-migrants between counties.

These diagnostic variables concisely present the counties' demographic potential and its changes brought about by demographic processes. For the purposes of this analysis, variables x_1, x_2 and x_4 were assumed to be stimulants

---

[10] Another possibility is to weight variables, but this approach is not used in this study.

[11] Hellwig's measure from which E. Antczak (2013) developed her measure of spatial development is not discussed in this research. For the applications of Hellwig's taxonomic measure of development and its modification in demographic research see the monograph by A. Majdzińska (2016).

[12] Variables were selected using the statistical criteria, such as relatively high variability and weak correlations.

[13] For the purposes of standardisation, the age structure of Polish population as on 31 Dec. 2016 was taken as a standard.

and variable x_3 to be a destimulant (converted into a stimulant using a ratio formula[14]).

Variable x_1 stands for the current level of demographic potential of a county, which is basically represented by population in the mobile working age group[15]. Variable x_2 denotes the reproduction level (i.e. the population's ability to reproduce itself), variable x_3 is a measure of population loss due to deaths, and variable x_4 represents migrations that frequently significantly contribute to fluctuations in the size of a population.

The analysis is performed on the 2005 and 2016 data with a view to ranking counties according to their demographic potential. The rankings are first produced using the spatial measures of development based on the Sokołowski and Zając method and then the spatial measures proposed by the non-reference method of sums. The outcomes of both approaches are compared and discussed.

In the first step, the values of the synthetic measure of development are calculated for counties based on the Sokołowski and Zając method[16] (formula 4) to be used as the criterion for assigning counties to four typological groups[17].

In 2005 and 2016, the most favourable levels of demographic potential were observed in the first group of counties, most of which are contiguous to or surround large, thriving cities of Warsaw, Tri-City, Poznań, Wrocław, Kraków, Toruń and Bydgoszcz (Figures 1-2). The majority of them were receiving in-migrants and were characterised by relatively high shares of people aged 15-44 years and quite advantageous reproduction rates (due to comparatively high fertility rates and low standardized death rates), although none of the counties fully matched the reference model of development. The most similar to it in both 2005 and 2016 were Gdański, Kartuski and Poznański counties (Table 1), whereas the counties in central, eastern and south-western Poland were the most distant from it. In 2005 these were Łódzki and Kutnowski counties (Łódzkie voivodeship), Hajnowski and Siemiatycki (Podlaskie voivodeship), Wałbrzyski (Dolnośląskie voivodeship), Krasnostawski (Lubelskie voivodeship) and in 2016 and Hajnowski county (Podlaskie voivodeship), Kłodzki, Ząbkowicki and the town of Wałbrzych (Dolnośląskie voivodeships), and Ostrowiecki and Skarżyski (Świętokrzyskie voivodeship).

---

[14] The ratio formula is written as (Panek, 2009, p. 36):
$x_j^S = \frac{1}{x_j^D}$; where: x_j^S is a stymulant and a x_j^D is a destimulant.

[15] According to the Polish law, the working age is 18-59 years for women and 18-64 years for men (*Ustawa z dnia 16 listopada 2016 r. …*). The mobile working age is 18-44 years (GUS, 2017b, p. 143). The age of entry into labour force as adopted by the Labour Force Survey (LFS) is 15 years (GUS, 2017a, p.16).

[16] Sokołowski and Zając define a reference unit of development as an abstract unit that has attained the highest level of development. In this study, the values of the diagnostic variables for the reference model are the following:
- 2005: $x_1 = 0.48$; $x_2 = 1.87$; $x_3 = 63.84$; $x_4 = 3.62$ (values noted for Pszczyński, Kartuski, Rzeszowski and Poznański counties);
- 2016: $x_1 = 0.46$; $x_2 = 2.03$; $x_3 = 77.78$; $x_4 = 3.76$ (Gdański, Kartuski, Tarnowski and Poznański counties).

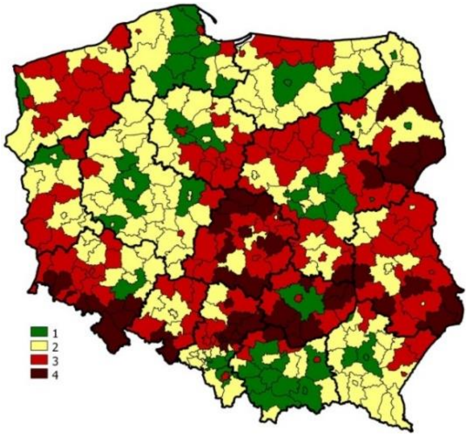
[17] Type (I) $z_i \geq \bar{z} + s_z$; (II) $\bar{z} + s_z > z_i \geq \bar{z}$; (III) $\bar{z} > z_i \geq \bar{z} - s_z$; (IV) $z_i < \bar{z} - s_z$, where: z_i – the value of the synthetic index for the *i*-th object, $\bar{z}$ – the arithmetic average of the synthetic measure, s_z – standard deviation from the value of the synthetic measure (Nowak, 1990, p. 93; Majdzińska, 2016, pp. 32-33). If a standard deviation of 0.5 was adopted, 8 typological groups would have been produced and the results would have been more detailed.

The similarity between the counties' rankings in 2005 and 2016 was corroborated statistically by Pearson's linear correlation coefficient of 0.85 (p<0.05) and Spearman's rank coefficient of 0.81.

Table 1. Counties' rankings according to their demographic potential – the synthetic measure based on the original method by the Sokołowski and Zając method, 2005 and 2016

counties with the lowest values of the measure				counties with the highest values of the measure			
2005		2016		2005		2016	
county	z_i	county	z_i	county	z_i	County	z_i
Łódź	0.14	Hajnowski	0.13	Kartuski	0.76	Poznański	0.76
Hajnowski	0.17	Wałbrzych	0.14	Poznański	0.75	Kartuski	0.74
Kutnowski	0.17	Kłodzki	0.16	Gdański	0.69	Gdański	0.72
Wałbrzyski	0.17	Ostrowiecki	0.17	Toruński	0.68	Wrocławski	0.69
Krasnostawski	0.17	Ząbkowicki	0.17	Bydgoski	0.68	Wielicki	0.67
Siemiatycki	0.18	Skarżyski	0.17	Wejherowski	0.68	Wołomiński	0.65
Pińczowski	0.19	Sosnowiec	0.17	Piaseczyński	0.67	Piaseczyński	0.64
Kazimierski	0.20	Głubczycki	0.18	Policki	0.62	Rzeszów	0.58
Sosnowiec	0.20	Piekary Śląskie	0.18	Pucki	0.59	Wejherowski	0.58
Poddębicki	0.21	Kutnowski	0.18	Wielicki	0.59	Bydgoski	0.55
Sopot	0.21	Częstochowa	0.19	Wrocławski	0.58	Legionowski	0.55
Lipski	0.21	Wałbrzyski	0.20	Nowosądecki	0.57	Grodziski	0.54
Łęczycki	0.21	Jelenia Góra	0.20	Myślenicki	0.57	Nowosądecki	0.54
Skarżyski	0.22	Hrubieszowski	0.20	Rzeszów	0.57	Limanowski	0.53
Zawierciański	0.22	Łódź	0.20	Warszawski Zach.	0.56	Rzeszowski	0.53

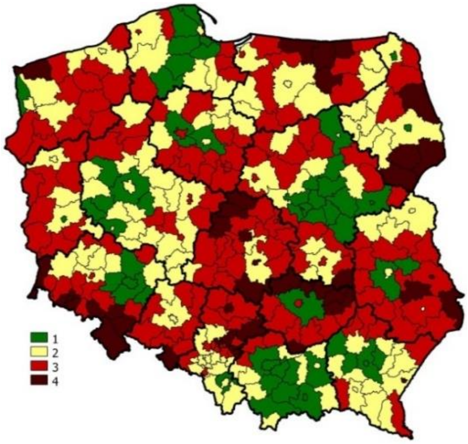
Source: GUS (BDL); created by the author.



$\bar{z} = 0.37,\ s_z = 0.10,\ min = 0.14,\ max = 0.76$

Source: GUS (BDL); created by the author.

Figure 1. Differences in regions' demographic potential in 2005 (based on the synthetic measure derived from the Sokołowski and Zając method)



$\bar{z} = 0.35,\ s_z = 0.09, min = 0.13,\ max = 0.76$

Source: GUS (BDL); created by the author.

Figure 2. Differences in regions' demographic potential in 2016 (based on the synthetic measure derived from the Sokołowski and Zając method)

Because the rankings characterise counties in a synthetic manner (Table 2)[18], the counties in particular typological groups have different values of the diagnostic variables, even though their distance from the reference county is similar. In almost all groups, variables x_1, x_2 and x_3 differ to a relatively small extent[19]).

[18] For the sake of comparison, the same set of diagnostic variables (x_1–x_4) and classical Hellwig's taxonomic measure of development (see Hellwig, 1968) were used to produce the counties' rankings for 2016. The values of Hellwig's measure and of the measure based on the Sokołowski and Zając method were similar; Pearson's linear correlation coefficient of 0.98 (p<0.05) and Spearman's rank correlation coefficient of 0.97 showed that the similarity was high and statistically significant.

[19] The coefficient of variance for x_4 was high, reaching 0.52 for the full sample (the 2016 value of this variable ranged from 0.3 to 3.8).

Table 2. Statistics characterising the groups of counties (the Sokołowski and Zając synthetic method, 2016).

Typological group	Average values of variables				Coefficients of variance			
	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
1	0.43	1.50	97.5	1.59	0.03	0.10	0.07	0.48
2	0.43	1.37	102.1	0.91	0.03	0.08	0.08	0.31
3	0.42	1.29	107.1	0.70	0.03	0.08	0.07	0.29
4	0.40	1.18	111.1	0.67	0.03	0.08	0.06	0.31

Source: GUS (BDL); created by the author.

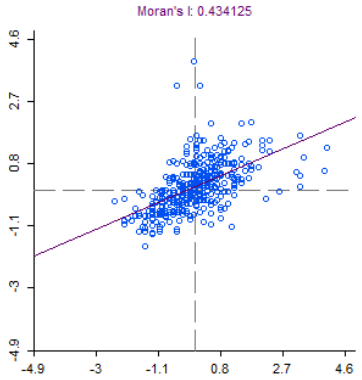
The spatial dependency of the measure derived from the Sokołowski and Zając method was tested using Moran's I statistics. Its 2005 and 2016 values of 0.43 and 0.45[20], respectively, show that the global spatial autocorrelation for first-order contiguity was positive, moderate and statistically significant ($p<0.001$)[21] (Figures 3 and 5).

According to local Moran's I_i statistics, in 2005 the largest clusters of counties with relatively high and significant levels of demographic potential were in Pomorskie, Wielkopolskie and Małopolskie voivodeships and in 2016 also in Mazowieckie voivodeship. The majority of the regions where the levels of demographic potential were rather unfavourable were situated in central, eastern and south-western Poland (Figures 4 and 6).

The spatial autocorrelation was also evaluated for the case of second-order contiguity in 2016. The global autocorrelation turned out to be slightly weaker compared with that determined for first-order contiguity (Moran's I statistics 0.17; $p<0.001$).

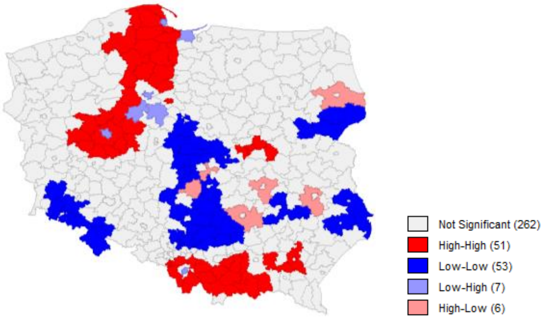
[20] Calculations were performed using the GeoDa software.

[21] The global spatial autocorrelation of the counties was determined based on a first-order row-standardised contiguity matrix. Moran's I statistics for the selected variables and years were the following:
2005 – x_1: 0.46; x_2: 0.45; x_3: 0.37; x_4: 0.21 ($p<0.001$);
2016 – x_1: 0.37; x_2: 0.50; x_3: 0.36; x_4: 0.24 ($p<0.001$).



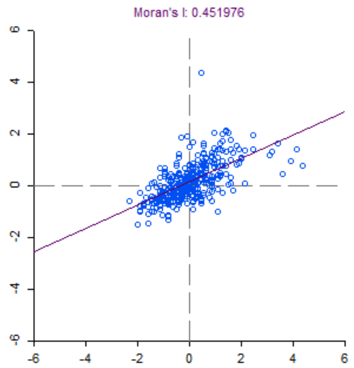
Source: GUS (BDL); created by the author using GeoDa software.

Figure 3. Global spatial autocorrelation for counties' demographic potential in 2005 (based on the Sokołowski and Zając synthetic method**;** first-order contiguity)



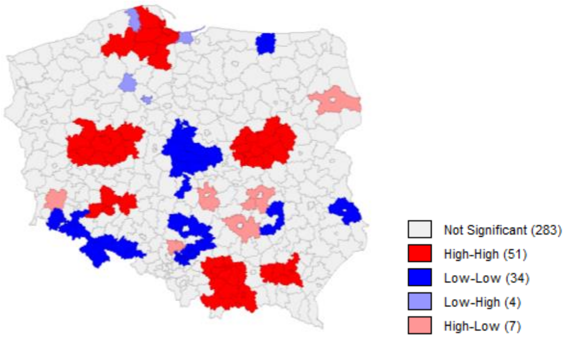
Source: GUS (BDL); created by the author using GeoDa software.

Figure 4. Local spatial autocorrelation for counties' demographic potential in 2005 (based on the Sokołowski and Zając synthetic method; first-order contiguity)



Source: GUS (BDL); developed by the author using GeoDa software.

Figure 5. Global spatial autocorrelation for counties' demographic potential in 2016 (based on the Sokołowski and Zając synthetic method; first-order contiguity)



Source: GUS (BDL); developed by the author using GeoDa software.

Figure 6. Local spatial autocorrelation for counties' demographic potential in 2016 (based on the Sokołowski and Zając synthetic method; first-order contiguity)

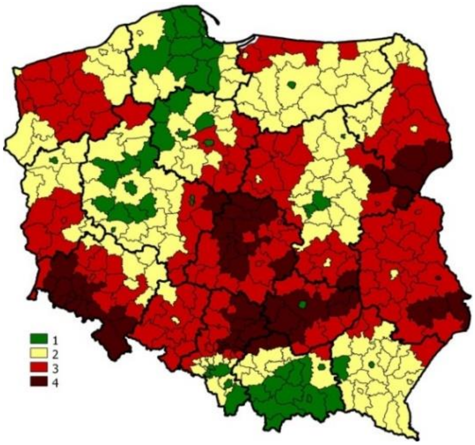
In the next step, counties were ranked and grouped using the spatial measure based on the Sokołowski and Zając method described in Section 2.2 and indicators transformed according to formulas (11)-(12). This measure accounts for the influence of intra-regional dependencies, so it produced different rankings than the classical measure did (formula 4). Some counties that had previously ranked relatively low moved up as a result of spatial interactions modifying the course of demographic processes, while others moved down because of the adjacency of counties with a relatively low levels of demographic potential (Tables 1 and 3 and Graphs 7 and 8).

Most large cities improved their rankings, because higher fertility rates and positive net migration rates in the contiguous counties strengthened their demographic potential eroded by depopulation processes.

Table 3. Counties' (regions) rankings according to their demographic potential – based on the spatial measure developed from the Sokołowski and Zając method, 2005 and 2016

counties with the lowest value of the measure				counties with the highest value of the measure			
2005		2016		2005		2016	
county	z_i	county	z_i	county	z_i	county	z_i
Kłodzki	0.15	Kłodzki	0.16	Poznań	0.85	Poznań	0.65
Dąbrowa Górnicza	0.20	Jelenia Góra	0.16	Toruń	0.77	Świdwiński	0.57
Zduńskowolski	0.20	Kamiennogórski	0.18	Bydgoszcz	0.76	Wałbrzyski	0.50
Łęczycki	0.20	Wałbrzych	0.18	Nowy Sącz	0.63	Pruszkowski	0.47
Siemianowice Śląskie	0.20	Chorzów	0.20	Kościerski	0.63	Warszawa	0.46
Zgierski	0.21	Siemianowice Śląskie	0.20	Lęborski	0.62	Kościerski	0.45
Świętochłowice	0.21	Ząbkowicki	0.20	Wejherowski	0.58	Wrocław	0.45
Ząbkowicki	0.21	Świętochłowice	0.20	Warszawa	0.58	Rzeszów	0.44
Kamiennogórski	0.22	Dąbrowa Górnicza	0.20	Gdynia	0.57	Nowy Sącz	0.43
Chorzów	0.22	Jeleniogórski	0.21	Pucki	0.57	Kraków	0.43
Kielecki	0.22	Kielecki	0.21	Pruszkowski	0.57	Bydgoszcz	0.43
Piekary Śląskie	0.23	Częstochowski	0.21	Tatrzański	0.57	Lęborski	0.43
Zawierciański	0.23	Szydłowiecki	0.21	Średzki	0.56	Gdynia	0.43
Bielski	0.23	Kętrzyński	0.21	Limanowski	0.55	Grodziski	0.42
Będziński	0.23	Łęczycki	0.22	Rzeszów	0.55	Bocheński	0.42

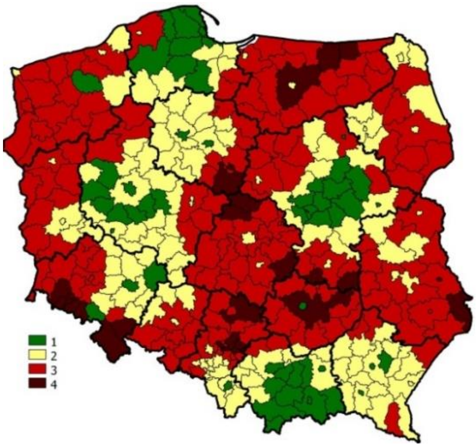
Source: GUS (BDL); created by the author.



$\bar{z} = 0.38, \quad s_z = 0.10, \quad min = 0.15, \quad max = 0.85$

Source: GUS (BDL); created by the author.

Figure 7. Differences in regions' demographic potential in 2005 (based on the spatial measure derived from the Sokołowski and Zając method)



$\bar{z} = 0.29, \quad s_z = 0.06, \quad min = 0.16, \quad max = 0.65$

Source: GUS (BDL); created by the author.

Figure 8. Differences in regions' demographic potential in 2016 (based on the spatial measure derived the Sokołowski and Zając method)

Subsequently, a spatial measure of development with equally weighted units (SMD-EW) (formulas 11 and 13) was calculated for counties and the areas around them (Table 4 and Figures 9 and 10). Because of their weights, some 'insular' counties in the previous classification (Graphs 7 and 8) joined areas formed by the adjacent counties, in spite of significantly different values of diagnostic variables. The results thus obtained for individual counties can be interpreted as representing the demographic potential of a region consisting of a county and the adjacent units.

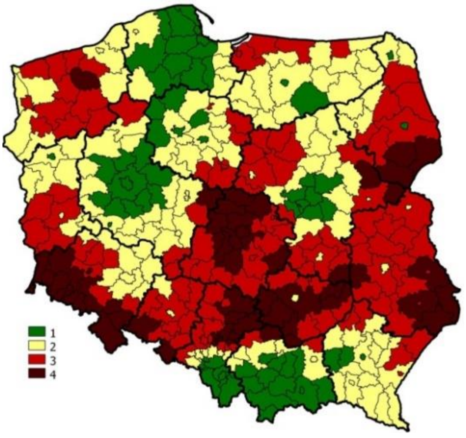
The highest levels of demographic potential in both 2005 and 2016 occurred in Pomorskie, Wielkopolskie, Małopolskie and Mazowieckie voivodeships, and the lowest in Łódzkie, Świętokrzyskie, Dolnośląskie, Podlaskie and Lubelskie voivodeships.

The highest levels of demographic potential in 2016 were determined for Poznań (its relatively low fertility rate and negative net migration were offset by high demographic potential of Poznański county) and Rzeszów (Rzeszowski county and the town of Rzeszów enjoyed a relatively good demographic situation because of low mortality rates and definitely more people seeking residence in the area than leaving it). The other end of the scale is represented by Kłodzki county (a very low fertility rate and a negative migration rate in this region were accompanied by a relatively high standardised mortality rate) and the town of Wałbrzych (its demographic potential was as low as in the county).

Table 4. Counties' (regions) rankings according to their demographic capital – the SMD-EW, 2005 and 2016

counties with the lowest value of the measure				counties with the highest value of the measure			
2005		2016		2005		2016	
county	z_i	county	z_i	county	z_i	county	z_i
Wałbrzyski	0.14	Kłodzki	0.12	Bydgoszcz	0.80	Poznań	0.79
Świdwiński	0.26	Wałbrzych	0.12	Poznań	0.79	Rzeszów	0.75
Kłodzki	0.27	Jelenia Góra	0.13	Wejherowski	0.78	Wielicki	0.71
Ząbkowicki	0.32	Wałbrzyski	0.16	Kościerski	0.78	Pruszkowski	0.71
Łęczycki	0.32	Ząbkowicki	0.18	Pucki	0.76	Warszawa	0.71
Dąbrowa Górnicza	0.32	Kamiennogórski	0.19	Toruń	0.76	Kościerski	0.70
Siemianowice Śląskie	0.33	Chorzów	0.19	Nowy Sącz	0.74	Wrocław	0.69
Chorzów	0.33	Świętochłowice	0.20	Lęborski	0.74	Kraków	0.67
Świętochłowice	0.33	Dąbrowa Górnicza	0.20	Kartuski	0.74	Kartuski	0.66
Hajnowski	0.34	Siemianowice Śląskie	0.20	Rzeszów	0.74	Grodziski	0.65
Jelenia Góra	0.34	Sosnowiec	0.20	Pruszkowski	0.73	Wołomiński	0.65
Kamiennogórski	0.34	Jeleniogórski	0.21	Warszawa	0.73	Bocheński	0.65
Kutnowski	0.34	Hajnowski	0.21	Limanowski	0.71	Myślenicki	0.64
Piekary Śląskie	0.35	Głubczycki	0.21	Myślenicki	0.71	Wejherowski	0.64
Zawierciański	0.35	Ostrowiecki	0.22	Wielicki	0.71	Piaseczyński	0.63

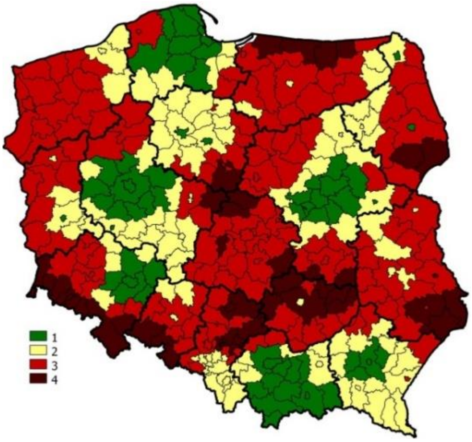
Source: GUS (BDL); created by the author.



$\bar{z} = 0.51,\quad s_z = 0.10,\quad min = 0.14,\quad max = 0.80$

Source: GUS (BDL); created by the author.

Figure 9. Differences in regions' demographic potential in 2005 (according to SMD-EW)



$\bar{z} = 0.39,\quad s_z = 0.12,\quad min = 0.12,\quad max = 0.79$

Source: GUS (BDL); created by the author.

Figure 10. Differences in regions' demographic potential in 2016 (according to SMD-EW)

In the next step, the spatial measure of development with intra-regionally weighted units (SMD-IRW) (formulas 14-15) was used.

The clusters of counties in Figures 11 and 12 and in Figures 9 and 10 are relatively similar, despite different rankings of their constituent counties (regions) determined from the SMD-EW and SMD-IRW methods (Tables 4 and 5).

In both 2005 and 2016, the highest average demographic potential was observed in the regions centred around Grodziski, Średzki and Kościański counties (Wielkopolskie voivodeship; Table 5), which significantly improved their positions compared with the previous ranking, having "absorbed" much of the demographic potential of Poznański county. However, the 2016 position of Poznań itself proved considerably lower (Table 4 and 5), because the intra-regional weight increased its "contribution" to the demographic potential of the city and the county[22]. The lowest levels of demographic potential occurred in counties in central and south-western Poland (in 2016 the worst situation in that respect was noted in the towns of Wałbrzych and Jelenia Góra and in Kłodzki county).

After intra-regional weights were assigned to counties, the most populous units gained slight advantage over other units in the same region (large units, especially large cities, tend to exert strong economic, social and demographic influence on the adjacent areas). The main reason for which many cities moved

[22] In 2016, Poznań accounted for 59.1% of the total population living in the city and the county (GUS, 2017b).

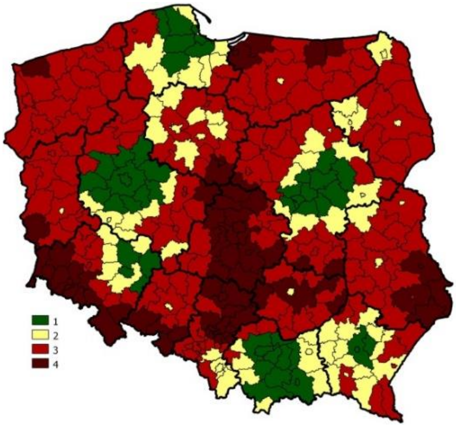
down in this ranking was relatively lower fertility rates and negative migration rates.

The relative similarity between the 2005 and 2016 rankings of the counties obtained from the SMD-IRW was confirmed by Pearson's correlation coefficient of 0.88 ($p<0.05$) and Spearman's rank coefficient of 0.86.

Tabale 5. Counties' (regions) rankings according to their demographic potential – the SMD-IRW, 2005 and 2016

counties with the lowest value of the measure				counties with the highest value of the measure			
2005		2016		2005		2016	
county	z_i	county	z_i	county	z_i	county	z_i
Łódź	0.12	Wałbrzych	0.09	Średzki	0.82	Grodziski	0.83
Zgierski	0.13	Kłodzki	0.11	Grodziski	0.82	Średzki	0.80
Kłodzki	0.15	Jelenia Góra	0.13	Kościański	0.78	Kościański	0.77
Łódzki Wschodni	0.15	Wałbrzyski	0.15	Obornicki	0.78	Szamotulski	0.76
Pabianicki	0.17	Ząbkowicki	0.16	Śremski	0.77	Śremski	0.76
Wałbrzyski	0.18	Kamiennogórski	0.18	Lęborski	0.77	Obornicki	0.73
Dąbrowa Górnicza	0.19	Dąbrowa Górnicza	0.18	Szamotulski	0.77	Nowotomyski	0.72
Ząbkowicki	0.19	Sosnowiec	0.19	Kościerski	0.76	Rzeszów	0.71
Siemianowice Śląskie	0.20	Jeleniogórski	0.19	Wejherowski	0.72	Wrzesiński	0.71
Jelenia Góra	0.21	Świętochłowice	0.20	Nowy Sącz	0.72	Kościerski	0.70
Chorzów	0.21	Łódź	0.20	Gnieźnieński	0.72	Gnieźnieński	0.69
Łęczycki	0.21	Częstochowa	0.20	Wrzesiński	0.71	Poznań	0.69
Świętochłowice	0.22	Chorzów	0.20	Nowotomyski	0.71	Lęborski	0.67
Kamiennogórski	0.22	Mysłowice	0.21	Wągrowiecki	0.70	Bocheński	0.65
Kutnowski	0.22	Częstochowski	0.21	Rzeszów	0.69	Wielicki	0.64

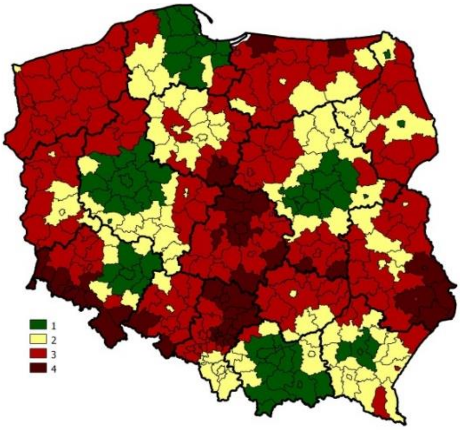
Source: GUS (BDL); developed by the author.



$\bar{z} = 0.43,$ $s_z = 0.13,$ $min = 0.12,$ $max = 0.82$

Source: GUS (BDL); created by the author.

Figure 11. Differences in regions' demographic potential in 2005 (according to SMD-IRW)



$\bar{z} = 0.40,$ $s_z = 0.12,$ $min = 0.09,$ $max = 0.83$

Source: GUS (BDL); created by the author.

Figure 12. Differences in regions' demographic potential in 2016 (according to SMD-IRW)

Finally, the spatial measure of development based on the non-reference method of sums was constructed, as well as its modifications allowing for counties' weights and intra-regional weights. The outcomes of these three measures will not be discussed in detail, because they are basically similar to those produced by Sokołowski and Zając method. One thing that is noteworthy, however, is that the use of intra-regional weights affected the counties' rankings (Table 5 and 6). In 2016, the most favourable levels of demographic potential were noted for Grodziski and Średzki counties in Wielkopolskie voivodeship, whereas in Tomaszowski county in Lubelskie voivodeship and in Kłodzki and Ząbkowicki counties in Dolnośląskie voivodeship the levels were the lowest (Table 6).

The different rankings of individual counties (i.e. regions) produced by both measures should be mainly attributed to the differences in their construction; in the case of the Sokołowski and Zając measure counties were compared with the reference model and in the case of the measure based on the non-reference method of sums with each other.

Table 6. Counties' (region) rankings according to their demographic potential – the SMD-IRW based on the non-reference method of sums, 2005 and 2016.

counties with the lowest value of the measure				counties with the highest value of the measure			
2005		2016		2005		2016	
county	z_i	county	z_i	county	z_i	county	z_i
Głubczycki	0.44	Tomaszowski	0.41	Grodziski	1.00	Grodziski	1.00
Kłodzki	0.44	Ząbkowicki	0.42	Średzki	0.99	Średzki	0.99
Ząbkowicki	0.45	Kłodzki	0.42	Obornicki	0.96	Szamotulski	0.96
Kędzierzyńsko-kozielski	0.45	Sandomierski	0.42	Szamotulski	0.96	Obornicki	0.95
Prudnicki	0.46	Węgorzewski	0.42	Kościański	0.94	Śremski	0.94
Krapkowicki	0.46	Lubaczowski	0.42	Śremski	0.94	Kościański	0.94
Zabrze	0.46	Stalowowolski	0.42	Nowotomyski	0.90	Nowotomyski	0.89
Gliwice	0.46	Zamojski	0.43	Gnieźnieński	0.88	Wrzesiński	0.88
Wałbrzyski	0.46	Hrubieszowski	0.43	Wrzesiński	0.88	Gnieźnieński	0.87
Opolski	0.47	Tarnobrzeg	0.43	Wągrowiecki	0.86	Poznań	0.82
Strzelecki	0.47	Wałbrzyski	0.43	Grodziski	0.85	Wągrowiecki	0.81
Bytom	0.47	Braniewski	0.44	Chełmiński	0.85	Grodziski	0.81
Chorzów	0.48	Leski	0.44	Poznań	0.81	Warszawa	0.79
Mysłowice	0.48	Wałbrzych	0.44	Lęborski	0.80	Kościerski	0.78
Nyski	0.48	Głubczycki	0.44	Bocheński	0.79	Bocheński	0.78

Source: GUS (BDL); created by the author.

4. Conclusion

Spatial measures of development allow comparing territorial units for the intensity of the selected phenomenon and evaluating the extent to which the intensity in one unit is determined by the contiguous units. The measures can be derived from any classical synthetic measure (with or without a reference unit) on condition that each variable is globally spatially autocorrelated. It is also vital to remember that the measures' results should be interpreted as allowing for the type of interactions (weights) adopted for the studied objects.

In this study, three approaches to constructing a spatial measure of development are proposed, all derived from the Sokołowski and Zając method utilising a reference unit and the non-reference method of sums. They were subsequently used to rank Polish counties according to their demographic potential in 2005 and 2016.

The first of them, based on the classical first-order contiguity matrix and the concept proposed by E. Antczak (2013), assumes that the level of the investigated phenomenon in a unit is only determined by the values of variables determining the phenomenon in the contiguous units. The situation in the first unit is thus omitted.

The second approach is the author's modification of the first one, involving the extension of the contiguity matrix to account for the territorial units' own weights. This change causes that the result obtained for a unit can be interpreted as the average level of the phenomenon under consideration in a region made up of this and the contiguous units. This modification has been called a spatial measure of development with equally weighted units (SMD-EW) and its variant (i.e. the third approach) a spatial measure of development with intra-regionally weighted units (SMD-IRW). Both measures can also be used to evaluate a territorial unit's impact on the contiguous units in terms of different criteria.

Other transformations of the measures are also possible, for instance by adopting matrices of higher orders or by assigning weights to the diagnostic variables.

A natural consequence of the three measures having different characteristics is different interpretation of their results. In the author's opinion, SMD-EW and SMD-IRW better evaluate the demographic potential of various territorial units, because, in addition to considering interactions between them, they also allow for their internal situation. In other words, they present the combined demographic potential of a unit and the units around it (i.e. of a region). The only inconvenience is that they are slightly more difficult to apply (a significant global autocorrelation of each variable is required) and the interpretation is different compared with the classical synthetic measure.

Nevertheless, spatial measures of development taking account of territorial units' weights are useful for evaluating the demographic potential of regions as defined in this article, particularly in studies investigating the demographic potential of the largest cities that lose inhabitants because of suburbanization and other processes, etc., while receiving from contiguous areas working-age people, who come every day to the city to work or study[23]. For instance, a relatively high value of a spatial measure obtained for a depopulating city indicates the demographic potential of the region made up of the city and the contiguous units (i.e. the city's capacity to take in population capital from the surrounding areas). In the case of a suburban unit that has a good demographic situation, a low value of the measure shows to what extent much the unit "supplies" the adjacent units.

[23] For more than two decades now, the majority of the largest Polish cities (in population terms) have been shrinking due to natural causes (declining fertility rates and relatively steady mortality rates) and relocation of many of their residents mainly to suburban areas.

This information is of practical importance in analysing the present and future of regional labour markets.

The created typology of Polish counties showed that the largest groupings of regions with the relatively best demographic potential are concentrated in Mazowieckie, Pomorskie, Wielkopolskie and Małopolskie voivodeships. This is probably due to the voivodeships' capital cities and suburbs being perceived favourably as in-migration areas and their fairly high fertility levels. Regions with the lowest demographic potential (resulting from low fertility rates, high shares of the elderly population and negative net migration rates) are mostly found in Łódzkie, Świętokrzyskie and Lubelskie voivodeships.

It is predicted that in the next several years the demographic potential of most regions in Poland will decline, the main reasons for which will be low fertility rates and the increasing proportion of elderly people. It seems, therefore, advisable that areas characterised by the lowest demographic potential be flagged as in need of attention from the government and of actions mitigating the negative impacts of changes in the population size and structure.

Both SMD-EW and SMD-IRW are worth considering as the tools of spatial analysis enhancing the demographic regionalisation methods and the forecasts of regions' demographic potential. At the government level, they can be used to support the creation of regional and local population policies and labour market policies.

REFERENCES

ANTCZAK, E., (2013). Przestrzenny taksonomiczny miernik rozwoju. Wiadomości Statystyczne, 7, pp. 37–53.

CIPD, (2017). Human capital analytics and reporting: exploring theory and evidence. Research report. Available through: https://www.cipd.co.uk/Images/human-capital-analytics-and-reporting_tcm18-22281.pdf [Accessed 19.03.2018].

DAWID, P. A., (2001). Knowledge, Capabilities and Human Capital Formation in Economic Growth. New Zealand Treasury Working Paper, 01/13.

EDIEV, D., (2001). Application of the Demographic Potential Concept to Understanding the Russian Population History and Prospects. Demographic Research, Vol. 4, pp. 289–336.

GAZIŃSKA, M., (2003). Potencjał demograficzny w regionie. Analiza ilościowa, Rozprawy i Studia, t. 448, Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.

GETIS, A., (2010). Spatial Autocorrelation. In: M.M. Fischer, A. Getis, eds. 2010, Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications, Springer-Verlag Berlin Heidelberg.

GIZA-POLESZCZUK, A., MARODY, M., (2000). Kapitał ludzki i systemowy. In: A. Giza-Poleszczuk, M. Marody, A. Rychard, eds. 2000, Strategie i system. Polacy w obliczu zmiany społecznej, Warszawa: Wydawnictwo IFiS PAN.

GOLDIN, C., (2014). Human Capital. In: C. Diebolt, M. Haupert, ed. 2014. In: Handbook of Cliometrics, Springer-Verlag, Berlin, pp. 55–86.

GRIFFITH, D. A., (2003). Spatial autocorrelation and spatial filtering. Berlin: Springer.

GUS, Bank Danych Lokalnych (BDL). Available through: https://bdl.stat.gov.pl/BDL/start [Accessed 30.09.2017].

GUS, (2017a). Aktywność ekonomiczna ludności Polski. II kwartał 2017, Warszawa.

GUS, (2017b). Rocznik Demograficzny 2017. Warszawa.

HELLWIG, Z., (1968). Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. Przegląd Statystyczny, 4, pp. 308–327.

JOHNSON, R., (1984). Elementary statistics, 4th edition. Boston: PWS Publishers

KOTARSKI, H., 2013. Kapitał ludzki i kapitał społeczny a rozwój województwa podkarpackiego. Rzeszów: Wydawnictwo Uniwersytetu Rzeszowskiego.

MAJDZIŃSKA, A., (2016). Regionalizacja demograficzna. Wybrane metody i ich aplikacje, Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

MONTELLO, D. R., (2008). Geographic regions as brute facts, social facts and institutional facts. In: B. Smith, D.M. Mark I. Ehrlich, eds. 2008, The mystery of capital and the construction of social reality, Chicago. Available through: https://geog.ucsb.edu/~montello/pubs/mystery.pdf [Accessed 10.02.2018].

NERMEND, K., (2009). Vector Calculus in Regional Development Analysis. Physica-Verlag, Heidelberg.

NOWAK, E., (1990). Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych. Warszawa: PWN.

OBRĘBALSKI, M., (2017). Demographic potential in functional areas of the selected medium-sized cities in Poland and the Czech Republic. GeoScape, Vol. 11, issue 1, pp. 16–24. Available through: https://content.sciendo.com/view/journals/geosc/11/1/article-p16.xml?rskey=lCVdj7&result=1 [Accessed 20.06.2018].

PANEK, T., (2009). Statystyczne metody wielowymiarowej analizy porównawczej. Warszawa: SGH.

PASTUSZKA, S., (2017). Regional Differentiation of the Demographic Potential in Italy and Poland. Comparative Economic Research, Vol. 20, No. 3.

ROSZKOWSKA, S., (2013). Kapitał ludzki a wzrost gospodarczy w Polsce. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

SCARPACI, J., (1984). A methodological note on location-allocation models. American Journal of Public Health, Vol. 74, No. 10, pp. 1155–1157.

SOKOŁOWSKI, A., ZAJĄC K., (1987). Rozwój demograficzny a rozwój gospodarczy. Warszawa: PWE.

SUCHECKI, B., OLEJNIK, A., (2010). Miary i testy statystyczne w eksploracyjnej analizie danych przestrzennych. In: B. Suchecki, ed. 2010, Ekonometria przestrzenna, Metody i modele analizy danych przestrzennych, Warszawa: Wydawnictwo C.H. Beck.

SUCHECKI, B., LEWANDOWSKA-GWARDA, K., OLEJNIK A., (2010). Wprowadzenie do przestrzennych analiz ekonomicznych. In: B. Suchecki B., ed. 2010, Ekonometria przestrzenna, Metody i modele analizy danych przestrzennych, Warszawa: Wydawnictwo C.H. Beck.

THE HUMAN CAPITAL REPORT, (2015). World Economic Forum. Available through: http://www3.weforum.org/docs/WEF_Human_Capital_Report_2015.pdf [Accessed 10.02.2018].

Ustawa z dnia 16 listopada 2016 r. o zmianie ustawy o emeryturach i rentach z Funduszu Ubezpieczeń Społecznych oraz niektórych innych ustaw (Dz.U. 2017, poz. 38).

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 477–493, DOI 10.21307/stattrans-2018-026

ANOTHER LOOK AT THE STATIONARITY OF INFLATION RATES IN OECD COUNTRIES: APPLICATION OF STRUCTURAL BREAK-GARCH-BASED UNIT ROOT TESTS

OlaOluwa S. Yaya[1]

ABSTRACT

The need to understand the stationarity property of inflation of any country is paramount in the design of monetary targeting policy. In this paper, unit root hypotheses of inflation rates in 21 OECD countries are investigated using the newly proposed GARCH-based unit root tests with structural break and trend specifications. The results show that classical tests over-accept unit roots in inflation rates, whereas these tests are not robust to heteroscedasticity. As it is observed from the pre-tests, those tests with structural break reject more null hypotheses of unit roots of most inflation series than those without structural breaks. By applying variants of GARCH-based unit root tests, which include those with structural breaks and time trend regression specifications, we find that unit root tests without time trend give most rejections of the conventional unit roots. Thus, care should be taken while applying variants of the new unit root tests on weak trending time series as indicated in this work. Batteries of unit root tests for discriminating between stationarity and nonstationarity of inflation rates are recommended, since in the case of over-differenced series, wrong policy decision will be made, particularly when inflation series is considered in a cointegrating relationship with other variables.

Key words: heteroscedasticity, inflation rate, structural breaks, unit root, OECD countries.

1. Introduction

Various unit root tests are documented in time series econometrics for pre-testing series stationarity before further model estimation. The first test is the Dickey-Fuller (DF) unit root test of Fuller (1976) and Dickey and Fuller (1979), which assumed serial un-correlation of the first differences of the time series, whereas first differences of most time series are serially correlated. The augmented component was added to the test regression model to control for the serial correlation. Augmented Dickey Fuller (ADF) (see Dickey and Fuller, 1981) and Phillips-Perron (PP) (Phillips and Perron, 1988) unit root tests were proposed simultaneously to control for serial correlation in the testing frameworks. Other

[1] Economic and Financial Statistics Unit, Department of Statistics, University of Ibadan, Nigeria. E-mail: os.yaya@ui.edu.ng; o.s.olaoluwa@gmail.com.

unit root tests of similar testing procedures are the GLS-detrended Dickey-Fuller (Elliot, Rothenberg, and Stock, 1996), Kwiatkowski, Phillips, Schmidt and Shin (KPSS, 1992), Schmidt and Phillips (S-P, 1992) Elliott, Rothenberg and Stock (ERS, 1996) and Ng and Perron (NP, 2001) unit root tests.[2] All these tests are limited in the sense that they lack power in the presence of structural breaks, which is often the case in economic and financial series.

These series, at times, are stationary around a deterministic time trend, which has undergone a permanent structural shift. Perron (1989), therefore, observed that failure of the unit root tests to account for existing structural breaks could lead to serious bias and lead to false acceptance of unit root hypothesis in the usual ADF testing framework. Thus, an exogenous structural break dummy is allowed in the ADF test regression to control for the effect of the break as detailed in Perron (1989) unit root test. Similarly, Zivot and Andrews (Z-A, 1992), Lumsdaine and Papell (LP, 1997), Lee and Strazicich (LS, 2003), Perron (2006) proposed other versions of the structural break unit root test, which allowed for one or more structural breaks to be determined along with the unit root decision. The Zivot-Andrews (Z-A) unit root test allows for only one endogenous structural break to be determined from the data. Both LP and LS unit root tests were developed by extending the endogenous structural break of Z-A (1992) test to allow for two endogenous structural breaks, whereas these two unit root tests are still gaining their popularity among other structural break unit root tests. A more popular unit root test which allows for one endogenously determined structural break at trend and intercept levels is the Perron (2006) unit root test, developed by extending the work of Ng-Perron (2001) unit root. The test considers both innovative and additive outlier-break types.

All these unit root tests are still lacking in their inability to capture a very salient property of economic and financial series at different time frequencies. Although, the application of ADF unit root test remains regardless of the time frequency of the data, however, when the data at hand are daily, weekly or monthly frequencies, it is not appropriate to use white noise assumption for the ADF-type tests in order to avoid size distortion problem. Thus, series of misleading inferences might have been made on data such as oil price, stocks, inflation, exchange rate, bonds, among others, since these tend to exhibit heteroscedasticity of any form. This observation was first documented in Kim and Schmidt (1993) and examined by Ling, Li and McAleer (2003) and Cook (2008). These heteroscedasticity-robust unit root tests are classified as Generalized Autoregressive Conditional Heteroscedasticity (GARCH)-based unit root tests, which allow for a GARCH process in the DF test regression instead of the white noise error process in the ADF-type unit root tests. Cook (2008) based the GARCH-based unit root test on the initial work of Kim and Schmidt (1993) and Haldrup (1994), whereas these tests have their shortcomings in their inability to account simultaneously for structural breaks, which is a major concern in high frequency economic and financial data. Using Cook (2008) unit root test (and others) in the presence of structural breaks could lead to wrong inference. Following the initial work of Narayan and Popp (NP, 2010) on structural break unit root test, three other new structural break-GARCH-based unit root tests are proposed: a two-exogenously determined structural break-GARCH-based unit

[2] Ng-Perron, (2001) unit root test is a modification of Philips-Perron testing procedure.

root test by Narayan and Liu (2011, NL); a two-endogenously determined structural break-trend-GARCH-based unit root test by Narayan and Liu (2015), and a two-endogenously determined structural break-GARCH-based unit root test by Narayan, Liu and Westerlund (2016, NLW).

Specifically in this paper, we re-investigate unit root hypothesis of inflation rates using structural break-GARCH-based unit root tests, with data from among the Organization of Economic Cooperation and Development (OECD) countries. We consider the OECD list since this will allow us a poll of inflation rates of countries of world's interest, with number of variables large enough to re-investigate unit root tests. Secondly, each of the time series is long enough to provide reliable estimates. As part of the strategy, we also carry out robustness checks on the tests by varying the size of the GARCH process.

There is the need to understand and judge correctly the stationarity property of inflation of any country since inflation targeting has been one of the contents of monetary targeting policy designed by the central banks over few years (Chang, Ranjbar and Tang, 2013). For example, if inflation follows I(1) process, then shocks affecting the series will have permanent effects, thereby shifting inflation from one equilibrium level to another. Policy makers now require a very strong decision to revert inflation rates to its original level. In the alternative, if inflation is stationary I(0) process, the effects of the shocks will be temporary, and it will be easier for the policy makers to revert inflation rates to its original level. Stationary or mean reverting inflation rates implies that inflation incurs a lower cost in pursuing monetary policies by the concerned agency (Cecchetti and Debelle, 2006). Actually, stationarity/nonstationarity of the inflation rate is controversial. Some authors are of the opinion that this series follow I(0) the process based on the fact that consumer price index (price process), the generating process is I(1). Other authors are of the opinion that the series is nonstationary I(1) process, and should be included in cointegrating relationships (Gil-Alana, Shittu and Yaya, 2012).

The empirical investigation of review on unit root hypothesis considered in this paper starts from the account of Culver and Papell (1997). These authors applied the sequential trend break and panel data modelling to investigate unit root hypothesis in inflation rates of 13 OECD countries and found rejection for unit root of four inflation rates based on individual country tests, but on applying panel data modelling, unit root hypothesis were rejected in all the 13 inflation rates. As a follow-up, Basher and Westerlund (2008) applied a more powerful panel unit root tests to inflation rates in OECD countries and obtained evidence for stationarity of inflation rates. Romeo-Avila and Usabiaga (2009) investigated unit root in inflation rates of 13 OECD countries taking into consideration cross-sectional dependence and mean shifts over the periods 1957 to 2005 using panel unit root test. Their results point to stationarity of inflation rates once mean shifts in the time series are considered. Gregoriou and Kontonikas (2009) applied ADF and Ng-Perron tests to five OECD countries inflation rates and found ADF accepting the null hypothesis of unit root in inflation rates in all five countries while Ng-Perron test rejected the null hypothesis of unit root in two of the countries. Narayan and Narayan (2010) tested unit root hypothesis in 17 inflation rates from OECD countries using conventional unit root tests and the KPSS univariate test without structural breaks. The results obtained indicate non-rejection of unit root hypothesis in all the 17 inflation series, while with KPSS test rejection of hypothesis of unit root was observed in 10 of the cases. Further investigation using panel unit root test reveals strong evidence of the inflation rate for panels of the countries which are earlier picked to be nonstationary. Narayan and Popp (2011) applied modified seasonal unit root test with seasonal mean shifts proposed by Popp (2007) to inflation in G7 countries and found that none of the countries possessed seasonal unit root at monthly and annual frequencies, whereas a semi-annual unit root is found in the case of Germany. Noriega, Capistran and Ramos-Francia (2013) studied inflation persistence in 45 countries between 1960 and 2008 using a test for multiple changes in persistence and obtained mixed results on I(1)/I(0) dynamics of inflation in the countries. Lee (2015) considered unit root testing on inflation rates of 20 OECD countries using panel unit root test, taking into account cross-sectional dependence and smoothing structural changes of unknown form by the Fourier function. The ADF and other classical unit root tests indicated rejection of fewer null hypothesis of unit roots of inflation series, while on applying the panel unit root tests, all the null hypotheses of unit roots for inflation series for the 20 countries were rejected. Chang, Ranjbar and Tang (2013) applied a flexible Fourier stationary test to investigate mean reversion of inflation rates in 22 OECD countries between 1961 and 2011 and obtained evidence of mean reversion in all the countries, contrary to the mixed results obtained by the classical unit root tests. Zhou (2013) applied nonlinearity-based unit root testing procedure to examine the stationarity of inflation rates of 12 European countries that form the Euro-zone. The results obtained showed that classical unit root test hardly rejects the null hypothesis of unit root due to the fact that the time series are characterized with nonlinearity which needs to be considered during the testing procedure. Upon applying the nonlinearity-based unit root test, 10 of the 12 inflation rates appear to be stationary. Gil-Alana, Yaya and Solademi (2016) considered inflation rates in a group of 7 countries and investigated unit roots hypothesis based on classical tests and fractional persistence approach with structural breaks and nonlinearity. The results obtained first indicated mixed results by the ADF, PP and Kapetanios Schmidt and Shin (KSS) tests, while upon applying the fractional persistence approach, the results showed evidence of unit roots in the cases of the UK, Canada, France, Italy, Japan and the USA, and evidence of mean reversion in the case of Germany. Canarella and Miller (2017) investigated the dynamics of inflation persistence in Canada, Sweden, UK, Chile, Isreal and Mexico in order to establish cointegrating relationship of inflation in each country with inflation rates in Germany and the US. The authors obtained mixed results such that inflation rates in advanced countries (Canada, Sweden, UK, Germany and the US) are fractionally integrated in stationary mean reverting ranges, while inflation rates for Chile, Isreal and Mexico are nonstationary mean reverting.

Due to the importance of the decision of unit root tests in econometric time series modelling, there is the need to apply a more robust unit root tests, developed for specific time series areas. As noted, economic and financial time series often exhibits structural breaks and nonlinearity in the form of heteroskedasticity, thus this calls for new testing procedure other than the well-known ADF unit root tests. The GARCH-based and structural break-GARCH-based unit root testing frameworks are very new testing procedures. The basis for

GARCH-based unit root test is found in Kim and Schmidt (1993), who considered the first GARCH-type heteroscedasticity unit root tests for time series. Then, Haldrup (1994), Ling, Li and McAleer (2003) and Cook (2008) applied the framework in testing unit roots in heteroscedasticity-based time series, whereas the testing procedure is lacking in its ability to investigate simultaneously structural break and unit root in the series as in Perron (2006) and NP (2010) for the case of homoscedasticity.

Narayan and Liu (NL, 2011) presented the first structural break-GARCH-based unit root test which accommodates two structural breaks in the heteroscedastic time series. This procedure lays the foundation for Narayan's structural break GARCH-based unit root frameworks and empirical applications of these tests are found in Salisu and Mobolaji (2013), Salisu and Fasanya (2013) and Mishra and Smyth (2014). NL (2015) and NLW (2016) were developed from NL (2011). NL (2015) include both intercept and time trend in the structural break-GARCH-based unit root test, while in NLW (2016), time trend is absent while only constant is included. Other applications to NL (2015) and NLW (2016) are found in Salisu and Adeleke (2016) and Salisu et al. (2016).

None of the empirical works on the newly proposed unit root testing procedures have been applied on inflation rates. Recent developments suggest that conflicting decisions often emerge while testing stationarity of inflation dynamics, particularly inflation rates from developed and emerging non-African economies such as the G7, BRICS and the OECD countries.

The rest of the paper is therefore structured as follows: Section 2 presents the data and pre-test analyses, which include testing for trend, heteroscedasticity, structural breaks. Section 3 presents the structural break GARCH-based unit root tests considered in the paper. Section 4 presents the results of the unit root tests, while Section 5 renders the concluding remarks.

2. The Data and Pre-test

Monthly time series of 21 inflation rates applied in this paper are sourced from Main Economic Indicators of the Organization of Economic Cooperation and Development (OECD), available at https://data.oecd.org/price/inflation-cpi.htm. These countries are Austria, Belgium, Brazil, Canada, Denmark, Finland, France, Greece, Hungary, India, Ireland, Israel, Italy, Japan, Korea, Luxemburg, Netherlands, Norway, South Africa, UK and the USA. For convenience, we have used initial classification to rename these countries, for example, Austria (AUS). Data identification and coverage for each of the time series is presented in Table 1.

Table 1. Data identification and Coverage

Country	Inflation initial	Start date	End date
Austria	AUS	1967M01	2016M10
Belgium	BEL	1956M01	2016M10
Brazil	BRA	1980M12	2016M09
Canada	CAN	1950M01	2016M10
Denmark	DEN	1956M01	2016M10

Table 1. Data identification and Coverage (cont.)

Country	Inflation initial	Start date	End date
Finland	FIN	1956M01	2016M10
France	FRA	1956M01	2016M10
Greece	GRE	1956M01	2016M10
Hungary	HUN	1981M01	2016M10
India	IND	1958M01	2016M09
Ireland	IRE	1956M01	2016M10
Israel	ISR	1956M01	2016M10
Italy	ITL	1956M01	2016M10
Japan	JPN	1971M01	2016M09
Korea	KOR	1952M08	2016M10
Luxemburg	LUX	1956M01	2016M10
Netherlands	NLD	1960M01	2016M10
Norway	NOR	1956M01	2016M10
South Africa	XAF	1958M01	2016M09
UK	UK	1956M01	2016M10
USA	US	1956M01	2016M10

Note: Determined by the authors.

Descriptive statistics computed from the inflation series are presented in Table 2. These analyses include mean, maximum, minimum, standard deviation, skewness, kurtosis and Jarque-Bera (JB) test for normality. The average highest inflation rate is recorded for Brazil, while many countries in the Euro-zone, in Canada and in the USA indicated average inflation rates of 3 to 4%. Most of these countries with low inflation rates present negative minimum inflation of about -2% and maximum inflation rates of two-digit of less than 30% on the average. Based on the estimates of standard deviation, it is quite obvious to observe fluctuations in inflation time series of the countries considered. Estimates of skewness are positive in all the cases, with skewness values above 0.5 in 19 of the cases. Thus, the distributions of the time series are positively skewed in these 19 cases. Platykurtosis is also observed in 19 different cases, while leptokurtosis is observed in the remaining two cases (GRE and XAF). Generally, estimates of JB test conclude that the null hypothesis of normality of inflation rates should be rejected in all the cases.

Next, we report the presence of heteroscedasticity based on ARCH test, and we found rejection of null hypothesis of homoscedasticity in all the cases. Thus, it implies that conditional heteroscedasticity is present in the time series. This also further explains the need for GARCH process in the unit root testing frameworks.

Since the unit root testing frameworks considered in this work applied trend and structural break as part of the testing procedures, we estimate both ‘Trend’ and ‘Trend1’ as presented in the last two columns of Table 2. The significance of trend coefficient in ‘Trend’ implies the consideration of trend in the unit root testing procedure. Similarly, ‘Trend1’ includes the two dummy variables D1 and D2 for the two break dates T1 and T2 as determined based on Bai-Perron (BP) multiple structural breaks test results presented in Table 4. All the trend coefficients under ‘Trend’ are significant, while in ‘Trend1’ column, the trend coefficients computed for BRA, LUX and NLD are not significant.

We present in Table 3 the results of classical unit root tests for non-structural break-based and structural break-based unit root tests. Starting with the results of DF, ADF, PP, S-P and Ng-Perron unit root tests, we observe consistency in the stationarity decision of these unit root tests for AUS, BRA, IND, JPN and KOR on significance of at least two of the unit roots for each of the series. Thus, these series seem to be stationary-based on these tests. In the overall, these five unit root tests were able to reject null hypothesis of unit roots at 4, 4, 6, 4 and 6 cases of each of the tests, respectively. Looking at the results of 1-structural break unit root tests by Zivot and Andrew (Z-A) and Perron 2006, we observed more rejections of unit roots in the inflation rates. The five inflation series observed to be stationary based on DF, ADF, PP, S-P and Ng-Perron are also found to be stationary based on these structural break-based unit root tests. Based on the results of 2-structural break unit root test of NP (2010) with M1 test model, we observed 8 rejections of null hypothesis of unit root of inflation series, and the decisions reached were different from those obtained by the classical unit root tests, even though there are consistencies with those results obtained based on 1-structural break unit root tests. Due to the fact that the inclusion of structural break in the unit root testing procedure increased the rejection rates of the unit roots in time series, it implies that ADF and other similar tests over-accept unit roots in the presence of structural breaks in the time series.

Table 2. Descriptive Statistics Inflation rates

Country	Inflation Rate	Mean	Maximum	Minimum	S.D.	Skewness	Kurtosis	JB	ARCH(5)	Trend	Trend1
Austria	AUS	3.6415	26.9113	-15.3006	7.2044	0.3944	3.9313	37.1156***	500.159***	-0.0097***	0.0064***
Belgium	BEL	3.4548	16.3127	-1.6809	2.8138	1.6722	6.6653	748.855***	711.190***	-0.0038***	-9.9E-04***
Brazil	BRA	370.47	6821.32	1.6454	934.75	4.0091	21.244	7115.07***	380.906***	-2.1210***	0.1644
Canada	CAN	3.6338	13.0081	-2.1127	3.1844	1.1878	3.6686	203.535***	755.520***	-0.0022***	-0.0045***
Denmark	DEN	4.5340	16.8290	-1.2884	3.6909	1.0782	3.3631	145.460***	671.389***	-0.0078***	-0.0045***
Finland	FIN	5.0010	19.2405	-1.5511	4.4027	1.1082	3.6436	162.018***	682.474***	-0.0112***	-0.0118***
France	FRA	4.4451	18.7812	-0.7253	3.9606	1.2677	3.8198	215.975***	710.973***	-0.0093***	-0.0059***
Greece	GRE	8.1507	33.8028	-2.8523	8.0860	0.9006	2.7708	100.287***	679.202***	-0.0040***	-0.0154***
Hungary	HUN	10.635	38.5766	-1.4793	8.9045	1.0973	3.3552	88.5458***	413.031***	-0.0325***	-0.0343***
India	IND	7.5915	34.6422	-11.287	5.4557	0.7122	6.8442	493.700***	656.765***	-0.0020***	0.0056***
Ireland	IRE	5.4606	24.1542	-6.5637	5.5742	1.3450	4.5622	294.331***	672.813***	-0.0093***	-0.0047***
Israel	ISR	15.605	102.174	-8.2569	17.916	1.7653	6.1027	671.979***	697.151***	-0.0210***	-0.0136***
Italy	ITL	5.7936	25.2351	-2.0140	5.5538	1.4447	4.2386	300.589***	708.034***	-0.0075***	-0.0123***
Japan	JPN	2.6448	24.9000	-2.5000	4.5150	2.5982	11.124	2127.26***	517.569***	-0.0189***	-0.0041***
Korea	KOR	11.380	122.100	-11.984	16.270	3.7056	20.245	11318.6***	665.605***	-0.0394***	-0.0172***
Luxemburg	LUX	3.2441	11.8099	-1.4192	2.6814	1.1507	3.9317	187.515***	691.650***	-0.0028***	7.8E-05
Netherlands	NLD	3.3544	11.1029	-2.7601	2.6287	0.8440	3.1919	82.0137***	632.191***	-0.0063***	-0.0008
Norway	NOR	4.5201	15.1194	-1.8341	3.3072	0.9192	3.1048	103.126***	665.700***	-0.0058***	-0.0031***
South Africa	XAF	7.8949	20.9424	-1.9993	4.9328	0.2860	2.0516	36.0325***	621.500***	0.0021***	0.0035***
UK	UK	5.1252	26.8670	-0.8167	4.9468	1.9945	7.1307	1003.001***	709.280***	-0.0075***	-0.0060***
USA	US	3.7033	14.7565	-2.0972	2.8336	1.5811	5.6511	517.917***	710.941***	-0.0028***	-0.0031***

Descriptive statistics such as mean, standard deviation (S.D.), maximum, minimum, skewness and kurtosis values are reported. Test of normality by Jarque-Bera (JB) test is presented. The ARCH test is the test of homoscedasticity of the time series against possible heteroscedasticity. 'Trend' is an OLS regression model with time trend only. 'Trend1' is an OLS regression with trend and structural break dummies D1 and D2 for T1 and T2 obtained based on Bai-Perron multiple structural break tests presented in Table 4 below.

*** *indicates significance of all the tests as well as that of trend term at 5% level.*

Table 3. Results of Non-GARCH-based unit root tests

Country	Inflation initial	Non-Structural break Unit root tests					1-Structural break Unit root tests		2- Structural break Unit root tests
		DF	ADF	PP	S-P	Ng-Perron	Z-A	Perron	NP 2010 M1
Austria	AUS	-3.49[12]***	-3.78[12]***	-4.41[8]***	-4.18[18]***	-28.54[12]***	-5.56[4]***	-4.82[12]	-3.92
Belgium	BEL	-2.13[16]	-2.55[16]	-2.99[12]	-2.23[18]	-10.30[16]***	-4.07[4]	-4.65[16]	-3.77
Brazil	BRA	-4.07[2]***	-4.40[2]***	-3.95[8]***	-3.20[17]***	-33.85[2]***	-7.57[4]***	-10.46[17]***	-3.51
Canada	CAN	-2.16[12]	-2.35[12]	-3.32[15]	-2.76[18]	-7.06[12]	-4.83[4]	-5.09[12]	-3.79
Denmark	DEN	-2.69[13]	-2.62[13]	-3.17[7]	-2.69[18]	-12.78[13]***	-6.06[4]***	-5.96[13]***	-5.39***
Finland	FIN	-2.54[13]	-2.61[13]	-3.23[14]	-2.80[18]	-12.56[13]	-4.71[4]	-4.33[13]	-4.16***
France	FRA	-1.56[13]	-2.96[13]	-3.27[16]	-2.35[18]	-5.32[13]	-5.97[4]***	-4.67[14]	-3.34
Greece	GRE	-1.56[14]	-1.70[14]	-2.17[14]	-2.03[18]	-5.29[14]	-5.48[4]***	-5.16[14]	-3.96
Hungary	HUN	-1.31[2]	-2.08[2]	-2.10[7]	-1.66[17]	-3.73[2]	-4.18[4]	-4.14[17]	-1.85
India	IND	-4.32[13]***	-4.64[13]***	-4.77[15]***	-4.27[18]***	-52.79[13]***	-5.56[4]***	-5.80[19]***	-3.13
Ireland	IRE	-1.80[12]	-2.23[12]	-2.64[12]	-2.16[18]	-7.02[12]	-4.41[4]***	-5.10[12]	-4.40***
Israel	ISR	-1.64[16]	-1.97[16]	-2.70[14]	-2.21[18]	-5.90[16]	-6.52[4]***	-5.91[16]***	-3.14
Italy	ITL	-1.06[13]	-1.70[13]	-2.23[14]	-1.80[18]	-2.61[13]	-5.10[4]***	-3.92[13]	-2.86
Japan	JPN	-2.99[12]***	-3.15[12]	-3.27[11]	-3.12[18]***	-20.10[12]***	-4.32[4]	-6.83[12]***	-2.42
Korea	KOR	-0.89[17]	-4.36[17]***	-6.09[16]***	-2.15[18]	-0.78[17]	-8.92[4]***	-9.14[15]***	-6.06***
Luxemburg	LUX	-1.55[12]	-2.17[12]	-3.08[11]	-2.30[18]	-4.97[12]	-4.36[4]	-5.74[18]***	-3.59
Netherlands	NLD	-1.12[13]	-3.28[13]	-3.49[4]***	-2.28[18]	-2.77[13]	-4.88[4]	-4.20[13]	-4.90***
Norway	NOR	-1.62[12]	-2.57[12]	-3.66[6]***	-2.67[18]	-5.14[12]	-5.32[4]***	-4.82[12]	-4.77***
S. Africa	XAF	-1.53[12]	-1.76[12]	-2.32[3]	-2.02[18]	-5.14[12]	-5.22[4]***	-4.26[12]	-4.42***
UK	UK	-2.19[14]	-2.42[14]	-2.59[15]	-2.27[18]	-10.62[14]	-5.20[4]***	-5.05[14]	-4.28***
USA	US	-1.84[13]	-2.44[13]	-2.99[11]	-2.28[18]	-7.35[13]	-4.73[4]	-6.63[16]***	-2.35
No. of rejections		4	4	6	4	6	13	8	8

Acronyms for the unit root tests: DF (Dickey-Fuller), ADF (Augmented Dickey Fuller), PP (Phillips-Perron), S-P (Schmidt-Phillips), Ng-Perron, (Z-A) Zivot-Andrews, Perron and NP(2010). Note, for PP test, the corresponding bandwidth value is in squared bracket, while for other tests the corresponding lag length for the test model information criterion is in squared bracket. Critical values for NP2010 test at 5% level of significance is 4.08, while critical levels for the other unit root tests are given in respective tables by the authors.

**** indicates significance of the tests at 5% level. The 'BLUE' denotes evidence of stationarity of inflation series in those countries by non-structural break-unit root tests, while the 'GREEN' denotes evidence of stationarity of inflation rates in those countries by structural break-unit root tests.*

Table 4 presents the results of the number of significant breaks as well as two break dates in the inflation series based on Bai-Perron multiple structural break test. Two significant break dates were identified for BRA, IND, ISR, KOR and LUX while AUS and HUN present 5 significant break dates. Since two break dates are found in all the inflation rates, the use of structural break-GARCH-based unit root tests by Cook (2008), NL(2011), NL(2015) and NLW(2016) is further justified.

Table 4. Bai and Perron (2003) multiple structural breaks test

Country	Inflation initial	T_1	T_2	NSB
Austria	AUS	1974M06	1982M08	5
Belgium	BEL	1971M09	1985M08	3
Brazil	BRA	1989M08	1994M12	2
Canada	CAN	1972M09	1983M05	3
Denmark	DEN	1973M02	1983M07	3
Finland	FIN	1973M01	1984M06	3
France	FRA	1973M09	1985M07	3
Greece	GRE	1973M06	1995M02	4
Hungary	HUN	1988M01	1998M08	5
India	IND	1972M12	1999M05	2
Ireland	IRE	1973M01	1984M07	3
Israel	ISR	1973M11	1986M04	2
Italy	ITL	1973M04	1984M10	4
Japan	JPN	1977M11	1985M02	3
Korea	KOR	1962M03	1982M05	2
Luxemburg	LUX	1970M02	1984M12	2
Netherlands	NLD	1969M01	1982M12	4
Norway	NOR	1970M01	1989M01	3
South Africa	XAF	1973M03	1993M08	4
UK	UK	1973M07	1982M08	3
USA	US	1968M06	1982M09	4

Note: NSB denotes the number of significant structural breaks in each time series. The critical values of this test for the five break dates are l = 1, 2, 3, 4, 5 are 8.58, 10.13, 11.14, 11.83, 12.25, and $\hat{T}_1$ *and* $\hat{T}_2$ *denote the two longest break sub-samples.*

3. The Structural break-GARCH-based Unit root test

Since the classical ADF test is not robust to structural breaks, and inference made based on the testing procedure is not valid in the presence of structural breaks, Narayan and Poop (NP, 2010) introduced a modified ADF-type tests, in two models, both allowing for two structural breaks. The first model, termed M1, allows for two structural breaks in the intercept of the time series only, while the second model, which is termed M2, allows to simultaneously test two breaks in the intercept and trend of the time series. Thus, M1 test model of NP (2010) is the basis for GARCH-based unit root tests applied in this paper, and this is specified as:

$$\Delta X_t^{M1} = \alpha_0 + \alpha_1 t + \delta X_{t-1} + \phi_1 DU'_{1,t-1} + \phi_2 DU'_{2,t-1} + \theta_1 D\left(T_B{}'\right)_{1,t} + \theta_2 D\left(T_B{}'\right)_{2,t} + \sum_{j=1}^{k} \alpha_{3,j} \Delta X_{t-j} + \varepsilon_t \qquad (1)$$

where X_t is the time series at time *t*; ΔX_t is the first difference of the series; ΔX_{t-j} is the lagged first differences of the series under the augmentation with parameters $\alpha_{3,i}$ ($j=1,\ldots,k$); $DU'_{i,t}=1\left(t>T'_{B,i}\right)$, $DT'_{i,t}=1\left(t>T'_{B,i}\right)\times\left(t-T_{B,i}\right)$, $t=1,2$, with $T'_{B,i}\left(i=1,2\right)$ as the break dates. The parameters, α_0 and α_1 are the intercept and coefficient of time trend, respectively, while δ determines the decision of the unit root. Thus, the null hypothesis $H_0 : \delta = 0$ for unit root is tested against the alternative hypothesis $H_1 : \delta < 0$ of no unit root with the aid of the t-statistic,

$$t_\delta = \frac{\delta}{s.e.(\delta)} \tag{2}$$

obtained from the test regression. From model M1 in (1), the parameters ϕ_i and θ_i $(i = 1, 2)$ denote the magnitude of the level and trend breaks, respectively. Similarly to ADF testing regression, the null hypothesis of a unit root is tested as $H_0 : \delta = 1$ for unit root against the alternative hypothesis $H_1 : \delta < 1$ for the test regression model.

As a result of non-normality of the residuals, which contradicts the OLS regression assumption, Narayan and Liu (NL, 2011) then proposed a GARCH-based unit root test by augmenting NP (2010) M1 test regression. Thus, the proposed NL(2011) test regression model is

$$\Delta X_t = \delta X_{t-1} + \phi_1 DU'_{1,t-1} + \phi_2 DU'_{2,t-1} + \varepsilon_t \tag{3}$$

for only the level breaks $DU'_{1,t}$ and $DU'_{2,t}$ for the break dates T_1 and T_2, respectively in the time series with

$$\varepsilon_t = \sigma_t z_t, \quad z_t \approx N(0,1), \tag{4}$$

$$\sigma_t^2 = a + \sum_{i=1}^{p} b_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} c_j \sigma_{t-j}^2, \tag{5}$$

where $b_i \left(i = 1, \ldots, p\right)$ and $c_j \left(j = 1, \ldots, q\right)$ are non-negative parameter values, and a is a strictly positive constant. The residual process ε_t is obtained as products of conditional standard deviation, σ_t and standardized normal variable, z_t. Next, by including both intercept and a time trend to the test regression model of NL(2011) in (3), together with (6) and 7), the model becomes

$$\Delta X_t = \alpha_0 + \alpha_1 t + \delta X_{t-1} + \phi_1 DU'_{1,t-1} + \phi_2 DU'_{2,t-1} + \varepsilon_t. \tag{6}$$

which is a testing procedure proposed for modelling trending series in NL(2015). Actually, the authors found that this testing procedure outperforms NL(2011) and Cook (2008) GARCH-based unit root tests. In the case of non-trending/weak trending series, Narayan, Liu and Westerlund (NLW, 2016) silenced the trend component in NL(2015) test regression. Thus, the model included only the intercept:

$$\Delta X_t = \alpha_0 + \delta X_{t-1} + \phi_1 DU'_{1,t-1} + \phi_2 DU'_{2,t-1} + \varepsilon_t . \tag{7}$$

Cook (2008) GARCH-based unit root test regression is obtained from (7) by excluding the structural break components to obtain

$$\Delta X_t = \alpha_0 + \delta X_{t-1} + \varepsilon_t . \tag{8}$$

The scope of the work was further extended by carrying out robustness checks by varying the orders of the GARCH (p,q) model as (1,2), (2,1) and (2,2).

4. GARCH-based Unit root results

The results for GARCH-based unit root tests discussed above are presented in Tables 5, 6a and 6b. Table 5 presents the standard tests based on GARCH(1,1) process, while Tables 6a and 6b present the robustness tests by varying the orders of GARCH model. From Table 5, we observe unit root rejection rates for 11, 14, 11 and 10 inflation rates corresponding to Cook (2008), NL(2011), NL(2015) and NLW(2016), respectively, and we further observed similar decision on the rejection of unit roots, which include five inflation rates (AUS, BRA, IND, JPN and KOR) picked to be stationary by classical unit root tests. We further observed similar decision on unit roots of five inflation rates (AUS, BRA, IND, JPN, KOR) based on the classical tests. Generally, in all the four GARCH-based unit root tests, null hypothesis of unit roots were rejected in the cases of BEL, BRA, DEN, IND, IRE, ISR, KOR and NOR.

Table 5. Results of GARCH-based unit root tests

Country	Inflation Initial	Cook (2008)	NL (2011)	NL (2015)	NLW (2016)
Austria	AUS	-3.53***	-3.20***	-3.29	-3.38
Belgium	BEL	-3.00***	-2.89***	-4.10***	-4.12***
Brazil	BRA	7.10***	6.85***	9.11***	9.11***
Canada	CAN	-2.21	-2.48	-3.90***	-3.56
Denmark	DEN	-8.20***	-12.75***	-8.60***	-7.39***
Finland	FIN	-2.14	-2.72	-3.03	-2.65
France	FRA	-4.26***	-1.90	-7.20***	-4.92***
Greece	GRE	-2.24	-3.14***	-3.82	-2.87
Hungary	HUN	-1.98	-3.97***	-3.25	-1.76
India	IND	-3.54***	-3.64***	-4.21***	-4.25***
Ireland	IRE	-3.81***	-3.72***	-4.06***	-5.34***
Israel	ISR	-2.69	-4.29***	-4.99***	-5.14***
Italy	ITL	-3.00***	-0.93	-3.07	-2.18
Japan	JPN	-3.48***	-3.83***	-4.09***	-3.61
Korea	KOR	-3.76***	-6.05***	-5.74***	-4.95***
Luxemburg	LUX	-2.28	-3.52***	-3.65	-3.61

Table 5. Results of GARCH-based unit root tests (cont.)

Country	Inflation Initial	Cook (2008)	NL (2011)	NL (2015)	NLW (2016)
Netherlands	NLD	-2.11	-2.64	-3.33	-3.20
Norway	NOR	-2.97***	-3.23***	-4.17***	-4.23***
South Africa	XAF	-2.60	-3.39***	-3.82	-3.78***
UK	UK	-2.64	-2.35	-2.67	-2.74
USA	US	-2.40	-1.75	-2.98	-2.86
No. of rejections		11	14	11	10

Note: For consistency, we only made inference on the tests at 5% significant levels. Thus, critical values for Cook (2008), NL(2011), NL(2015) and NLW(2016) unit root tests are -2.86, -2.87, -3.89 and -3.66, respectively.

**** denotes statistical significance of the unit root tests.*

Consistency and robustness of the unit root test are investigated by varying the lag lengths of the GARCH process as GARCH(1,2), GARCH(2,1) and GARCH(2,2). These results are presented in Tables 6a [Cook (2008) and NL(2011)] and Tables 6b [NL(2015) and NLW(2016)]. A critical look at the results indicates quite much consistency in the decision of the unit root tests based on NL(2011) unit root test. Recall that this testing regression does not include constant and time trend. Thus, NL(2011) test exhibits more robustness to lag lengths than any other GARCH-based unit root test since it appears to be insensitive to the lag order of the symmetric GARCH model. Based on consistency, Cook (2008) also outperformed the other remaining two GARCH-based unit root tests.

Table 6a. Robustness tests

Cook (2008)					NL (2011)				
Country	Inflation Initial	GARCH (1,2)	GARCH (2,1)	GARCH (2,2)	Country	Inflation Initial	GARCH (1,2)	GARCH (2,1)	GARCH (2,2)
Austria	AUS	-3.79***	-3.90***	-4.02***	Austria	AUS	-3.41***	-3.47***	-3.57***
Belgium	BEL	-3.01***	-3.03***	-3.03***	Belgium	BEL	-3.01***	-3.02***	-3.01***
Brazil	BRA	-0.04	76.11***	1.33	Brazil	BRA	6.38***	5.02***	5.06***
Canada	CAN	-2.20	-2.49	-2.43	Canada	CAN	-2.14	-2.64	-2.56
Denmark	DEN	-2.32	-3.65***	-5.80***	Denmark	DEN	-12.28***	-8.68***	-7.32***
Finland	FIN	-2.15	-2.11	-2.17	Finland	FIN	-2.95***	-2.84	-2.98***
France	FRA	-3.04***	-2.54	-2.55	France	FRA	-1.68	-1.50	-1.02
Greece	GRE	-2.55	-2.32	-2.21	Greece	GRE	-3.07***	-2.69	-2.58
Hungary	HUN	-2.09	-2.13	-2.33	Hungary	HUN	-4.25***	-4.81***	-4.40***
India	IND	-3.57***	-3.58***	-3.57***	India	IND	-3.60***	-3.58***	-3.58***
Ireland	IRE	-3.90***	-0.52	-29.71***	Ireland	IRE	-4.10***	-3.00	-1.47
Israel	ISR	-2.57	-2.69	-2.56	Israel	ISR	-4.23***	-5.02***	-5.80***
Italy	ITL	-2.95***	-3.16***	-3.15	Italy	ITL	-0.85	-0.77	-0.73
Japan	JPN	-3.50***	-3.50***	-3.52***	Japan	JPN	-3.77***	-3.80***	-3.79***
Korea	KOR	-3.76***	-3.71***	-3.72***	Korea	KOR	-6.00***	-5.91***	-5.90***
Luxemburg	LUX	-2.06	-1.95	-1.83	Luxemburg	LUX	-3.39***	-3.13	-2.72
Netherlands	NLD	-2.07	-2.04	-1.98	Netherlands	NLD	-2.64	-2.61	-81.83***

Table 6a. Robustness tests (cont.)

Cook (2008)					NL (2011)				
Country	**Inflation Initial**	**GARCH (1,2)**	**GARCH (2,1)**	**GARCH (2,2)**	**Country**	**Inflation Initial**	**GARCH (1,2)**	**GARCH (2,1)**	**GARCH (2,2)**
Norway	NOR	-2.87***	-2.88***	-3.08***	Norway	NOR	-3.20***	-3.22***	-3.17***
South Africa	XAF	-2.50	-2.29	-2.22	South Africa	XAF	-3.40***	-3.54***	-3.57***
UK	UK	-2.83***	-2.71	-2.91***	UK	UK	-0.49	-0.74	-0.41
USA	US	-2.39***	-2.56	-2.51	USA	US	-1.82	-1.96	-1.95
No. of rejections		11	9	9	No. of rejections		15	11	13

*Note: For consistency, we only made inference on the tests at 5% significant levels. Thus, critical values for Cook (2008), NL(2011), NL(2015) and NLW(2016) unit root tests are -2.86, -2.87, -3.89 and -3.66, respectively. Statistical significance of the test is therefore denoted by ***. Thus, decision on the stationarity of inflation series is reached based on rejection of at least three null hypotheses of the four tests, at 5% level of significance, and these rejections always included that of NL(2015) test.*

Table 6b. Robustness tests (cont'd)

NL (2015)					NLW (2016)				
Country	**Inflation Initial**	**GARCH (1,2)**	**GARCH (2,1)**	**GARCH (2,2)**	**Country**	**Inflation Initial**	**GARCH (1,2)**	**GARCH (2,1)**	**GARCH (2,2)**
Austria	AUS	-3.61	-3.64	-3.77	Austria	AUS	-3.60	-3.63	-3.76***
Belgium	BEL	-4.10***	-4.12***	-4.11***	Belgium	BEL	-4.12***	-4.14***	-4.13***
Brazil	BRA	-4.59***	-1.33	15.55***	Brazil	BRA	2.30	9.48***	-2.57
Canada	CAN	-3.49	-4.12***	-4.04***	Canada	CAN	-3.12	-3.74***	-3.66***
Denmark	DEN	-8.51***	-6.49***	-5.56***	Denmark	DEN	-5.70***	-4.72***	-2.77
Finland	FIN	-3.00	-2.84	-2.98	Finland	FIN	-2.61	-2.42	-2.69
France	FRA	-5.99***	-5.84***	-5.92***	France	FRA	-4.04***	-3.53	-3.68***
Greece	GRE	-4.01***	-3.52	-3.27	Greece	GRE	-3.16	-2.77	-2.55
Hungary	HUN	-3.41	-1.20	-2.09	Hungary	HUN	-1.87	-2.12	-2.70
India	IND	-4.16***	-4.15***	-4.15***	India	IND	-4.23***	-4.22***	-4.22***
Ireland	IRE	-4.39***	-1.01	-3.12	Ireland	IRE	-3.77***	-1.55	-5.40***
Israel	ISR	-5.16***	-5.08***	-5.17***	Israel	ISR	-5.30***	-5.21***	-5.30***
Italy	ITL	-2.88	-2.87	-2.86	Italy	ITL	-2.08	-2.13	-2.11
Japan	JPN	-4.25***	-4.72***	-4.68***	Japan	JPN	-3.76***	-4.10***	-4.10***
Korea	KOR	-5.68***	-5.70***	-5.68***	Korea	KOR	-4.86***	-4.78***	-4.78***
Luxemburg	LUX	-3.45	-3.40	-3.27	Luxemburg	LUX	-3.41	-3.36	-3.15
Netherlands	NLD	-3.23	-3.04	-3.39	Netherlands	NLD	-3.10	-2.93	-4.30***
Norway	NOR	-4.15***	-4.10***	-4.67***	Norway	NOR	-4.22***	-4.16***	-4.71***
South Africa	XAF	-3.68	-3.56	-3.52	South Africa	XAF	-3.61	-3.45	-3.40
UK	UK	-2.65	-2.86	-2.70	UK	UK	-2.73	-2.93	-2.79
USA	US	-2.95	-3.14	-3.11	USA	US	-2.85	-3.05	-3.02
No. of rejections		11	9	10	No. of rejections		9	9	11

*Note: For consistency, we only made inference on the tests at 5% significant levels. Thus, critical values for Cook (2008), NL(2011), NL(2015) and NLW(2016) unit root tests are -2.86, -2.87, -3.89 and -3.66, respectively. Statistical significance of the test is therefore denoted by ***. Thus, decision on the stationarity of inflation series is reached based on rejection of at least three null hypotheses of the four tests, at 5% level of significance, and these rejections always included that of NL(2015) test.*

5. Concluding remarks

The unit root hypothesis of inflation rates in 21 OECD countries was investigated using structural break GARCH-based unit root tests newly proposed in the literature. These unit root tests are the NL(2011), NL(2015) and NLW(2016) for without both intercept and trend specification, with intercept and trend specification, and the specification without trend only, respectively. These tests are based on the initial propositions of Cook (2008) for GARCH-based unit root test and NP(2010) two exogenous structural break regression test. Combining the ideas of the two strategies, NL(2011) obtained the first structural break GARCH-based unit root test.

Firstly, the pre-tests results to describe the data were obtained, the level of stationarity based on classical unit root tests, the trend, heteroscedasticity and structural break tests were determined. The results pointed to the usage of the newly proposed structural break GARCH-based unit root tests as better tests than the earlier proposed tests in the presence of heteroscedasticity and structural breaks.

It was found that classical ADF unit root test and other similar tests over-accept the null hypotheses of unit roots in inflation series in the presence of structural breaks and heteroscedasticity. However, pre-tests results indicated significant trend in the presence of structural breaks, but unit root analyses indicated that the test of NL(2011) without both intercept and trend gave the best unit root decision, with the highest number of rejections of the null hypothesis. Thus, care should be taken in applying the structural break GARCH-based unit root tests, particularly in a weak and significant trend case. Also, in the case of ADF unit root testing framework for no intercept, trend only, and intercept and trend, the three tests (NL2011, NLW2016 and NL2015) are recommended to be carried out simultaneously on a weak trended time series such as inflation rates in order to properly establish the nonstationarity/stationarity level of the series. This work still agrees with Narayan and Liu (2015) and Salisu and Adeleke (2015) in the cases of trended time series, noting that the series applied in the papers are strongly trended. To the monetary agency and other time series experts, we recommend using batteries of unit root tests on inflation rates to discriminate between stationarity and nonstationarity, since in the case of over-differenced series, wrong policy decision will be reached, particularly when the inflation rate is used in a cointegrating relationship.

REFERENCES

BAI, J., PERRON, P., (2003). Computation and Analysis of Multiple Structural Change Models. Journal of Applied Econometrics, 18, pp. 1–22.

BASHER, S. A., WESTERLUND, J., (2008). Is there really a unit root in the inflation rate? More evidence from panel data models. Applied Economics Letters, 15 (3), pp. 161–164.

BOLLERSLEV, T., (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, pp. 307–327.

CANARELLA, G., MILLER, S., (2017). Inflation targeting and inflation persistence. New evidence from fractional integration and cointegration. Journal of Economics and Business, 92, pp. 45–62.

CECCHETTI, S., DEBELLE, G., (2006). Has the inflation process changed? Economic Policy, 21 (46), pp. 311–352.

CHANG, T., RANJBAR, O., TANG, D. P., (2013). Revisiting the mean reversion of inflation rates for 22 OECD countries. Economic Modelling, 30, pp. 245–252.

COOK, S., (2008). Joint maximum likelihood estimation of unit root testing equations and GARCH processes: some finite-sample issues. Mathematics and Computers in Simulation, 77, pp. 109–116.

CULVER, S. E., PAPELL, D. H., (1997). Is there a unit root in the inflation rate? Evidence from sequential break and panel data model models. Journal of Applied Econometrics, 12, pp. 435–444.

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

DICKEY, D. A., FULLER, W. A., (1981). Distribution of the estimators for autoregressive time series with a unit root. Econometrica 49, pp. 1057–1072.

ELLIOT, G., ROTHENBERG, T. J., STOCK, J. H., (1996). Efficient tests for an Autoregressive unit root. Econometrica, 64, pp. 813–836.

FULLER, W. A., (1976). Introduction to statistical Time Series. Wiley, New York.

GIL-ALANA, L. A., SHITTU, O. I., YAYA, O. S., (2012). Long memory, Structural breaks and Mean shifts in the Inflation rates in Nigeria. African Journal of Business Management, 6 (3), pp. 888–897.

GIL-ALANA, L. A., YAYA, O. S., SOLADEMI, E. A., (2016). Testing unit roots, structural breaks and linearity in the inflation rates in the G7 countries with fractional dependence techniques. Applied Stochastic Models in Business and Industry, 32, pp. 711–724.

GREGORIOU, A., KONTONIKAS, A., (2009). Modelling the behaviour of inflation deviations from the target. Economic Modelling, 26, pp. 90–95.

HALDRUP, N., (1994). Heteroscedasticity in non-stationary time series, some Monte Carlo evidence. Statistical Papers, 35, pp. 287–307.

HUANG, H.-C., LIN, P.-C., YEH, C.-C., (2010). Price level convergence across cities? Evidence from panel unit root tests. Applied Economics Letters, 18 (1), pp. 87–93.

KIM, K., SCHMIDT, P., (1993). Unit root tests with conditional heteroscedasticity. Journal of Econometrics, 26, pp. 409–432.

KWIATKOWSKI, D., PHILLIPS, P. C. B., SCHMIDT, P., SHIN, Y., (1992). Testing the null hypothesis of stationarity against the alternative of unit root. Journal of Econometrics, 54, pp. 159–178.

LEE, J., STRAZICICH, M. C., (2003). Minimum Lagrange multiplier unit root test with two structural breaks. Review of Economic Statistics, 85, pp. 1082–1089.

LEE, Y-J., (2015). The stationarity of the inflation rate: Evidence from OECD countries. Master thesis. Institute of Economics, National Sun Yat-sen University.

LING, S., LI, W. K., MCALEER, M., (2003). Estimation and testing for unit root process with GARCH(1,1) errors: theory and Monte evidence. Econometric Reviews, 22, pp. 179–202.

LUMSDAINE, R. L., PAPELL, D. H., (1997). Multiple trend breaks and the unit-root hypothesis. Rev. Econ. Stat. 79, pp. 212–218.

MISHRA, V., SMYTH, R., (2014). Is monthly US natural gas consumption stationary? New evidence from a GARCH unit root test with structural breaks, Energy Policy, in press.

NARAYAN, P. K., NARAYAN, S., (2010). Is there a unit root in the inflation rate? New evidence from panel data models with multiple structural breaks. Applied Economics, 42 (13), pp. 1661–1670.

NARAYAN, P. K., POPP, S., (2010). A new unit root test with two structural breaks in level and slope at unknown time. Journal of Applied Statistics, 37, pp. 1425–1438.

NARAYAN, P. K., POPP, S., (2011). An application of a new seasonal unit root test to inflation. International Review of Economics and Finance, 20, pp. 707–716.

NARAYAN, P. K., LIU, R., (2011). Are shocks to commodity prices persistent? Applied Energy, 88, pp. 409–416.

NARAYAN, P. K., LIU, R., (2015). A unit root model for trending time-series energy variables. Energy Economics, 50, pp. 391–402.

NARAYAN, P. K., LIU, R., WESTERLUND, J., (2016). A GARCH model for testing market efficiency. Journal of International Financial Markets Institutions and Money, 41, pp. 121–138.

NG, S., PERRON, P., (2001). Lag length selection and the construction of unit root tests with good size and power. Econometrica, 69, pp. 1519–1554.

NORIEGA, A. E., CAPISTRAN, C., RAMOS-FRANCIA, M., (2013). On the dynamics of inflation persistence around the world. Empirical Economics, 44, pp. 1243–1265.

PERRON, P., (1989). The Great Crash, the oil price shocks and the unit root hypothesis. Econometrica, 57, pp. 1361–1401.

PERRON, P., (2006). Dealing with structural breaks. Palgrave Handbook of Econometrics, 1, pp. 278–352.

PHILLIPS, P. C. B., PERRON, P., (1988). Testing for a unit root in time series regression. Biometrika, 75, pp. 335–346.

POPP, S., (2007). Modified seasonal unit root test with seasonal level shifts at unknown time. Economics Letters, 97 (2), pp. 111–117.

ROMERO-ÁVILA, D., USABIAGA, C., (2009). The Hypothesis of a Unit Root in OECD Inflation Revisited. Journal of Economics and Business, 61, pp. 153–161.

SALİSU, A. A., FASANYA, I. O., (2013). Modelling oil price volatility with structural breaks. Energy Policy, 52, pp. 554–62.

SALISU, A. A., MOBOLAJI, H., (2013). Modelling returns and volatility transmission between oil price and US-Nigeria exchange rate. Energy Economics, 39, pp. 169–76.

SALISU, A. A., ADELEKE, A. I., (2016). Further Application of Narayan and Liu (2015) unit root model for trending time series. Economic Modelling, 55, pp. 305–314.

SALISU, A. A., NDAKO, U. B., OLOKO, T. F., AKANNI, L. O., (2016). Unit root modelling for trending stock market series. Borsa Istanbul Review, 16-2, pp. 82–91.

SCHMIDT, P., PHILLIPS, P., (1992). LM Tests for a Unit Root in the Presence of Deterministic Trends. Oxford Bulletin of Economics and Statistics, 54, pp. 257–287.

ZHOU, S., (2013). Nonlinearity and stationarity of inflation rates: Evidence from the Euro-zone countries. Applied Economics, 45, pp. 849–856.

ZIVOT, E., ANDREWS, D. W. K., (1992). Further evidence on Great Crash, the oil price shock and the unit root hypothesis. Journal of Bus. Econ. Stat. 10, pp. 251–270.

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 495–506, DOI 10.21307/stattrans-2018-027

DISCRIMINANT COORDINATES ANALYSIS IN THE CASE OF MULTIVARIATE REPEATED MEASURES DATA

Mirosław Krzyśko[1], Wojciech Łukaszonek[2], Waldemar Wołyński[3]

ABSTRACT

The main aim of the paper is to adapt the classical discriminant coordinates analysis to multivariate repeated measures data. The proposed solution is based on the relationship between the discriminant coordinates and canonical variables. The quality of these new discriminant coordinates is examined on some real data.

Key words: discriminant coordinates analysis, repeated measures data (doubly multivariate data), Kronecker product covariance structure, maximum likelihood estimates.

1. Introduction

Let us consider a case where samples originate from K groups (classes). We would often like to present them graphically, to see their configuration. However, it may be difficult to produce such a presentation even only three variables are observed. A different method must therefore be sought for presenting multidimensional data originating from multiple groups. That is the role of the discriminant coordinates (Seber (1984), p. 269). They are also sometimes called canonical variates (Krzanowski (2000), p. 370; Srivastava (2002), p. 257), but this name is misleading, because canonical variates with completely different properties occur in canonical correlation analysis. Another name used is discriminant functions (Rencher (1998), p. 202; Fujikoshi et al. (2010), p. 255) - this is inappropriate because discriminant functions are surfaces that separate K groups from one another.

The aim of the classical discriminant coordinates technique is to replace the input variables by a smaller number of independent coordinates in such a way that the separation among groups (classes) is maximum in the reduced space. In the case of two classes we obtain only one discriminant coordinate, coinciding with the well-known Fisher's linear discriminant function (Fisher (1936)). Generalization on $K > 2$ classes was shown by Rao (1948). The space of discriminant coordinates

[1]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. Faculty of Management, President Stanisław Wojciechowski Higher Vocational State School, Poland. E-mail: mkrzysko@amu.edu.pl

[2]Faculty of Management, President Stanisław Wojciechowski Higher Vocational State School, Poland. E-mail: w.lukaszonek@g.pl

[3]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. E-mail: wolynski@amu.edu.pl

is a space convenient for the use of various classification methods (methods of discriminant analysis).

In the present paper, we adapt the classical discriminant coordinates analysis to multivariate repeated measures data. Suppose that we have a sample of n objects characterized by p-variables measured in T different time points or physical conditions. Such data are referred to in the statistical literature as multivariate repeated data or doubly multivariate data. Analysis of such data is complicated by the existence of correlation among the measurements of different variables as well as correlation among measurements taken at different time points.

The proposed methods are particularly useful when the number of time points is small and the data sets are also small. Such situations concern, from example, the observation of variables in groups of territorial units, the number of which is fixed.

In the case of a large number of variables, a large number of time points and large data sets (e.g. in modern on-line economy), the alternative may be to use discriminant coordinates for functional data (see e.g. Górecki et al. (2018)).

In practice, the use of classical discriminant coordinates described in Section 2 requires the fulfillment of the condition $\max\{n_1,\dots,n_K\} > pT$, where n_i is the sample size derived from the ith group, $i = 1,\dots,K$. This condition is very restrictive and requires large samples. If it is not satisfied, then our problem can be partially solved using an existing relationship between the discriminant coordinates and canonical variables (Krzyśko (1979)). The construction of the discriminant coordinates as the canonical variables of the $\boldsymbol{X}$-space is described in Section 3. In this case the condition $n > pT + K - 1$ is required, where $n = n_1 + \dots + n_K$. Note that the condition $n > pT + K - 1$ is a condition much weaker than the condition $\max\{n_1,\dots,n_K\} > pT$, especially for a small number of groups K. If $n \leq pT + K - 1$, then we can construct the discriminant coordinates with the additional condition imposed on the covariance matrix. This construction is presented in Section 4. Section 5 illustrates the approaches presented in the paper on a real data set.

2. Classical discriminant coordinates: a review

Let us consider the multivariate discriminant problem with K groups. We observe $(\boldsymbol{X},Y)$, $\boldsymbol{X} = (\boldsymbol{X}_1,\dots,\boldsymbol{X}_p)$, $\boldsymbol{X}_i \in \mathbf{R}^T$, where $\mathrm{vec}\boldsymbol{X} \in \mathbf{R}^{pT}$ is a predictor vector, and $Y \in \{1,\dots,K\}$ is a categorical response variable representing the group membership. We are interested in predicting the class membership Y based on the p variables measured in T different time points or physical conditions. Suppose that group i has group mean vector $\boldsymbol{\mu}_i \in \mathbf{R}^{pT}$, a common (within-group) $pT \times pT$ covariance matrix $\boldsymbol{\Sigma}$ and associated group probability $q_i > 0$, $i = 1,\dots,K$. That is $\mathrm{E}(\mathrm{vec}(\boldsymbol{X})|Y = i) = \boldsymbol{\mu}_i$, $\mathrm{Var}(\mathrm{vec}(\boldsymbol{X})|Y = i) = \boldsymbol{\Sigma} > 0$, for $i = 1,\dots,K$ and $P(Y = i) = q_i > 0$, $q_1 + \dots + q_K = 1$. Discriminant coordinates are then defined to be the linear combination $U = \boldsymbol{u}' \mathrm{vec}(\boldsymbol{X})$, which maximizes the ratio of the between-group variance to the within-group variance.

Specifically, let $\boldsymbol{\Delta}$ be the between-group covariance matrix defined by

$$\boldsymbol{\Delta} = \sum_{i=1}^{K} q_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})',$$

for

$$\boldsymbol{\mu} = \sum_{i=1}^{K} q_i \boldsymbol{\mu}_i.$$

The between-group covariance matrix $\boldsymbol{\Delta}$ is nonnegative definite matrix. Then, the ratio of the between-group variance to the within-group variance is equal to

$$J(\boldsymbol{u}) = \frac{\boldsymbol{u}'\boldsymbol{\Delta}\boldsymbol{u}}{\boldsymbol{u}'\boldsymbol{\Sigma}\boldsymbol{u}}, \tag{1}$$

provided that $\boldsymbol{u} \in \mathbf{R}^{pT} \neq \mathbf{0}$.

If $\boldsymbol{u}_1$ is the vector which maximizes (1), we call the corresponding linear combination $U_1 = \boldsymbol{u}_1' \operatorname{vec}(\boldsymbol{X})$ the first discriminant coordinate. In particular, $\boldsymbol{u}_1$ can be obtained by solving

$$\max_{\boldsymbol{u} \in \mathbf{R}^{pT}} \boldsymbol{u}'\boldsymbol{\Delta}\boldsymbol{u}$$

subject to

$$\boldsymbol{u}'\boldsymbol{\Sigma}\boldsymbol{u} = 1.$$

Since $\boldsymbol{\Sigma}$ is a nonsingular matrix, then $\boldsymbol{u}_1$ is the eigenvector of $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}$ corresponding to its largest eigenvalue λ_1.

The second discriminant coordinate maximizes the measure $J(\boldsymbol{u})$ and satisfies the conditions:

$$\boldsymbol{u}_2'\boldsymbol{\Sigma}\boldsymbol{u}_2 = 1, \quad \boldsymbol{u}_1'\boldsymbol{\Sigma}\boldsymbol{u}_2 = 0.$$

Continuing this process, we can define the k-th discriminant coordinate as far as maximizing the measure $J(\boldsymbol{u})$, which must also comply with conditions:

$$\boldsymbol{u}_k'\boldsymbol{\Sigma}\boldsymbol{u}_l = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases}$$

$k,l = 1,\dots,s = \operatorname{rank}\boldsymbol{\Delta}$.

This means that the discriminant coordinates are uncorrelated and have unit variance.

The vectors $\boldsymbol{u}_k$, which maximize the measure $J(\boldsymbol{u})$, fulfill the equality

$$(\boldsymbol{\Delta} - \lambda_k\boldsymbol{\Sigma})\boldsymbol{u}_k = \mathbf{0},$$

where $\lambda_1 \geq \dots \geq \lambda_s > \lambda_{s+1} = \dots = \lambda_{pT} = 0$ are the eigenvalues of the matrix $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}$, $k = 1,\dots,s = \operatorname{rank}\boldsymbol{\Delta}$.

Note that the construction of discriminant coordinates requires knowledge of the vectors $\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_K$, prior probabilities $q_1,\ldots,q_K$, and the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$. In practice these parameters are not known, and we need to use their estimates from the sample.

Let $\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{in_i}$ be a sample derived from the ith group, where $i=1,\ldots,K$, and let $n=n_1+\cdots+n_K$. Then

$$\hat{q}_i = \frac{n_i}{n},$$

$$\hat{\boldsymbol{\mu}}_i = \bar{\boldsymbol{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathrm{vec}(\boldsymbol{x}_{ij}),$$

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{K} n_i\bar{\boldsymbol{x}}_i,$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-K}\boldsymbol{W}, \quad \text{where} \quad \boldsymbol{W}=\sum_{i=1}^{K}\boldsymbol{A}_i,\ \boldsymbol{A}_i=\sum_{j=1}^{n_i}(\mathrm{vec}(\boldsymbol{x}_{ij})-\bar{\boldsymbol{x}}_i)(\mathrm{vec}(\boldsymbol{x}_{ij})-\bar{\boldsymbol{x}}_i)',$$

$$\hat{\boldsymbol{\Delta}} = \frac{1}{K-1}\boldsymbol{B}, \quad \text{where} \quad \boldsymbol{B}=\sum_{i=1}^{K} n_i(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})',\ i=1,\ldots,K.$$

Note that the matrix $\boldsymbol{W}$ is the sum of the matrices $\boldsymbol{A}_1,\ldots,\boldsymbol{A}_K$. The matrix $\boldsymbol{A}_i$ is positive definite with probability 1 if and only if $n_i > pT$, $i=1,\ldots,K$. Then $\boldsymbol{W}$ is also positive definite if and only if $\max\{n_1,\ldots,n_K\} > pT$ (Banerjee and Roy (2004), p. 418; Giri (1996), p. 93). Therefore, if $\max\{n_1,\ldots,n_K\} > pT$, then the estimate $\hat{\boldsymbol{\Sigma}}$ of the positive definite matrix $\boldsymbol{\Sigma}$ is positive definite with probability 1, and we can use the given estimates of unknown parameters. The condition $\max\{n_1,\ldots,n_K\} > pT$ is very restrictive and requires large samples. If it is not satisfied, then the problem of correct estimation of the matrix $\boldsymbol{W}$ can be partially solved using an existing relationship between the discriminant coordinates and canonical variables.

3. The relationship between the discriminant coordinates and canonical variables

In the case where $\max\{n_1,\ldots,n_K\} \le pT$ estimates of the unknown parameters will be calculated using the relationship between discriminant coordinates and canonical variables (Krzyśko (1979)).

Let the q-dimensional vector $\boldsymbol{Y}$ be a vector of dummy variables defined as follows:

$$Y_i = \begin{cases} 1, & \text{if the matrix } \boldsymbol{X} \text{ is observed in the } i\text{th group,} \\ 0, & \text{in other cases,} \end{cases}$$

$i=1,\ldots,q=K-1$.

Let

$$\boldsymbol{Z} = \begin{bmatrix} \mathrm{vec}(\boldsymbol{X}) \\ \boldsymbol{Y} \end{bmatrix}$$

and let

$$\mathrm{Var}(\boldsymbol{Z}) = \begin{bmatrix} \mathrm{Var}(\mathrm{vec}(\boldsymbol{X})) & \mathrm{Cov}(\mathrm{vec}(\boldsymbol{X}),\boldsymbol{Y}) \\ \mathrm{Cov}(\boldsymbol{Y},\mathrm{vec}(\boldsymbol{X})) & \mathrm{Var}(\boldsymbol{Y}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix} = \boldsymbol{\Omega},$$

where $\boldsymbol{\Omega}_{21} = \boldsymbol{\Omega}_{12}'$ and $\boldsymbol{\Omega}$ is positive definite.

The estimate $\hat{\boldsymbol{\Omega}}$ of the positive definite matrix $\boldsymbol{\Omega}$ is positive definite with probability 1 if and only if $n > pT + q$, where $n = n_1 + \cdots + n_K$. If the matrix $\hat{\boldsymbol{\Omega}}$ is positive definite, then the matrices $\hat{\boldsymbol{\Omega}}_{11}$ and $\hat{\boldsymbol{\Omega}}_{22}$ are non-singular.

Let

$$\hat{\boldsymbol{\Gamma}} = \hat{\boldsymbol{\Omega}}_{11}^{-1}\hat{\boldsymbol{\Omega}}_{12}\hat{\boldsymbol{\Omega}}_{22}^{-1}\hat{\boldsymbol{\Omega}}_{21}.$$

Consider the equation

$$(\hat{\boldsymbol{\Gamma}} - r^2\boldsymbol{I})\boldsymbol{m} = \boldsymbol{0}. \tag{2}$$

Variables $V_k = \boldsymbol{m}_k' \mathrm{vec}(\boldsymbol{X})$, where $\boldsymbol{m}_k \in \mathbf{R}^{pT}$ are eigenvectors of the matrix $\hat{\boldsymbol{\Gamma}}$ satisfying equation (2), are called sample canonical variables of the $\boldsymbol{X}$-space.

The following relationships are satisfied (Krzyśko (1979)):

$$\boldsymbol{W} = n\hat{\boldsymbol{\Omega}}_{11} - (n\hat{\boldsymbol{\Omega}}_{12})(n\hat{\boldsymbol{\Omega}}_{22})^{-1}(n\hat{\boldsymbol{\Omega}}_{21}),$$
$$\boldsymbol{B} = (n\hat{\boldsymbol{\Omega}}_{12})(n\hat{\boldsymbol{\Omega}}_{22})^{-1}(n\hat{\boldsymbol{\Omega}}_{21}).$$

Thus, equation (2) is equivalent to the equation

$$(\boldsymbol{B} - \lambda\boldsymbol{W})\boldsymbol{m} = \boldsymbol{0},$$

where $\lambda = r^2(1-r^2)^{-1}$.

This means that the discriminant coordinates $U_k = \boldsymbol{u}_k' \mathrm{vec}(\boldsymbol{X})$ are proportional to the canonical variables of the $\boldsymbol{X}$-space $V_k = \boldsymbol{m}_k' \mathrm{vec}(\boldsymbol{X})$, where $\boldsymbol{Y}$ is a vector of dummy variables. Note that the condition $n > pT + q$ is a condition much weaker than the condition $\max\{n_1,\ldots,n_K\} > pT$, especially for a small number of groups $K = q+1$.

4. The special structure of the matrix $\boldsymbol{\Omega}$

If $n \leq pT + q$, then we can construct the discriminant coordinates with the additional condition that assumes that

$$\boldsymbol{\Omega}_{11} = \boldsymbol{U} \otimes \boldsymbol{V},$$

where $\boldsymbol{U} > 0, \boldsymbol{V} > 0$.

The matrix $\boldsymbol{U}$ represents the covariance between all p-variables on a given object and for a given time point. Likewise, $\boldsymbol{V}$ represents the covariance between repeated measures on a given object and for a given variable. The above covariance structure is subject to an implicit assumption that for all variables the correlation structure between repeated measures remains the same, and that covariance between variables does not depend on time and remains constant for all time points.

Estimates of the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, and thus the matrix $\boldsymbol{\Omega}_{11}$ can be obtained using the method given by Srivastava et al. (2008).

Let

$$\bar{\boldsymbol{x}}^{\star} = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j,$$

$$\boldsymbol{x}_{j,c} = \boldsymbol{x}_j - \bar{\boldsymbol{x}}^{\star},\ j = 1,\ldots,n.$$

Then, estimates of the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are obtained iteratively with a system of equations

$$\hat{\boldsymbol{U}} = \frac{1}{nT}\sum_{j=1}^{n}\boldsymbol{x}'_{j,c}\hat{\boldsymbol{V}}^{-1}\boldsymbol{x}_{j,c},$$

$$\hat{\boldsymbol{V}} = \frac{1}{np}\sum_{j=1}^{n}\boldsymbol{x}_{j,c}\hat{\boldsymbol{U}}^{-1}\boldsymbol{x}'_{j,c}.$$

In this case, the matrix $\hat{\boldsymbol{\Omega}}_{11} = \hat{\boldsymbol{U}} \otimes \hat{\boldsymbol{V}}$ is positive definite with probability 1 if and only if $n > \max(p,T)$.

Note that the fact that the matrix $\hat{\boldsymbol{\Omega}}_{11}$ is positive definite with probability 1 does not always guarantee that the matrix $\hat{\boldsymbol{\Omega}}$ is positive definite with probability 1. However, the fact that the matrix $\hat{\boldsymbol{\Omega}}_{11}$ is positive definite with probability 1 allows us to determine the discriminant coordinates on the basis of the matrix $\hat{\boldsymbol{\Gamma}}$ because then the matrix $\hat{\boldsymbol{\Omega}}_{11}^{-1}$ exists.

5. Example

The described methods were employed here to build the discriminant coordinates based on the annual data on the 38 European countries in the period 2009-2015. These countries were divided into 4 regions purposes by the United Nations Statistics Division: (1) Northern Europe, (2) Western Europe, (3) Eastern Europe, (4) Southern Europe. The list of countries used in the discriminant coordinates analysis is contained in Table 1.

We used the data published by the World Economic Forum (WEF) in its annual reports (`http://www.weforum.org`). Those are comprehensive data, describing exhaustively various socio-economic conditions or spheres of individual states. For statistical analysis, we chose 2 of 12 pillars of variables: technological readiness (consists of 4 variables) and higher education and training (consists of 6 variables). Table 2 describes the pillars used in the analysis.

Table 1: Countries included in analysis

	Country	Group		Country	Group
1	Albania (AL)	4	20	Lithuania(LT)	1
2	Austria (AT)	3	21	Luxembourg (LU)	3
3	Belgium (BE)	3	22	Macedonia FYR (MK)	4
4	Bosnia and Herzegovina (BA)	4	23	Malta (MT)	4
5	Bulgaria (BG)	2	24	Montenegro (ME)	4
6	Croatia (HR)	4	25	Netherlands (NL)	3
7	Cyprus (CY)	4	26	Norway (NO)	1
8	Czech Republic (CZ)	2	27	Poland (PL)	2
9	Denmark (DK)	1	28	Portugal (PT)	4
10	Estonia (EE)	1	29	Romania (RO)	2
11	Finland (FI)	1	30	Russian Federation (RU)	2
12	France (FR)	3	31	Serbia (XS)	4
13	Germany (DE)	3	32	Slovak Republic (SK)	2
14	Greece (GR)	4	33	Slovenia (SI)	4
15	Hungary (HU)	2	34	Spain (ES)	4
16	Iceland (IS)	1	35	Sweden (SE)	1
17	Ireland (IE)	1	36	Switzerland (CH)	3
18	Italy (IT)	4	37	Ukraine (UA)	2
19	Latvia (LV)	1	38	United Kingdom (GB)	1

Table 2: Variables used in analysis

Pillars	Variables
Technological readiness	Availability of latest technologies (X_1)
	Firm-level technology absorption (X_2)
	FDI and technology transfer (X_3)
	Internet users (X_4)
Higher education and training	Quality of the educational system (X_1)
	Quality of math and science education (X_2)
	Quality of management schools (X_3)
	Internet access in schools (X_4)
	Local availability of specialized research and training services (X_5)
	Extent of staff training (X_6)

In both cases $\max\{n_1,\ldots,n_K\} \leq pT$, and we could not use the classical discriminant coordinates algorithm. The unknown parameters were calculated using the relationship between discriminant coordinates and canonical variables described in Section 3. The technological readiness pillar consists of 4 variables ($p = 4$). In this case $n > pT + q$. The discriminant coordinates are uncorrelated. However, they are not orthogonal. In practice, however, the usual procedure is to plot discriminant coordinates on a rectangular coordinate system. The resulting distortion is generally not serious. Projection of the 38 European countries on the plane $(\hat{U}_1, \hat{U}_2)$ is presented in Figure 1.

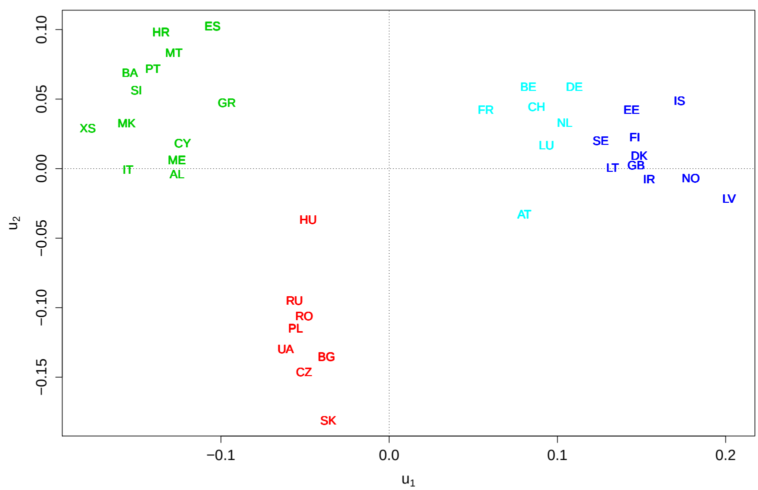


Figure 1: Technological readiness. Projection of the 38 European countries on the plane $(\hat{U}_1, \hat{U}_2)$. Regions used for statistical processing purposes by the United Nations Statistics Division: ■ – Northern Europe, ■ – Western Europe, ■ – Eastern Europe, ■ – Southern Europe

Figure 1 confirms that the European countries in terms of four characteristics of forming the technological readiness pillar are divided into four groups. However, the difference between the countries of Western Europe and the countries of Northern Europe is small.

The contribution of each variable to the discriminant coordinate is not the same. The correlation between each variable and a discriminant coordinate is widely recommended as a useful measure of variable importance in the discriminant coordinate. These correlations, sometimes called structure coefficients, are provided in many software packages. However, it turns out that these correlations do not show the multivariate contribution of each variable, but rather provide only univariate information, showing how each variable by itself separates the groups, ignoring the presence of the other variables. The better measure are the absolute values of standardized coefficients, because these coefficients show the contribution of the variables in the presence of each other, that is, these coefficients provide a multivariate interpretation that allows for the correlations among the variables (Rencher (1998), p. 214). If we denote the rth coefficient in the qth discriminant coordinate by m_{qr} then the standardized form is $m^{\star}_{qr} = s_r m_{qr}$, where s_r is the within-group standard deviation of the rth variable obtained from the diagonal of $(n-K)^{-1}\boldsymbol{W}$.

The absolute values of the standardized coefficients can be used to rank the variables in order of their contribution to separating the groups. Tables 3 and 4 show the standardized discriminant coordinate coefficients of the first and second discriminant coordinate, respectively, for the technological readiness data.

Table 3: Technological readiness. The standardized coefficients of the first discriminant coordinate $\hat{U}_1$.

	2009	2010	2011	2012	2013	2014	2015
X_1	0.0864	-0.0537	-0.0963	-0.1803	0.1950	-0.1269	0.0924
X_2	0.0133	0.1138	-0.2164	0.2248	-0.1485	0.0876	-0.0004
X_3	0.0412	-0.1393	-0.0500	0.1100	0.0013	-0.0387	0.0841
X_4	-0.0055	-0.0324	0.0696	0.0228	-0.0400	-0.0082	0.0397

Table 4: Technological readiness. The standardized coefficients of the second discriminant coordinate $\hat{U}_2$.

	2009	2010	2011	2012	2013	2014	2015
X_1	-0.0830	0.1067	-0.0487	-0.0289	-0.1082	0.3113	-0.0686
X_2	-0.0618	0.0751	-0.0677	0.1204	-0.0575	0.0096	-0.0629
X_3	-0.0829	0.1346	-0.1693	0.1705	-0.0329	-0.1491	0.1036
X_4	-0.0358	0.0071	0.0471	-0.0439	-0.0060	-0.0108	0.0245

The higher education and training pillar consists of 6 variables ($p = 6$). In this case $n \leq pT + q$, and we made the calculation with the additional condition $\boldsymbol{\Omega}_{11} = \boldsymbol{U} \otimes \boldsymbol{V}$. A projection of the 38 European countries on the plane $(\hat{U}_1, \hat{U}_2)$ is presented in Figure 2.

Figure 2 shows three clusters. The countries of Western Europe and Northern Europe in fact form a single group. The outliers countries in this group are Lithuania (LT) and Latvia (LV).

Tables 7 and 8 show the standardized discriminant coordinate coefficients of the first and second discriminant coordinate, respectively, for the higher education and training data.

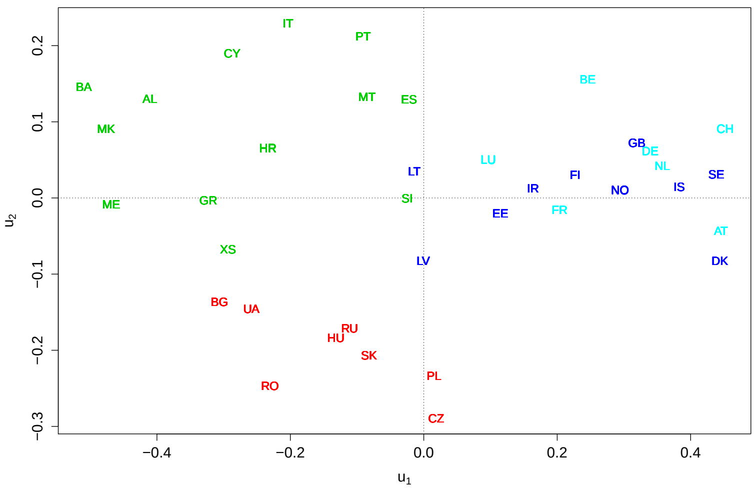


Figure 2: Higher education and training. Projection of the 38 European countries on the plane $(\hat{U}_1, \hat{U}_2)$. Regions used for statistical processing purposes by the United Nations Statistics Division: ■ – Northern Europe, ■ – Western Europe, ■ – Eastern Europe, ■ – Southern Europe

Table 5: Higher education and training. The standardized coefficients of the first discriminant coordinate $\hat{U}_1$.

	2009	2010	2011	2012	2013	2014	2015
X_1	0.1232	-0.2098	0.1197	-0.1592	0.2263	-0.1633	0.0831
X_2	-0.0356	-0.0918	0.1149	0.1016	-0.0958	0.0126	-0.0290
X_3	-0.0261	0.1900	-0.1579	0.0123	-0.0119	0.0524	-0.0450
X_4	0.1023	-0.0006	0.0861	-0.0974	-0.0743	-0.0163	0.0537
X_5	0.0128	-0.0820	0.0797	-0.0054	0.0534	-0.1114	0.1008
X_6	0.0338	-0.0280	-0.0398	0.0350	0.0367	0.0169	-0.0320

Table 6: Higher education and training. The standardized coefficients of the second discriminant coordinate $\hat{U}_2$.

	2009	2010	2011	2012	2013	2014	2015
X_1	-0.0725	0.2119	-0.1886	0.1291	-0.1413	0.1061	-0.0254
X_2	0.0638	-0.2120	0.1114	-0.1125	0.1617	-0.0979	0.0661
X_3	-0.0615	0.0122	0.0464	0.0027	0.0159	0.0551	-0.0473
X_4	-0.0810	0.1384	-0.0576	0.0334	-0.0219	0.0517	-0.0722
X_5	0.1000	-0.1485	0.0805	-0.0799	0.0571	-0.0048	-0.0037
X_6	0.0296	-0.0453	0.0292	0.0333	-0.1020	0.0755	-0.0132

During the numerical calculation process we used R software (R Core Team (2015)).

REFERENCES

BANERJEE, S., ROY, A., (2014). Linear Algebra and Matrix Analysis for Statistics. CRC Press, Boca Ratop.

FISHER, R. A., (1936). The use of multiple measurements in taxonomic problem. Annals of Eugenics 7, pp. 179–188.

FUJIKOSHI, Y., ULYANOV, V. V., SHIMIZU, R., (2010). Multivariate statistics. High-Dimensional and Large-Sample Approximations. Hoboken, New Jersey.

GIRI, C. G., (1996). Multivariate Statistical Analysis. Marcel Dekker, Inc., New York.

GÓRECKI, T., KRZYŚKO, M., WASZAK, Ł., WOŁYŃSKI, W., (2018). Selected statistical methods of data analysis for multivariate functional data, Statistical Papers 59, pp. 153–182.

KRZANOWSKI, W. J., (2000). Principles of Multivariate Analysis, Revised Edition. Oxford University Press, Oxford.

KRZYŚKO, M., (1979). Discriminant variables. Biometrical Journal 21 (3), pp. 227–241.

RAO, C. R., (1948). The utilization of multiple measurements in problem of biological classification. Journal of the Royal Statistical Society Series B, 10 (2), pp. 159–203.

R CORE TEAM, (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

RENCHER, A. C., (1998). Multivariate Statistical Inference and Applications. Wiley, New York.

SEBER, G. A. F., (1984). Multivariate Observations. John Wiley and Sons, New York.

SRIVASTAVA, M. S., (2002). Methods of Multivariate Statistics. Wiley, New York.

SRIVASTAVA, M. S., VON ROSEN, T., VON ROSEN, D., (2008). Models with a Kronecker product covariance structure: estimation and testing. Mathematical Methods of Statistics 17 (4), pp. 357–370.

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 507–526, DOI 10.21307/stattrans-2018-028

A MATHEMATICAL PROGRAMMING APPROACH FOR OBTAINING OPTIMUM STRATA BOUNDARIES USING TWO AUXILIARY VARIABLES UNDER PROPORTIONAL ALLOCATION

Faizan Danish[1]

ABSTRACT

Optimum stratification is the method of choosing the best boundaries that make the strata internally homogenous. Many authors have attempted to determine the optimum strata boundaries (OSB) when a study variable is itself a stratification variable. However, in many practical situations fetching information regarding the study variable is either difficult or sometimes not available. In such situations we find help in the variable (s) closely related to the study variable. Using auxiliary information many authors have formulated the problem as a MPP by redefining the problem as the problem of optimum strata width, and developed a solution procedure using dynamic programming technique. By using many distributions they worked out the optimum strata boundary points for the population under different allocation. In this paper, under proportional allocation OSBs are determined for the study variable using two auxiliary variables as the basis of stratification with uniform, right-triangular, exponential and lognormal frequency distribution by formulating the problems which are executed by using dynamic programming. Empirical studies are presented to illustrate the computation details of the solution procedure and its comparison with the existing literature.

Key words: optimum stratification, multistage decision problem, mathematical programming problem.

Mathematical Classification: 62D05

1. Introduction

Stratified random sampling is the most commonly used sampling technique for estimating population parameters with greater precision in sample surveys. In order to use the stratified random sampling the sample needs to choose the best boundary points such that the strata internally homogenous and the variance of the estimator within the strata be as small as possible. However, when the single characteristic is under study and its frequency distribution is known, one could use this information effectively to achieve the best boundary strata boundaries.

[1] Division of Statistics and Computer Science, Faculty of Basic Sciences, SKUAST-Jammu, Main campus, Chatha-180009 (J&K), India. E-mail: danishstat@gmail.com.

If in many situations the frequency distribution of the study variable is unknown, it may be approximated from the past experience or using some prior knowledge obtained in a recent study. This problem was pioneered by Dalenius (1950) and he obtained a set of minimal equations that could be solved for obtaining the optimum stratification points. However, the equations so obtained could not be solved provided the number of strata is small. Since then, several steps have been made for obtaining the stratification points such as Dalenius and Gurney (1951), Mahalanobis (1952), Aoyama (1954), Dalenius and Hodges (1959), Singh and Sukhatme (1969, 1973), Singh (1977), etc. Most of the authors suggested different approaches and obtained the calculus equations in terms of stratum mean and stratum variance for determining the strata boundaries.

Buhler and Deutler (1975) formulated the problem of optimum strata boundaries (OSB) as an optimization problem and developed a computational technique to solve the problem using dynamic programming. Khan *et al*. (2002, 2008) applied their procedure to determine OSB to the population various distributions. Danish et al. (2017a) made an attempt to present all the developed methods introduced for construction of stratification points using mathematical programming technique. Also, Danish *et al*. (2017b) proposed a method for determining OSB for single study variable with one auxiliary variable when the cost of every unit varies in the whole strata.

In this study, a procedure has been produced for constructing stratification points under proportional allocation for two auxiliary variables with uniform, exponential, right triangular and lognormal distributions.

2. Formulation of problem

Let us assume we have a population consisting of ‘N’ units stratified into L×M strata on the basis of two auxiliary variables ‘X’ and ‘Z’ when the estimation of the mean of the study variable ‘Y’ is of interest. We divide the whole population into the L×M (say) number of strata, such that each stratum is homogenous within itself and heterogeneous between strata with respect to the character under study such that the number of units in the (h, k)th stratum is N_{hk}, so that $\sum_{h=1}^{L}\sum_{k=1}^{M} N_{hk} = N$.

A sample of size n_{hk} (h=1,2,...,L; k=1,2,...,M) is to be drawn from each such that $\sum_{h}\sum_{k} n_{hk} = n$. The population unit in the (h , k)th stratum can be expressed as $Y = \sum_{h}\sum_{k}\sum_{i} y_{hki}$. We know, under stratified random sampling, the unbiased estimator of the population mean $\overline{Y}_N$ is

$$\overline{y}_{st} = \sum_{h}\sum_{k} W_{hk}\,\overline{y}_{hk}$$

where $W_{hk} = \frac{N_{hk}}{N}$ denotes the weight of the (h, k)th stratum and $\bar{y}_{hk} = \frac{1}{n_{hk}} \sum_i y_{hki}$

However, for an unbiased estimator $\bar{y}_{st}$ we have

$$V\left(\bar{y}_{st}\right) = \sum_h \sum_k \left(\frac{1}{n} - \frac{1}{N}\right) W_{hk}^2 \sigma_{hky}^2$$

where σ_{hky}^2 is the variance for the $(h,k)^{th}$ stratum (h = 1, 2, ..., L; k = 1,2,..., M). If finite population is ignored (fpc), we have

$$V\left(\bar{y}_{st}\right) = \sum_h \sum_k \frac{W_{hk}^2 \sigma_{hky}^2}{n}$$

Since ‘n’ is constant, thus it is sufficient to minimize

$$V\left(\bar{y}_{st}\right) = \sum_h \sum_k W_{hk}^2 \sigma_{hky}^2 \tag{2.1}$$

Let us assume the regression model of the study variable on auxiliary variables is of the form as:

$$Y = \lambda\left(x, z\right) + \varepsilon \tag{2.2}$$

where $\lambda\left(x, z\right)$ is a linear or non-linear function of ‘X’ and ‘Z’ and ‘ε’ denotes the error term such that its conditional expectation is zero and variance is finite and equal to $\phi\left(x, z\right)$ for all x and z.

For (h,k)th stratum the mean ‘μ_{hky}’ and the stratum variance ‘σ_{hky}^2’ can be written as

$$\mu_{hky} = \mu_{hk\lambda}$$

and

$$\sigma_{hky}^2 = \sigma_{hk\lambda}^2 + \mu_{hk\phi} \tag{2.3}$$

where $\mu_{hk\phi}$ are the expected values of $\phi\left(x, z\right)$ and $\mu_{hk\lambda}$ & $\sigma_{hk\lambda}^2$ denote the mean variance of $\lambda(x, z)$ in the (h, k)th stratum.

If 'λ' and 'ε'are uncorrelated, then in the model (2.2) then 'σ^2_{hky}' can be expressed as

$$\sigma^2_{hky} = \sigma^2_{hk\lambda} + \sigma^2_{hk\varepsilon}$$

where $\sigma^2_{hk\varepsilon}$ is the variance of error term in (h, k)th stratum.

Let the joint density function of (Y, X, Z) in the super population be f(y, x, z) and let f(x,z) be the joint function of X and Z, and f(x) & f(z) be the frequency function of the auxiliary variables X and Z, respectively, defined in the interval [a, b] and [c, d].

For determining the strata boundaries is to cut up the ranges $d_x = b - a$ and $t_z = d - c$, at (L-1) and (M-1) intermediate points as $a = x_0 \le x_1 \le \ldots \le x_{L-1} \le x_L = b$ and $c = z_0 \le z_1 \le \ldots \le z_{M-1} \le z_M = d$, respectively, such that the equation (2.1) is minimum.

Thus, while using (2.3), we have

$$\sum_h \sum_k W^2_{hk} \left(\sigma^2_{hk\lambda} + \mu_{hk\phi} \right) \tag{2.4}$$

$W_{hk}, \sigma^2_{hk\lambda}$ and $\mu_{hk\phi}$ can be obtained as a function of boundary points $\left(x_{h-1}, x_h, z_{k-1}, z_k\right)$ if $f\left(x,z\right), \lambda\left(x,z\right)$ and $\phi\left(x,z\right)$ are known and also integrable. Then, by using the following expression

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f\left(x,z\right) \partial x \partial z \tag{2.5}$$

$$\sigma^2_{hk\lambda} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \lambda^2\left(x,z\right) f\left(x,z\right) \partial x \partial z - \mu^2_{hk\lambda} \tag{2.6}$$

and

$$\mu_{hk\phi} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \phi\left(x,z\right) f\left(x,z\right) \partial x \partial z$$

where $\mu_{hk\lambda} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \lambda\left(x,z\right) f\left(x,z\right) \partial x \partial z$ and $\left(x_h, x_{h-1}\right)$ & $\left(z_k, z_{k-1}\right)$

Thus, the objective function (2.4) could be expressed as the function of boundary points $\left(x_{h-1}, x_h, z_{k-1}, z_k\right)$ only.

Let

$$\phi_{hk}\left(x_h, x_{h-1}, z_k, z_{k-1}\right) = W^2_{hk}\left(\sigma^2_{hk\lambda} + \mu_{hk\phi}\right) \tag{2.7}$$

and the ranges as:

$$d_x = b - a = x_L - x_0 \tag{2.8}$$

$$t_z = d - c = z_M - z_0 \tag{2.9}$$

Then, in the bivariate stratification the problem of determining the strata boundaries (x_h, z_k) is to break up the ranges of (2.8) and (2.9) at intermediate points. Then, the reasonable criterion for determining optimum strata boundaries (OSB) (x_h, z_k) is to minimize

Minimize $\sum_h \sum_k \phi_{hk}(x_h, x_{h-1}, z_k, z_{k-1})$

Subject to

$$a = x_0 \le x_1 \le \ldots \le x_{L-1} \le x_L = b$$
$$c = z_0 \le z_1 \le \ldots \le z_{M-1} \le z_M = d \tag{2.10}$$

and

$$\sum_h \sum_k n_{hk} = n$$

Let $V_h = x_h - x_{h-1}$ and $U_k = z_k - z_{k-1}$ denote the total length or width of the (h, k)th stratum for rectangular stratification. Then, using (2.8) and (2.9), the ranges can be expressed as

$$\sum_h V_h = d_x \tag{2.11}$$

$$\sum_k U_k = t_z \tag{2.12}$$

The objective function in (2.3) suggests that, for determining two way stratification, a two-dimensional dynamic programming approach should be used, employing the general concept of dynamic programming with the state and decision variables by the pairs (h, k). Then, the problem of two-way optimum stratification can be expressed as to

$$\text{Minimize} \quad \sum_h \sum_k \phi_{hk}(x_h, x_{h-1}, z_k, z_{k-1})$$

$$\begin{aligned}
&\text{Subject to} \\
&(x_h, z_k) = (x_{h-1} + V_h, z_{k-1} + U_k) \\
&(x_h, z_k) \in [a, d] \times [c, d] \\
&(V_h, U_k) \in B_h(x_{h-1}) \times B_k(z_{k-1}) \\
&\quad = [0, b - x_{h-1}] \times [0, d - z_{k-1}] \\
&(x_0, z_0) = [a, c] \\
&h = 1, 2, \ldots, L \quad \text{and} \quad k = 1, 2, \ldots, M
\end{aligned} \tag{2.13}$$

We propose a simple approach which permits a solution to the problem (2.13) using the unidimensional dynamic programming iteratively. Before the first iteration, some trail values, say x_0 and z_0, such that $a = x_0 \le x_1 \le ... \le x_{L-1} \le x_L = b$ and $c = z_0 \le z_1 \le ... \le z_{M-1} \le z_M = d$ are chosen for the initial points of the stratification. Then, for the i^{th} iteration (i=1, 2, ...) the points of stratification z^{i-1} are first considered as fixed. Note that the points of stratification x^{i-1}could also be chosen instead of z^{i-1}. Fixing the values of z^{i-1} has in fact the effect of reducing the problem exactly to the one of two-way optimum stratification with one categorical stratification variable. This can be seen by comparing the formulation (2.13) to the one which is defined on univariate auxiliary variable used as stratification variable with the values of the points of stratification Z taken as constant in (2.13).

Let $\phi^*_{x_h}\left(x_{h-1}, z^{i-1}\right)$ be the optimal value for the objective function (2.10) for the strata (h, k) to (L, k) for all k = 1, 2, ..., M given that the lower bound for the strata (h, k) for k = 1, 2, ..., M is x_{h-1}.The functional equation of Bellman with respect to the first part of the i^{th} iteration is then given by

$$\phi^*_{x_h}\left(x_{h-1}, z^{i-1}\right)$$

$$= \underset{V_h \in B_h\left(x_{h-1}\right)}{Minimize} \left\{ \sum_{k=1}^{M} \phi\left(x_{h-1}, x_h, z^{i-1}_{k-1}, z^{i-1}_k\right) + \phi^*_{x_{h+1}}\left(x_h, z^{i-1}\right) \middle| x_h = x_{h-1} + V_h \right\}$$

where $B_h\left(x_{h-1}\right)$ is defined in (5.1.17).

Restating the problem of determining OSB as the problem of determining optimum points $\left(V_h, U_k\right)$, adding equation (2.11) and (2.12) as a constraint, the problem (2.10) can be treated as an equation problem of determining Optimum Strata Width (OSW),$V_1, V_2, ..., V_L \quad and \quad U_1, U_2, ..., U_M$, and expressed as the following Mathematical Programming Problem (MPP):

$$Minimize \sum_h \sum_k \phi_{hk}\left(x_h, x_{h-1}, z_k, z_{k-1}\right)$$

Subject to (2.14)

$$\sum_h V_h = d_x$$

$$\sum_k U_k = t_z \text{ , h = 1, 2, ..., L and k = 1, 2, ..., M}$$

and

$$V_h \ge 0 \quad and \quad U_k \ge 0$$

Therefore, the first term $\phi_{11}\left(x_1, x_0, z_1, z_0\right)$ in the objective function (2.14) is the function of $\left(V_1, U_1\right)$ alone as $\left(x_0, z_0\right)$ are initially known, once the $\left(V_1, U_1\right)$ is known. The second term $\phi_{22}\left(x_2, x_1, z_2, z_1\right)$ will be the function of $\left(V_2, U_2\right)$ alone, and so on. Due to the special nature of function, MPP (2.14) may be treated as the function of $\left(V_h, U_k\right)$ and can be expressed as

Minimize $\sum_h \sum_k \phi_{hk}\left(V_h, U_k\right)$

Subject to (2.15)

$$\sum_h V_h = d_x$$

$$\sum_k U_k = t_z \text{ , h=1, 2, ..., L and k=1,2,...,M}$$

and $V_h \geq 0 \quad and \quad U_k \geq 0$

3. Proportional Allocation

Proportional allocation was originally proposed by Bowley (1926), which is very common in practice because of its simplicity, when no other information other than N_{hk}, which denotes the total number of units in the $\left(h, k\right)^{th}$ stratum, is available, the allocation of a given sample size ‘n’ to different strata is done in proportion to their sizes, i.e. in the $\left(h, k\right)^{th}$ stratum

$$n_{hk} = \frac{n}{N} N_{hk}$$

This means that the sampling fraction is the same in all strata. It gives a self-weighing sample by which numerous estimates can be made with greater speed and a higher degree of precision.

Under proportional allocation the variance is given by

$$V\left(\bar{y}_{st}\right) = \frac{\left(1-f\right)}{n} \sum_h \sum_k W_{hk} \sigma_{hky}^2$$

where $f = \frac{n}{N}$ is sampling fraction. If the finite population correction is ignored, we get

$$V\left(\bar{y}_{st}\right) = \frac{1}{n} \sum_h \sum_k W_{hk} \sigma_{hky}^2$$

Minimizing this function is equivalent to minimizing

$$\sum_{h}\sum_{k} W_{hk}\sigma_{hky}^{2} \tag{3.1}$$

Using the same procedure as discussed in the case of general and equal allocation, we need to replace the equation the objective function by $\sum_{h}\sum_{k} W_{hk}\sigma_{hky}^{2}$, Thus, MPP that we have to minimize is

Minimize $\sum_{h}\sum_{k} W_{hk}\sigma_{hky}^{2}$

Subject to

$$\begin{aligned} &\sum_{h} V_h = d_x \\ &\sum_{k} U_k = t_z \end{aligned} \tag{3.2}$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{aligned} h &= 1,2,\dots,L \\ k &= 1,2,\dots,M \end{aligned}$$

4. The solution procedure

The problem (2.15) is a problem of multistage decision in which the objective function and the constraints are separable functions of $\left(V_h, U_k\right)$, which allows us to use a dynamic programming technique, and a dynamic programming model is generally a recursive equation. These recursive equation links to different stages of the problem.

Consider the following sub-problem of equation (2.15) for first $\left(L_1 \times M_1\right)$ strata, where $\left(L_1 \times M_1\right) \leq \left(L \times M\right)$, i.e. $L_1 < L, M_1 < M$

Minimize $\sum_{h=1}^{L_1}\sum_{k=1}^{M_1} \phi_{hk}\left(x_{h-1}, x_h, z_{k-1}, z_k\right)$

Subject to (4.1)

$$\sum_{h=1}^{L-1} V_h = d_{L_1}$$

$$\sum_{k=1}^{M-1} U_k = t_{M_1} \qquad , \text{h=1, 2, ..., } L_1 \text{ and k =1, 2, ..., } M_1$$

and $\quad V_h \geq 0 \quad and \quad U_k \geq 0$

where $\quad d_{L_1} < d_x, t_{M_1} < t_z$

Note: If $d_{L_1} = d_x$ and $t_{M_1} = t_z$ then $(L_1 \times M_1) = (L \times M)$

The transformation functions are given by

$$d_{L_1} = V_1 + V_2 + ... + V_{L_1}$$

$$d_{L_1 - 1} = V_1 + V_2 + ... + V_{L_1 - 1} = d_{L_1} - V_{L_1}$$

$$\vdots$$

$$d_1 = V_1 = d_2 - V_2$$

Similarly, we have

$$t_{M_1} = U_1 + U_2 + ... + U_{M_1}$$

$$t_{M_1 - 1} = U_1 + U_2 + ... + U_{M_1 - 1} = t_{M_1} - U_{M_1}$$

$$\vdots$$

$$t_1 = U_1 = t_2 - U_2$$

Let the minimum value of the objective function of the equation (4.1) be denoted as

$$\phi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = Min\left[\sum_{h=1}^{L_1-1}\sum_{k=1}^{M_1-1} \phi_{hk}\left(V_h, U_k\right) \middle| \sum_{h=1}^{L_1-1} V_h = d_{L_1-1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1-1}\right] = A_1$$

$$and \quad V_h \geq 0, U_k \geq 0; h = 1,2,3,...,L_1 \quad ; \quad k = 1,2,3,...,M_1$$

with the above definition of $\phi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right)$, MPP (2.15) is equivalent to finding $\phi_{L \times M}\left(d_X, t_Z\right)$ recursively by defining $\phi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right)$ for $L_1 = 1, 2, ..., L$ and $M_1 = 1, 2, ..., M$; $0 \le d_{L_1} \le V, 0 \le t_{M_1} \le U$.

$$\phi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = Min \left[A_1 + \left[\sum_{h=1}^{L_1-1} \sum_{k=1}^{M_1-1} \phi_{hk}\left(V_h, U_k\right) \middle| \sum_{h=1}^{L_1-1} V_h = d_{L_1} - V_{L_1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1} - U_{M_1} \right] \right] \tag{4.2}$$

$$and \quad V_h \ge 0, U_k \ge 0; h = 1, 2, 3, ..., L_1 \quad and \quad k = 1, 2, 3, ..., M_1$$

For fixed value of $\left(V_{L_1}, U_{M_1}\right)$, $0 \le d_{L_1} \le V$, $0 \le t_{M_1} \le U$.

$$\phi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = A_1 + Min \left[\sum_{h=1}^{L_1-1} \sum_{k=1}^{M_1-1} \phi_{hk}\left(V_h, U_k\right) \middle| \sum_{h=1}^{L_1-1} V_h = d_{L_1} - V_{L_1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1} - U_{M_1} \right]$$

and

$$V_h \ge 0 \quad , h = 1, 2, ..., L_1, U_k \ge 0 \quad , k = 1, 2, ..., M_1, 1 \le L_1 \le L \quad , 1 \le M_1 \le M$$

Using the same procedure to write the forward recursive equation of the dynamic programming technique and could obtain OSB.

Let the estimation variable and the stratification variables take the regression model defined in (2.2) be of the form as

$$Y = \alpha + \beta x + \gamma z + \varepsilon \tag{4.3}$$

then
$$\sigma_{hky}^2 = \beta^2 \sigma_{hkx}^2 + \gamma^2 \sigma_{hkz}^2$$

The weight and variance of the (h, k)th stratum having auxiliary variables as 'X' and 'Z'.

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f\left(x, z\right) \partial x \partial z \tag{4.4}$$

$$\sigma_{hkx}^2 = \frac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} x^2 f\left(x\right) \partial x \partial z - \mu_{hkx}^2 \tag{4.5}$$

$$\sigma_{hkz}^2 = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} z^2 f\left(z\right) \partial z \partial x - \mu_{hkz}^2 \tag{4.6}$$

where $\mu_{hkx} = \dfrac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} x f(x) \partial x \partial z$,

$$\mu_{hkz} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} z f(z) \partial z \partial x$$

Thus, under proportional allocation with the model of the form given in (4.3), MPP will take the form as

Minimize $\sum_h \sum_k W_{hk} \left(\beta^2 \sigma_{hkx}^2 + \gamma^2 \sigma_{hkz}^2 \right)$

Subject to

$$\begin{aligned} &\sum_h V_h = d_x \\ &\sum_k U_k = t_z \\ &\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{matrix} h = 1, 2, ..., L \\ k = 1, 2, ..., M \end{matrix} \end{aligned} \tag{4.7}$$

5. Empirical study

I:Let the variable X follow a distribution with pdf as

$$f(x) = \begin{cases} 2(2-x) & ; 1 \leq x \leq 2 \\ 0 & ; otherwise \end{cases} \tag{5.1}$$

and the other auxiliary variable Z follow truncated exponential distribution with pdf

$$f(z) = \begin{cases} e^{-z+1} & ; \quad 1 \leq z \leq 6 \\ 0 & ; otherwise \end{cases} \tag{5.2}$$

In order to obtain OSB under proportional allocation with the pdf's of the auxiliary variables defined in (5.1) and (5.2), we need to obtain the value of W_{hk} and σ_{hky}^2 , for which we have to substitute (5.1) and (5.2) in equations (4.4)-(4.6), and get

$$W_{hk} = V_h e^{-z_k + 1} \left(e^{U_k} - 1 \right) \left(4 - V_h - 2x_{h-1} \right) \tag{5.3}$$

$$\sigma_{hkx}^2=\frac{a_1U_ke^{-z_k+1}}{\left(a_1e^{-z_k+1}\right)^2}\left\{\frac{4}{3}\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-\frac{1}{2}\left[\begin{matrix}\left(V_h+2x_{h-1}\right)\\ \left[V_h\left(V_h+2x_{h-1}\right)+x_{h-1}\left(1+x_{h-1}\right)\right]\end{matrix}\right]\right\}-4U_k^2 \tag{5.4}$$

and $$\sigma_{hkz}^2=\frac{a_1a_2-e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1}{a_1^2} \tag{5.5}$$

where $a_1=\left(e^{U_k}-1\right)\left(4-V_h-2x_{h-1}\right)$

$$a_2=z_{k-1}^2e^{U_k}-U_k^2-z_{k-1}^2-U_kz_{k-1}+2\left[e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1\right]$$

Substituting the values obtained in (5.3)-(5.5) in equation (4.7), we have MPP as

Minimize

$$\sum_h\sum_k\left(Sqrt\left(a_1V_he^{-z_k+1}\right)\right)$$
$$\left\{\beta^2\frac{a_1U_ke^{-z_k+1}}{6\left(a_1e^{-z_k+1}\right)^2}\left\{\left[8\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-3\left[\left(V_h+2x_{h-1}\right)\left[\begin{matrix}V_h\left(V_h+2x_{h-1}\right)\\+x_{h-1}\left(1+x_{h-1}\right)\end{matrix}\right]\right]\right]-4U_k^2\right\}\right.$$
$$\left.+\gamma^2\frac{a_1a_2-e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1}{a_1^2}\right\}$$

Subject to

$$\sum_h V_h=d_x$$
$$\sum_k U_k=t_z \tag{5.6}$$

$$\forall V_h\geq 0, U_k\geq 0 \quad , \quad \begin{matrix}h=1,2,...,L\\k=1,2,...,M\end{matrix}$$

By using the given pdf's a simulation has been done in R-software and the values of β = 0.576 and γ = 0.257 and have been obtained. Thus, using the values of β and γ, d_x = 1 and t_z = 5 as given above, the defined interval for X and Z respectively for total 6 (2×3) strata. Thus (5.6) can be written as

Minimize

$$\sum_h \sum_k \left[Sqrt\left(a_1 V_h e^{-z_k+1} \right) \right]$$

$$\left\{ (0.055) \frac{a_1 U_k e^{-z_k+1}}{\left(a_1 e^{-z_k+1}\right)^2} \left\{ \left[8\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - 3\left[\left(V_h + 2x_{h-1}\right) \left[\begin{array}{l} V_h\left(V_h + 2x_{h-1}\right) \\ + x_{h-1}\left(1 + x_{h-1}\right) \end{array} \right] \right] \right] - 4U_k^2 \right\} \right.$$

$$\left. + (0.666) \frac{a_1 a_2 - e^{U_k}\left(1 + z_{k-1}\right) - U_k - z_{k-1} - 1}{a_1^2} \right\}$$

Subject to

$$\sum_h V_h = 1$$

$$\sum_k U_k = 5 \qquad (5.7)$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1,2 \\ k = 1,2,3 \end{array}$$

Executing a computer programme for MPP (5.5.10) using LINGO software, we get OSB as given in tables below:

Table 5.1. OSB when the auxiliary variables X and Z are independent with right triangular and exponential distribution respectively

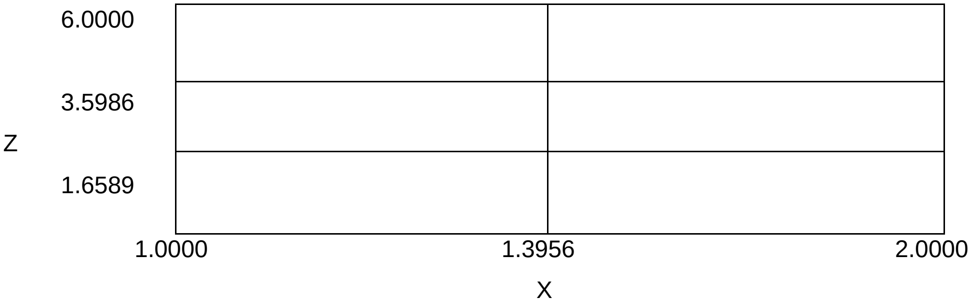


Table 5.2. OSB and Variance when the auxiliary variables X and Z are independent with right triangular and exponential distribution respectively

OSB (x_h, z_k)	Variance (Proposed method)	Variance (Thomson 1973)	% R.E.
(1.3956,1.6568) (2.0000,1.6568) (1.3956,3.5986) (2.0000,3.5986) (1.3956,6.0000) (2.0000,6.0000)	0.000864	0.00412	476.85

Thus, while making 2 strata along x-axis and 3 along z-axis when the auxiliary variables X and Z are having Right triangular and Exponential distribution respectively independently. The results obtained in Table 5.1 and 5.2 reveal that the variance obtained by the proposed method is much less than Thomson (1973), for which the percentage relative efficiency comes out to be 476.85. Thereby, it is revealed that the use of two auxiliary variables is better than using one auxiliary variable.

II: The log normal distribution is a positively skewed distribution. Surveyors may use the log normal distribution for a positive valued study variable, which might increase without limit, such as the value of securities in financial problem or the values of properties in real estate or the failure rate of electronic parts in the engineering problems.

Let us assume that one of the auxiliary variable, say X, follows log-normal distribution with pdf as

$$f(x)=\begin{cases}\dfrac{1}{\sigma x\sqrt{2\pi}}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} & ;x>0,\sigma>0\\ 0 & ;otherwise\end{cases} \tag{5.8}$$

and the other auxiliary variable Z with pdf as:

$$f(z)=\begin{cases}\dfrac{1}{b-a} & ,a\le z\le b\\ o & ,otherwise\end{cases} \tag{5.9}$$

Then, in order to estimate OSB we need to find the value of W_{hk} and σ^2_{hky}. Substituting the pdf's (5.8) and (5.9) in equations (4.4)-(4.6), we shall get

$$W_{hk} = \frac{U_k}{2(b-a)} E_1 \tag{5.10}$$

$$\sigma^2_{hkx} = \frac{(b-a)\left[e^{2\left(\sigma^2+\mu\right)}(E_2)(E_1)\right] - U_k^2 (b-a)^2 \left[e^{\frac{1}{2}\left(\sigma^2-2\mu\right)}(E_3)\right]^2}{E_1^2} \tag{5.11}$$

and $$\sigma^2_{hkz} = \frac{2E_1\left(U_k^2 + 3z_{k-1}^2 + 3U_k z_{k-1}\right) - 3V_h^2 \left(U_k + 2z_{k-1}\right)^2}{3E_1^2} \tag{5.12}$$

where $$E_1 = erf\left(\frac{\log(V_h + x_{h-1}) - \mu}{\sqrt{2\sigma^2}}\right) - erf\left(\frac{\log(x_{h-1}) - \mu}{\sqrt{2\sigma^2}}\right)$$

$$E_2 = erf\left(\frac{\log(V_h + x_{h-1}) - \mu - 2\sigma^2}{\sqrt{2\sigma^2}}\right) - erf\left(\frac{\log(x_{h-1}) - \mu - 2\sigma^2}{\sqrt{2\sigma^2}}\right)$$

and $$E_3 = erf\left(\frac{\log(V_h + x_{h-1}) - \mu - \sigma^2}{\sqrt{2\sigma^2}}\right) - erf\left(\frac{\log(x_{h-1}) - \mu - \sigma^2}{\sqrt{2\sigma^2}}\right)$$

It is to be noted here that the function 'erf', which repeats many times in the above result, is an error function, which is used to counter the integration with log-normal density function. It is defined as

$$erf(\omega) = \frac{2}{\sqrt{\pi}} \int_0^{\omega} e^{-j^2} \partial j$$

and some of its properties that need to be noted are

$$erf\left(-\omega\right)=-erf\left(\omega\right)$$
$$erf\left(0\right)=0$$
$$erf\left(\infty\right)=1$$
$$erf\left(-\infty\right)=1$$

Substituting values (5.10) to (5.12) in (4.7), we have MPP as

Minimize

$$\sum_h\sum_k\left[Sqrt\left(\frac{U_k}{2(b-a)}E_1\right)\right]\left(\beta^2\frac{(b-a)\left[e^{2\left(\sigma^2+\mu\right)}(E_2)(E_1)\right]-U_k^2(b-a)^2\left[e^{\frac{1}{2}\left(\sigma^2-2\mu\right)}(E_3)\right]^2}{E_1^2}+\gamma^2\frac{2E_1\left(U_k^2+3z_{k-1}^2+3U_kz_{k-1}\right)-3V_h^2\left(U_k+2z_{k-1}\right)^2}{3E_1^2}\right)$$

Subject to

$$\sum_h V_h=d_x$$
$$\sum_k U_k=t_z \qquad (5.13)$$

$$\forall V_h\geq 0, U_k\geq 0 \quad , \quad \begin{matrix} h=1,2,...,L \\ k=1,2,...,M \end{matrix}$$

In this case let us assume that the log-normal distribution is to be standardized, i.e. μ=0,σ=1 $z\in[0,1]i,e.z_0=0,z_M=1$ and the other variable $x\in[0,10],i,e.x_0=0,x_L=10$. Further, let us assume that the total strata to be made are 3×2(L×M)=6 and by simulation in R-software the value of β = 0.82 and γ = 0.437. Then, to obtain OSB we need to solve MPP

Minimize

$$\sum_{h=1}^{3}\sum_{k=1}^{2}\left[Sqrt\left(\frac{U_k}{2}E_1\right)\right]$$

$$\left((0.0722)\frac{\left[(7.389)\left(E_2^{'}\right)\left(E_1^{'}\right)\right]-U_k^2\left[(1.648)\left(E_3^{'}\right)\right]^2}{E_1^2} + (0.1909)\frac{2E_1\left(U_k^2+3z_{k-1}^2+3U_k z_{k-1}\right)-3V_h^2\left(U_k+2z_{k-1}\right)^2}{3E_1^2}\right)$$

Subject to

$$\sum_{h=1}^{3}V_h=10$$

$$\sum_{k=1}^{2}U_k=1 \qquad (5.14)$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{matrix} h=1,2,3 \\ k=1,2 \end{matrix}$$

where $E_1^{'}=erf\left(\frac{\log\left(V_h+x_{h-1}\right)}{1.141}\right)-erf\left(\frac{\log\left(x_{h-1}\right)}{1.141}\right)$

$$E_2^{'}=erf\left(\frac{\log\left(V_h+x_{h-1}\right)-2}{1.141}\right)-erf\left(\frac{\log\left(x_{h-1}\right)-2}{1.141}\right)$$

and $E_3^{'}=erf\left(\frac{\log\left(V_h+x_{h-1}\right)-1}{1.414}\right)-erf\left(\frac{\log\left(x_{h-1}\right)-1}{1.414}\right)$

By executing the computer programme of (5.14) MPP in LINGO, we get OSB value presented in the following tables.

Table 5.3. OSB when the auxiliary variables X and Z follow log-normal and uniform distribution respectively

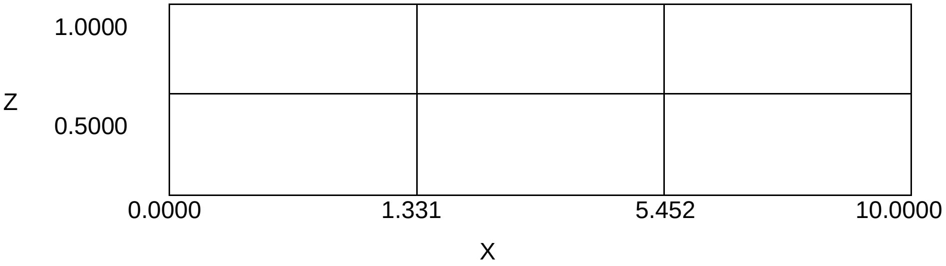


Table 5.4. OSB and Variance when the auxiliary variables X and Z follow log-normal and uniform distribution respectively

OSB (x_h, z_k)	Variance (Proposed method)	Variance (Khan *et al*. 2005)	% R.E.
(1.331,0.500) (5.452,0.500) (10.000,0.500) (1.331,1.000) (5.452,1.000) (10.000,1.000)	0.005916	0.014708	248.61

A perusal of Tab 5.4 indicates that the variance obtained by the proposed method is much less than not on Khan *et al*. (2005) and the percentage of relative efficiency comes out to be 248.61 of the proposed method over the other method in comparison. Thus, it may be concluded that using two auxiliary variables is better than using one auxiliary variable. In practice, the complete dataset of the study variable is unknown, which diminishes the uses of many stratification techniques. In such a situation, the proposed technique can be used as it requires only the values of parameters of the population, which can easily be available from the past studies

Conclusion

In this investigation, a scheme has been proposed to obtain the optimum strata boundaries (OSB) for two stratification variables highly related to the study variable. Numerical illustrations have been presented to explain the computational details of the application of the proposed method for two auxiliary variables. By using the frequency distribution the problem of constructing stratification points is formulated into the mathematical, programming problem, which results in a multistage decision problem, which is to be solved on a compromise distance.

In the empirical study I while comparing the proposed method with Thomson (1973), the percentage of relative efficiency comes out to be 476.85 when the auxiliary variables have right triangular and exponential distributions. Similarly, in study II the percentage of relative efficiency comes out to be 248.61 when comparing the proposed method with Khan *et al*. (2005). However, it is found that while obtaining the strata for two variables, when the frequency distributions are well-known, leads to the substantial gains in average relative efficiencies and have gains in precision of estimates. Thus, both the empirical studies suggest that the proposed method is more preferable than the existing methods.

REFERENCES

AOYAMA, H., (1954). A study of stratified random sampling. Annals of Institute of Statistical Mathematrica, Vol. 6, pp. 1–36.

BUHLER, W., DEUTLER, T., (1975). Optimal stratification and grouping by dynamic programming. Metrik, Vol. 22, pp. 161-175.

DALENIUS, T., (1950). The problem of optimum stratification. Skandinavisk Aktuarietidskrift, Vol. 33, pp. 203–213.

DALENIUS, T., HODGES, J. L., (1959). Minimum variance stratification. Journal of American Statistical Association, Vol. 54, pp. 88-101.

DALENIUS, T. AND GURNEY, M. (1951). The problem of optimum stratification II. Skandinavisk Aktuarietidskrift, Vol. 34, pp. 133–148.

DANISH, F., RIZVI, S. E. H., SHARMA, M. K., JEELANI, M. I., (2017a). Optimum Stratification Using Mathematical Programming Approach: A Review. Journal of Statistics Applications & Probability Letters, Vol. 4(3), pp. 123–129

DANISH, F., RIZVI, S. E. H., JEELANI, M. I., REASHI, J. A., (2017b). Obtaining Strata Boundaries under Proportional Allocation with Varying Cost of Every Unit. Pakistan Journal Of Statistics and Operations Research, Vol. 13(3), pp. 567–574

KHAN, E. A., KHAN, M. G. M., AHSAN, M. J., (2002). Optimum stratification: a mathematical programming approach. Calcutta Statistical Association bulletin, Vol. 52, pp. 323–333.

KHAN, M. G. M., NAND, N., AHMAD, N., (2008). Determining the optimum strata boundary points using dynamic programming. Survey Methodology, Vol. 34 (2), pp. 205–214.

KHAN, M. G. M., NAJMUSSEHAR, AHSAN, M. J., (2005). Optimum stratification for exponential study variable under Neyman allocation. Journal of the Indian Society of Agricultural Statistics, Vol. 59 (2), pp. 146–150.

MAHALANOBIS, P. C., (1952). Some aspect of design of sample surveys. Sankhya, Vol. 12, pp. 1–17.

SINGH, R., SUKHATME, B. V., (1969). Optimum stratification. Annals of Institute of Statistical Mathematrica, Vol. 21, pp. 515–528.

SINGH, R., SUKHATME, B. V., (1973). Optimum stratification with ratio and regression methods of estimation. Annals of Institute of Statistical Mathematrica, Vol. 25, pp. 627–633.

SINGH, R., (1977). A note on equal allocation with ratio and regression methods of estimation. Australian Journal of Statistics, Vol. 19, pp. 96–104.

THOMSEN, I., (1976). A comparison of approximately optimal stratification given proportional allocation with other methods of stratification and allocation. Metrika, Vol. 23, pp. 15–25.

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 527–549, DOI 10.21307/stattrans-2018-029

COMPARISON OF DIABETIC NEPHROPATHY ONSET TIME OF TWO GROUPS WITH LEFT TRUNCATED AND RIGHT CENSORED DATA

Alka Sabharwal[1], **Gurprit Grover**[2]

ABSTRACT

The present paper is concerned with the comparison of the nephropathy onset time of type-2 diabetic patients, grouped on the basis of gender and age at the time of diabetes diagnosis. Diabetic Nephropathy (DN) onset time is assumed to follow Weibull distribution with fixed left truncation. The likelihood ratio test is applied on uncensored cases and Thoman and Bain two sample tests is applied with generated left truncated Weibull distributions. To avoid the model validity issues for left truncated and right censored data (LTRC), the nonparametric approach, suggested by Kaplan and Meier, is used to compare the survival function of two groups over different time periods. Another method based on median survival time of the pooled group is applied to compare the survival function of two groups with LTRC data. The major advantage of developing methods for comparing the nephropathy onset times of DM patients is that the expected DN onset time of new DM patients can be predicted depending on the patient group.

Key words: Kaplan-Meier survival function; survival time; Weibull distribution.

1. Introduction

Diabetes is considered to be the primary cause of nephropathy if subjects develop diabetes after the age of 35 years and if diabetes was present in the subjects for more than 5 years before the initiation of renal replacement therapy (Meredith et al. 2009). Type-2 diabetes is known as adult-onset diabetes as it is primarily seen in middle-aged adults over the age of 40. (Brenner et al. 2003). It has been predicted by Viswanathan (2004) that worldwide the prevalence of diabetes in adults would increase to 5.4% by the year 2025 from the prevalence rate 4.0% in 1995. Rodby (1997) study suggested that type-2 diabetic males are at greater risk of developing nephropathy. Also, it has been shown by Wagle (2010) that serum creatinine levels in males are significantly higher than in females.

[1] Kirori Mal College, Department of Statistics, University of Delhi, India. E-mail: alkasabh@gmail.com.
[2] Department of Statistics, University of Delhi, India. E-mail: gurpritgrover@yahoo.com.

Survival time comparison is one of the main goals of survival studies. Biomedical studies often compare the distributions of failure/survival time variables among two or more groups. Rossing et al. (1996) study compared three levels of albuminuria in insulin dependent diabetic patients. Joss et al. (2002) work concludes that survival time of type-2 diabetic patients, once diabetic nephropathy has developed, becomes even worse after starting dialysis. Bruce, Sheppard and others (2004) compared survival times of three categories: no diabetes, diabetes without peripheral vascular disease and renal failure, and diabetes with peripheral vascular disease and/or renal failure. Ashfaq et al. (2006) compared survival time of diabetic and non-diabetic groups to observe the effect of vein graft intervention. In all the above mentioned studies the authors estimated the survival function by Kaplan-Meier method and applied log rank test to compare the survivability of groups. Villar et al. (2007) suggested Cox proportional hazard model to study the effect of renal replacement therapy on the survival times of type-1 diabetic, type -2 diabetic and non-diabetic patients. Jianguo Sun and others (2008) compared survival times of two groups by applying generalized log-rank test. They discuss a class of generalized log-rank tests for incomplete survival data and establish their asymptotic properties and illustrated their study with diabetic patient data.

There are two broad approaches to compare the survival distribution among two groups: non-parametric and parametric. The Weibull family is commonly used in the statistical analysis of lifetime or response time data from reliability experiments and survival studies. To compare two Weibull distributions likelihood ratio test can be used for small samples. This test is based on the identification of the likelihood function. In parametric problems, the likelihood is usually a well-defined quantity. Thoman and Bain (1969) proposed a test to compare shape parameters in two Weibull distributions with the scale parameters unknown, along with a procedure which tests the equality of scale parameters.

The main objectives of this paper are to compare the DN onset time of (i) male and female groups, and (ii) the groups whose age at diabetes diagnosis is less than or equal to 45 years and more than 45 years. In this paper we consider data sets representing the survival time until the occurrence of an event of interest which is diabetic nephropathy. Two methods, i.e. parametric and nonparametric methods are used to compare the survival times of two groups, in the above two cases. Under parametric method, we have compared two left truncated Weibull distributions by applying likelihood ratio test and the two sample test proposed by Thoman and Bain (1969), on the generated left truncated Weibull distributions. Nonparametric methods are applied with left truncated and right censored data. The survival functions are first estimated by applying Kaplan-Meier (KM) method for the groups in both the cases. Then, weighted KM method with appropriate weights is used to test the equality of survival functions, by comparing the difference in the survival functions over time. We have also modified Brookmeyer and Crowley (1982b) method, which is based on the median survival time, to test the equality of median survival of two groups in both the cases. The remainder of the paper is organized as follows. In section 2 development of the models is discussed. Section 3 applies the models to a data set of type-2 diabetic patients (diagnosed of diabetes as per ADA standards) from the data base of Dr. Lal's Path Lab, Delhi, India. Although some work has been done on the estimation of survival times of diabetic patients, to the best of our knowledge there is no study that has systematically compared the onset DN times of two groups with type-2 diabetes under both the methods: parametric and nonparametric . Some concluding remarks are made in section 4.

2. Development of the model

Observations are made on m mutually independent type-2 diabetic patients. These m patients, on certain criterion, are divided into two groups of sizes m_1 and m_2 , so that $m = m_1 + m_2$. Survival time, i.e. DN onset time is not known for all m individuals. It is known for n_1 from m_1 and n_2 from m_2 patients. Thus, survival time is known for $n = n_1 + n_2$ patients and $t_1, t_2, ...t_n$ denote the DN onset time for the combined data. The data at time t (end of study) from two groups can be summarized in a 2 X 2 table as given in Table 1.

Table 1. Distribution of diabetic patients among different groups

Group	Uncensored	censored	Total
Group 1	n_1	$m_1 - n_1$	m_1
Group 2	n_2	$m_2 - n_2$	m_2
Total	n	$m - n$	m

2.1. Likelihood ratio test to compare two left truncated Weibull distributions

Let $t_1, t_2, ...t_{m_1}$ and $t_1, t_2, ...t_{m_2}$ be the times of patients from two groups of sizes m_1 and m_2 . The disease time from two groups is assumed to follow Weibull distribution characterized by two parameters, shape (γ) and scale (λ), which are unknown. All the patients included in this study are patients with diabetic history of more than 5 years. The probability density function and survival function for the i^{th} patient belonging to the j^{th} group ($j = 1, 2$) are given as follows:

$$f_j(t_i) = \frac{\lambda_j \gamma_j \ (\lambda_j t_i)^{\gamma_j - 1} \ \exp(-(\lambda_j t_i)^{\gamma_j})}{\exp(-(5\lambda_j)^{\gamma_j})} \qquad t_i \geq 5, \ i = 1, 2......., m_j \tag{1}$$

$$S_j(T_i) = \frac{\exp(-(\lambda_j T_i)^{\gamma_j})}{\exp(-(5\lambda_j)^{\gamma_j})} \qquad \lambda_j > 0, \gamma_j > 0 \tag{2}$$

$$L_j = \prod_{1}^{m_j} (f(t_{ij}))^{\delta_{ij}} (S(T_{ij}))^{1-\delta_{ij}}$$

where δ_{ij} is zero if the i^{th} patient of the j^{th} group is censored and δ_{ij} is unity when the i^{th} patient of the j^{th} group is uncensored, i.e. DN onset time is known. $\mathrm{L_j}(\lambda_\mathrm{j},\gamma_\mathrm{j})$, $j = 1,2$, denoting the log likelihood functions of the two groups, are given by

$$L_j(\lambda_\mathrm{j},\gamma_\mathrm{j}) = \sum_1^{m_j} \delta_{ij}((\log\gamma_\mathrm{j} + \gamma_\mathrm{j}\log\lambda_\mathrm{j} - (5\lambda_\mathrm{j})^{\gamma_\mathrm{j}}) + (\gamma_\mathrm{j}\ \text{-}1)\log \mathrm{t_{ij}} - (\lambda_\mathrm{j}\mathrm{t_{ij}})^{\gamma_\mathrm{j}})$$
$$-\sum_1^{m_j}(1-\delta_j)((5\lambda_\mathrm{j})^{\gamma_\mathrm{j}} + (\lambda_\mathrm{j}T_\mathrm{ij})^{\gamma_\mathrm{j}}) \qquad (3)$$

The maximum likelihood estimates of λ_j and γ_j are found by partially differentiating the above function with respect to λ_j and γ_j, and equating the derivatives to zero. The resulting equations are:

$$\lambda_j^{\gamma_j}(\sum_1^{m_j}\delta_{ij}t_{ij}^{\gamma_j} + \sum_1^{m_j}(1\text{-}\delta_{ij})T_{ij}^{\gamma_j} - m_j 5^{\gamma_j}) - \sum_1^{m_j}\delta_{ij} = 0 \qquad (4)$$

$$\lambda_j^{\gamma_j}(\sum_1^{m_j}\delta_{ij}t_{ij}^{\gamma_j}\log(\lambda_j t_{ij}) + \sum_1^{m_j}(1\text{-}\delta_{ij})\log(\lambda_j T_{ij})T_{ij}^{\gamma_j} - n5^{\gamma_j}\log(5\lambda_j)) - \sum_{i=1}^{m_j}\delta_{ij}(\frac{1}{\gamma_j} + \log(\lambda_j t_{ij})) = 0$$
$$(5)$$

The maximum likelihood estimates of λ_j and γ_j are obtained as $\hat{\lambda}_\mathrm{j}$ and $\hat{\gamma}_\mathrm{j}$ using error and trial method.

To compare the two left truncated Weibull distributions using likelihood ratio test (Lee, 2003), the null hypothesis is given by

$$H_0 : \lambda_1 = \lambda_2 = \lambda \quad \text{and} \quad \mathrm{H_0}: \gamma_1 = \gamma_2 = \gamma \qquad (6)$$

where λ and γ are unknown. To test the above null hypothesis, we compute the statistic

$$X_L = 2(\mathrm{L_1}(\hat{\lambda}_1,\hat{\gamma}_1) + \mathrm{L_2}(\hat{\lambda}_2,\hat{\gamma}_2) - L(\tilde{\lambda},\tilde{\gamma},\tilde{\lambda},\tilde{\gamma})) \qquad (7)$$

$L(\tilde{\lambda},\tilde{\gamma},\tilde{\lambda},\tilde{\gamma})$, the log likelihood value of the combined group with $\tilde{\lambda}$ and $\tilde{\gamma}$ as maximum likelihood estimators, is defined as:

$$L(\tilde{\lambda},\tilde{\gamma},\tilde{\lambda},\tilde{\gamma}) = \sum_1^{uncen}\delta_{ij}((\log\tilde{\gamma} + \tilde{\gamma}\log\tilde{\lambda} - (5\tilde{\lambda})^{\tilde{\gamma}}) + (\tilde{\gamma}\ \text{-}1)\log \mathrm{t_{ij}} - (\tilde{\lambda}\mathrm{t_{ij}})^{\tilde{\gamma}})$$
$$-\sum_1^{censored}(1-\delta_j)((5\tilde{\lambda}_\mathrm{j})^{\tilde{\gamma}_\mathrm{j}} + (\tilde{\lambda}_\mathrm{j}T_\mathrm{ij})^{\tilde{\gamma}_\mathrm{j}}) \qquad (8)$$

This likelihood ratio test statistic defined in equation (7), first propounded by Fisher (1922), gives (twice) the log likelihood of the ratio of one hypothesis vs. the other. It is used to compare two left truncated distributions with fixed truncation time. We reject H_0 if $X_L > \chi^2_{2,\alpha}$, or equivalently, $P(\chi^2_2 > X_L) < \alpha$.

2.2. Likelihood ratio test to compare two left truncated Weibull distributions for uncensored cases

Let $t_1, t_2, ...t_{n_1}$ and $t_1, t_2, ...t_{n_2}$ be the observed exact DN onset times of patients from two groups of sizes n_1 and n_2. The DN onset time from two groups is assumed to follow Weibull distribution characterized by two parameters, shape (γ) and scale (λ), which are unknown. The probability density function for the i^{th} patient belonging to the j^{th} group ($j = 1,2$) is given above in equation 1 under section 2.1.

$\mathrm{L_j}(\lambda_j, \gamma_j)$, $j = 1,2$, denoting the log likelihood functions of the observed survival time from two groups, are given by

$$\mathrm{L_j}(\lambda_j, \gamma_j) = n_j(\log\gamma_j + \gamma_j \log\lambda_j - (5\lambda_j)^{\gamma_j}) + (\gamma_j - 1)\sum_{i=1}^{n_j} \log t_i - \sum_{i=1}^{n_j} (\lambda_j t_i)^{\gamma_j} ;$$

$$j = 1,2 \tag{9}$$

The maximum likelihood estimates of λ_j and γ_j are obtained as $\hat{\lambda}_j$ $\hat{\gamma}_j$, respectively.

To compare the two left truncated Weibull distributions using likelihood ratio test

$$H_0 : \lambda_1 = \lambda_2 = \lambda \quad \text{and} \quad H_0 : \gamma_1 = \gamma_2 = \gamma$$

where λ and γ are unknown. To test the above null hypothesis, we compute the statistic

$$X_L = 2(\mathrm{L_1}(\hat{\lambda}_1, \hat{\gamma}_1) + \mathrm{L_2}(\hat{\lambda}_2, \hat{\gamma}_2) - L(\tilde{\lambda}, \tilde{\gamma}, \tilde{\lambda}, \tilde{\gamma}))$$

$L(\tilde{\lambda}, \tilde{\gamma}, \tilde{\lambda}, \tilde{\gamma})$, the log likelihood value of the combined group of uncensored cases, is defined as:

$$L(\tilde{\lambda}, \tilde{\gamma}, \tilde{\lambda}, \tilde{\gamma}) = (n_1 + n_2)(\log\tilde{\gamma} + \tilde{\gamma}\log\tilde{\lambda} - (5\tilde{\lambda})^{\tilde{\gamma}}) + (\tilde{\gamma} - 1)(\sum_{i=1}^{n_1} \log t_i + \sum_{i=1}^{n_2} \log t_i) - (\sum_{i=1}^{n_1} (\tilde{\lambda} t_i)^{\tilde{\gamma}} + \sum_{i=1}^{n_2} (\tilde{\lambda} t_i)^{\tilde{\gamma}}) \tag{10}$$

This test is sufficient if the data concludes that the two distributions are significantly different but if the data concludes otherwise, an additional two-sample test proposed by Thoman and Bain (1969) for uncensored samples for comparing two Weibull distributions has been applied for left truncated distributions.

2.2. Thoman and Bain two-sample test for comparing two left truncated Weibull distributions

This test assumes that independent samples of equal sizes are obtained from left truncated Weibull distributions as given in 2.1. To compare the two distributions the equality of shape parameter is tested. The null and alternative hypothesis is given as:

$$H_0 : \gamma_1 = \gamma_2 \quad \text{against} \quad H_1 : \gamma_1 > \gamma_2 \qquad (11)$$

To test the above null hypothesis, the following statistic is computed

$$R = \frac{\hat{\gamma}_1 / \gamma_1}{\hat{\gamma}_2 / \gamma_2} \quad \text{and under } H_0, \quad R = \frac{\hat{\gamma}_1}{\hat{\gamma}_2} \quad \text{(Thoman and Bain, 1969)}$$

We reject H_0 if $R > \ell_\alpha$ (using Thoman and Bain, 1969) and conclude that the two Weibull distributions are significantly different. However, if the hypothesis $H_0 : \gamma_1 = \gamma_2$ is not rejected we test the equality of scale parameters.

$$H_0 : \lambda_1 = \lambda_2 \quad \text{against} \quad H_1 : \lambda_1 < \lambda_2 \qquad (12)$$

To test the above null hypothesis we compute the statistic given as follows:

$$G = 0.5(\hat{\gamma}_1 + \hat{\gamma}_2)(\log \hat{\lambda}_1 - \log \hat{\lambda}_2) \quad \text{(Thoman and Bain, 1969)} \qquad (13)$$

H_0 is rejected if $G > z_\alpha$, (using Thoman and Bain, 1969) where z_α is $P(G < z_\alpha \mid H_0) = 1 - \alpha$ and $\hat{\gamma}_1$, $\hat{\gamma}_2$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are the maximum likelihood estimators of γ_1, γ_2, λ_1 and λ_2 respectively.

2.3 Weighted Kaplan-Meier method to compare survival distributions of two groups

We consider the classical two-sample censored data with fixed left truncation survival analysis problem, with survival continuous and censoring independent of survival in each group. To compare the survival distributions of two groups of sizes m_1 and m_2, as defined in section 2, let $t_1 < t_2 < \ldots\ldots < t_n$ denote the ordered survival time of two groups taken together. Let O_{ij}, C_{ij}, Y_{ij} be, respectively, the number of events, the number of censored observations and the number at risk at time t, in the j^{th} group, $j = 1,2$. Let $\hat{S}_j(t)$ be the Kaplan-Meier (KM) estimator (Klien and Moeschberger, 2003; Pepe and Fleming, 1989) of the event distribution using data in the j^{th} group and $\hat{H}_j(t)$ be the KM estimator of the time to censoring in the j^{th} group, that is $\hat{H}_j(t) = \prod_{t_i \le t} [1 - C_{ij} / Y_{ij}]$, and $\hat{S}_p(t)$ be the KM estimator based on the combined group.

To test the equality of the two survival distributions the hypothesis is defined as

$$H_0 : S_1(t) = S_2(t) \quad \text{against} \quad H_1 : S_1(t) > S_2(t) \tag{14}$$

Or $$H_0 : S_1(t) = S_2(t) \quad \text{against} \quad H_1 : S_1(t) \neq S_2(t)$$

The test statistic is given by

$$Z = W_{KM} / \hat{\sigma}_p$$

where

$$W_{KM} = \sqrt{\frac{m_1 m_2}{m}} \sum_{i=1}^{D-1} [t_{i+1} - t_i] w(t_i) [\hat{S}_1(t_i) - \hat{S}_2(t_i)] \tag{15}$$

KM suggested a weight function, w(t), in the study period t_D defined as

$$w(t) = \frac{m \hat{H}_1(t) \hat{H}_2(t)}{m_1 \hat{H}_1(t) + m_2 \hat{H}_2(t)} \quad , \quad 5 \leq t \leq t_D$$

The variance ($\hat{\sigma}_p$) is given by

$$\hat{\sigma}_p^2 = -\sum_{i=1}^{D-1} [\hat{S}_p(t_i) - \hat{S}_p(t_{i-1})] \frac{T^2{}_i}{\hat{S}_p(t_i) \hat{S}_p(t_{i-1})} \frac{m_1 \hat{H}_1(t_{i-1}) + m_2 \hat{H}_2(t_{i-1})}{m \hat{H}_1(t_{i-1}) \hat{H}_2(t_{i-1})} \tag{16}$$

where

$$T_i = \sum_{k=i}^{D-1} [t_{k+1} - t_k] w(t_k) \hat{S}_p(t_k)$$

The sum in (12) has only nonzero contributions as t_i is the onset time and for censored observations $\hat{S}_p(t_{i-1}) - \hat{S}_p(t_i)$ =0.

The test statistic Z has an approximate standard normal distribution. Suppose the alternative hypothesis, $S_1(t) > S_2(t)$, reject H_0 if Z comes out to be greater than Z_α ($Z >$ Z_α) where Z_α is given by $P(Z > Z_\alpha \mid H_0) = \alpha$. If the alternative hypothesis, $S_1(t) \neq S_2(t)$, reject H_0 if Z comes out to be greater than $Z_{\alpha/2}$.

We suggest a new weight function defined as:

$$w_{AG}(t) = \frac{(n/m) \hat{H}_1(t) \hat{H}_2(t)}{(n_1/m_1) \hat{H}_1(t) + (n_2/m_2) \hat{H}_2(t)} \; ; \; 5 \leq t \leq t_D \tag{17}$$

to test the equality of the two survival distributions. In this case the test statistic is computed as follows:

$$Z = W_{AG} / \hat{\sigma}_{AG}$$

where W_{AG} and $\hat{\sigma}_{AG}$ can be computed like W_{KM} and $\hat{\sigma}_p$ by replacing m, m_1 and m_2 by $n/m, n_1/m_1$ and n_2/m_2, respectively.

2.4. Modified Brookmeyer and Crowley method to compare survival distributions of two groups

Brookmeyer and Crowley (1982b) have suggested an alternate method (Klien and Moeschberger, 2003) to test the equality of survival times of two groups. This method is based on median survival time rather than comparing the difference in the survival functions over time. Let $t_1, t_2 ...., t_m$ denote the diabetic duration of the two groups taken together, where O_{ij}, Y_{ij}, $\hat{S}_j(t)$ ($j = 1, 2$) and $\hat{S}_p(t)$ are the same as defined in the procedure given in section 2.3.

To test the equality of the two median survival times the null hypothesis is given as follows:

$$H_0 : M_d(S_1(t)) = M_d(S_2(t)) \text{ against } H_1 : M_d(S_1(t)) \neq M_d(S_2(t)) \tag{18}$$

and the test statistic is defined as follows:

$$U = \frac{m(S_1(\hat{M}) - 0.5)}{\sigma^2}$$

Brookmeyer and Crowley derived a method where median survival time is based on the common survival function, $S_w(t) = \dfrac{m_1\hat{S}_1(t) + m_2\hat{S}_2(t)}{m}$.

We modified this survival function, which meets the situation when the number of events and their onset and censoring times in $\hat{S}_1(t)$ and $\hat{S}_2(t)$ are different. In our case median survival time is based on common pooled survival function. To compute $S_1(\hat{M})$, σ^2 and $\hat{M}$, the algorithm used (Klien and Moeschberger, 2003) is given as follows:

- Arrange the onset time as $t_1 < t_2 < ...., t_n$ in a pooled sample. If for some t_i, $\hat{S}_p(t_i) =$ 0.5 , then $\hat{M} = t_i$.
- If no event time gives a value of $\hat{S}_p$ equal to 0.5 then let M_L be the largest event time with $\hat{S}_p(M_L)$ > 0. 5 and let M_U be the smallest event with $\hat{S}_p(M_U)$ < 0.5. Then, the median lies in the interval (M_L , M_U) and is obtained by linear interpolation, i.e.

$$\hat{M} = M_L + \frac{(0.5 - \hat{S}_p(M_L))(M_U - M_L)}{(\hat{S}_p(M_U) - \hat{S}_p(M_L))} \tag{19}$$

- Obtain the death time in the j^{th} sample such that $T_{Lj} \le \hat{M} < T_{Uj}$. The estimated probability of survival beyond $\hat{M}$ in the j^{th} sample is obtained by linear interpolation as

$$\hat{S}_j(\hat{M}) = \hat{S}_j(T_{Lj}) + \frac{(\hat{S}_j(T_{Uj}) - \hat{S}_j(T_{Lj}))(\hat{M} - T_{Lj})}{(T_{Uj} - T_{Lj})}, \quad j = 1,2 \tag{20}$$

- For obtaining σ^2, let t_{ij} denote the distinct death time in the j^{th} sample, d_{ij} the number of deaths at time t_{ij} and Y_{ij} the number at risk at time t_{ij}. Then, σ^2 is given as follows:

$$\sigma^2 = \frac{m_2^{\,2}}{m}(V_1 + V_2) \tag{21}$$

$$V_j = [\hat{S}_j(T_{Uj})(\frac{\hat{M} - T_{Lj}}{T_{Uj} - T_{Lj}})]^2 \sum_{t_{ij} \le T_{Uj}} \frac{O_{ij}}{Y_{ij}(Y_{ij} - O_{ij})} +$$

$$+\{[\hat{S}_j(T_{Lj})(\frac{T_{Uj} - \hat{M}}{T_{Uj} - T_{Lj}})]^2 + 2(\frac{(\hat{M} - T_{Lj})(T_{Uj} - \hat{M})}{(T_{Uj} - T_{Lj})^2})\hat{S}_j(T_{Uj})\hat{S}_j(T_{Lj})]\} \sum_{t_{ij} \le T_{Lj}} \frac{O_{ij}}{Y_{ij}(Y_{ij} - O_{ij})}$$

The test statistic U follows chi-square distribution with one degree of freedom. Reject H_0 if U comes out to be greater than $Z_{\alpha/2}$, where Z_α is given by $P(U > Z_\alpha \mid H_0) = \alpha$.

3. Application

The methods discussed in section 2 are applied to the data obtained through a house to house survey of diabetic patients who were referred for pathological tests to Dr. Lal path lab, Delhi, India. A retrospective study is conducted on the collected data. Since our study is focused on diabetic nephropathy only, the patient's data indicating effect on eyes, heart, etc. are excluded. Thus, a total of 132 patients were selected who were diagnosed as diabetics as per ADA standards with minimum 5 years duration. Out of these 132 patients, 60 were uncensored with diabetic nephropathy and 72 were censored/non-diabetic nephropathy, all aged between 44.45 ± 4.79 years (mean ± SD). The patients in the diabetic nephropathy group were distributed according to gender and age at the time of diagnosis as displayed in Table 3. The demographic and risk variables recorded were: age at the time of diagnosis, duration of disease, fasting blood glucose (FBG), diastolic blood pressure (DBP), systolic blood pressure (SBP), low-density lipoprotein (LDL) and serum creatinine (SrCr) as given in Table 2.

Table 2. Mean± SD of demographic and risk variables of 132 patients of two groups; non-diabetic nephropathy(NDN) and diabetic nephropathy(DN)

Variable	NDN=Censored	DN=Uncensored
Group size	72	60
Age at diabetes diagnosis (years)	44.01±4.36	45.00± 5.28
Duration of diabetes (years)	10.28±5.70	14.10±5.05
FBG (mg/dl)	133.80±17.48	142.04±14.39
DBP (mmHg)	82.39±6.08	91.97±9.42
SBP (mmHg)	125.12±12.40	142.82±13.88
LDL (mg/dl)	91.80±18.75	107.44±14.27
SrCr (mg/dl)	1.00±0.151	1.67±0.2833

* NDN denotes non-diabetic nephropathy and DN denotes diabetic nephropathy

Table 3. Distribution of patients among different groups based on gender and age at diabetes diagnosis.

Group	Uncensored	Censored	Total group size
Male	28	31	59
Female	32	41	73
Diagnosis≤45years	33	48	81
Diagnosis>45years	27	24	51
Total	60	72	132

3.1. Comparison of patients on the basis of gender and age at the time of diagnosis of diabetes by applying likelihood test

The motivation of our model development is to compare the time of type-2 diabetic patients on the basis of gender and age at the time of diagnosis of diabetes. Out of 132 cases, there were 59 males and 73 females. Available diabetic data has been used to test the equality of male and female disease time, which is assumed to follow Weibull distribution with fixed left truncation time as 5 years. To find the likelihood ratio test statistic, X_L, and to test the hypothesis (5), we have found the maximum likelihood estimators of unknown parameters for male, female and combined groups and their respective log-likelihood values as given in Table 4. The observed value of the likelihood ratio test statistic, X_L, came out to be 34.9082. Since test statistic X_L is found to be greater than $\chi^2_{2,\alpha}$, p < 0.001, we reject H_0 and conclude that the two Weibull distributions are significantly different.

Secondly, we have classified the time on the basis of age at diabetes diagnosis. Out of 132 cases with diabetic nephropathy, the age at diabetes diagnosis of 81 patients was less than or equal to 45 years and of 51 patients was greater than 45 years. Available data has been used to test the equality of the two groups, which are based on age at diabetes diagnosis and are assumed to follow Weibull distribution with fixed left truncation time as 5 years. To test the hypothesis (5), the above procedure is repeated to compute the log-likelihood ratio test statistic and its value came out to be X_L = 16.2872. Again, since test statistic X_L is greater than $\chi^2_{2,\alpha}$, p <0.001, we reject H_0 and conclude that the DN onset times of two groups are significantly different or the two Weibull distributions are significantly different. The log-likelihood values and maximum likelihood estimators of the parameters are given in Table 4.

Table 4. Maximum Likelihood estimators (MLE) of λ and γ along with log likelihood value of two groups when the grouping variable is (i) gender (ii) age at diabetes diagnosis

(i) Group Size	**Male** m_1 **=59**	**Female** m_2 **=73**	**Total** m_1+m_2 **=132**
MLE of shape parameter	$\hat{\gamma}_1$ =3.4115	$\hat{\gamma}_2$ =3.61382	$\tilde{\gamma}$ =4.04138
MLE of scale parameter	$\hat{\lambda}_1$ =0.0661	$\hat{\lambda}_2$ =0.0718	$\tilde{\lambda}$ =0.080238
log likelihood value	$L_1(\hat{\lambda}_1,\hat{\gamma}_1)$ =-46.1095	$L_2(\hat{\lambda}_2,\hat{\gamma}_2)$ =-60.1714	$L(\tilde{\lambda},\tilde{\gamma},\tilde{\lambda},\tilde{\gamma})$ =-123.735
(ii) Group Size	**Diagnosis ≤45yrs** m_1 **=81**	**Diagnosis >45yrs** m_2 **=51**	**Total** m_1+m_2 **=132**
MLE of shape parameter	$\hat{\gamma}_1$ = 3.4967	$\hat{\gamma}_2$ =3.8199	$\tilde{\gamma}$ =4.04138
MLE of scale parameter	$\hat{\lambda}_1$ =0.0616	$\hat{\lambda}_2$ =0.0.0527	$\tilde{\lambda}$ =0.080238
log likelihood value	$L_1(\hat{\lambda}_1,\hat{\gamma}_1)$ =-51.2137	$L_2(\hat{\lambda}_2,\hat{\gamma}_2)$ =-64.3777	$L(\tilde{\lambda},\tilde{\gamma},\tilde{\lambda},\tilde{\gamma})$ =-123.735

3.2. Comparison of DN onset times of patients on the basis of gender and age at the time of diagnosis of diabetes by applying likelihood test

Further, we have classified the survival time, i.e. DN onset time, on the basis of gender. Out of 60 uncensored cases with diabetic nephropathy, there were 28 males and 32 females. Available diabetic data has been used to test the equality of male and female DN onset time, which is assumed to follow Weibull distribution with fixed left truncation time as 5 years. To find the likelihood ratio test statistic, X_L, and to test the hypothesis (5), we have found the maximum likelihood estimators of unknown parameters for male, female and combined groups and their respective log-likelihood values as given in Table 5. The observed value of the likelihood ratio test statistic, X_L, came out to be 85.056. Since test statistic X_L, is found to be greater than $\chi^2_{2,\alpha}$, $p < 0.001$, we reject H_0 and conclude that the two Weibull distributions are significantly different. It has been shown graphically that the DN onset times of two groups are significantly different, as displayed in Figure 1.

DN onset time is classified on the basis of age at diabetes diagnosis. Out of 60 uncensored cases with diabetic nephropathy, the age at diabetes diagnosis of 33 patients was less than or equal to 45 years and the age at diabetes diagnosis of 27 patients was greater than 45 years. Available data has been used to test the equality of the two groups, which are based on age at diabetes diagnosis and are assumed to follow Weibull distribution with fixed left truncation time as 5 years. To test the hypothesis (5), the above procedure is repeated to compute the log-likelihood ratio test statistic and its value came out to be X_L = 14.538. Again, since test statistic X_L is greater than $\chi^2_{2,\alpha}$, p <0.001, we reject H_0 and conclude that the DN onset times of two groups are significantly different or the two Weibull distributions are significantly different. The log-likelihood values and maximum likelihood estimators of the parameters are given in Table 5. It has been also shown graphically that the DN onset times of the two groups are significantly different, as displayed in Figure 2.

Thus, log-likelihood ratio test is found to be sufficient, as in all the above cases the data conclude that the DN onset times of two groups are significantly different.

Table 5. Maximum Likelihood estimators (MLE) of λ and γ along with log likelihood value for two groups when the grouping variable is (i) gender (ii) age at diabetes diagnosis

(i) Group Size	**Male** n_1 **=28**	**Female** n_2 **=32**	**Total uncensored** n_1+n_2 **=60**
MLE of shape parameter	$\hat{\gamma}_1$ = 3.113	$\hat{\gamma}_2$ = 1.561	$\tilde{\gamma}$ =2.811
MLE of scale parameter	$\hat{\lambda}_1$ =0.075	$\hat{\lambda}_2$ =0.062	$\hat{\lambda}$ =0.080
log likelihood value	$L_1(\hat{\lambda}_1,\hat{\gamma}_1)$=-81.761	$L_2(\hat{\lambda}_2,\hat{\gamma}_2)$=-60.993	$L(\hat{\lambda},\tilde{\gamma},\hat{\lambda},\tilde{\gamma})$=-185.283
(ii) Group Size	**Diagnosis ≤45yrs** n_1 **=33**	**Diagnosis >45yrs** n_2 **=27**	**Total uncensored** n_1+n_2 **=60**
MLE of shape parameter	$\hat{\gamma}_1$ = 3.221	$\hat{\gamma}_2$ = 3.014	$\tilde{\gamma}$ =2.811
MLE of scale parameter	$\hat{\lambda}_1$ =0.077	$\hat{\lambda}_2$ =0.072	$\hat{\lambda}$ =0.080
log likelihood value	$L_1(\hat{\lambda}_1,\hat{\gamma}_1)$=-94.142	$L_2(\hat{\lambda}_2,\hat{\gamma}_2)$=-83.871	$L(\hat{\lambda},\tilde{\gamma},\hat{\lambda},\tilde{\gamma})$=-185.283

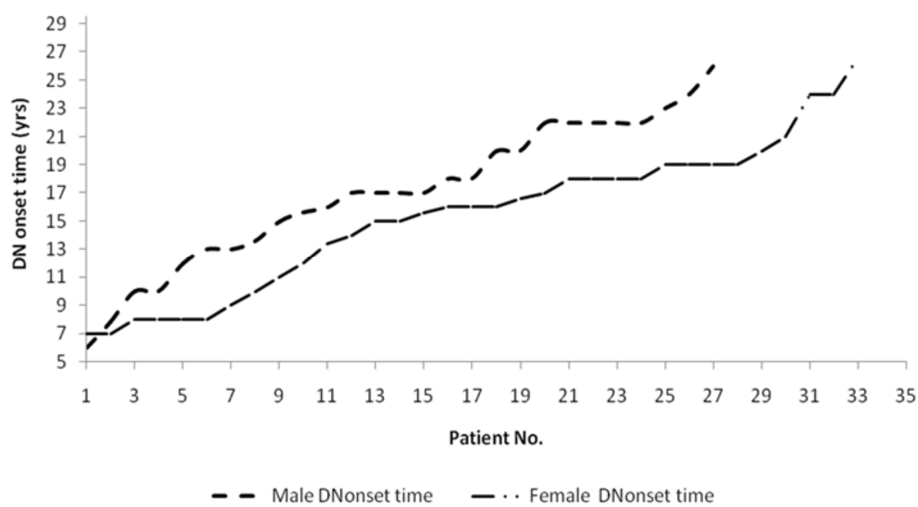


Figure 1. Onset time of diabetic nephropathy, uncensored cases; group1 =28 male patients; group2 = 32 female patients

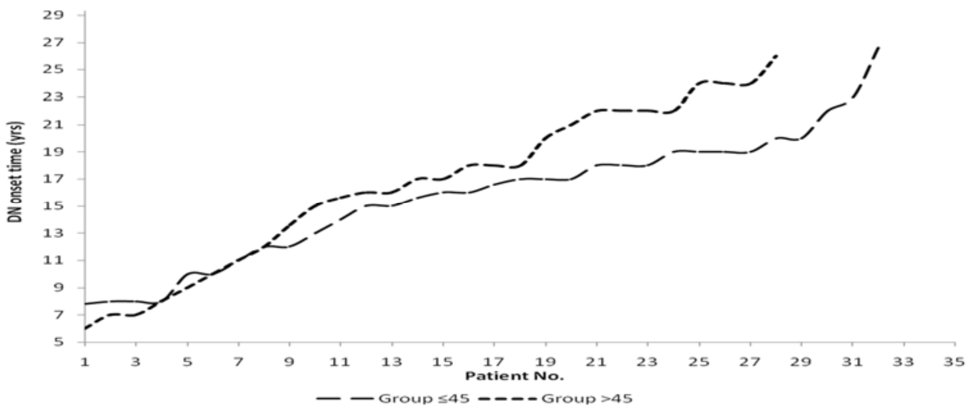


Figure 2. Onset time of diabetic nephropathy, uncensored cases; group1 =33, age at diagnosis less than or equal to 45years; group2 = 27, age at diagnosis greater than 45 years

3.2. Left truncated Weibull distributions are generated for the application of Thoman and Bain two sample test

In this section we have shown the application of two-sample test proposed by Thoman and Bain (1969), for cases when the likelihood ratio test is not sufficient. To compare the survival distributions of two groups of patients of same sizes with fixed left truncation, we generated three left truncated Weibull distributions with maximum likelihood estimators (MLE) of λ, γ parameters. The group1 is generated with MLE of λ, γ as $(\tilde{\lambda}, \tilde{\gamma})$ = (0.077, 3.221) for sizes 10, 15, 20, 30 and 40. The group2 is generated with MLE of λ, γ as $(\tilde{\lambda}, \tilde{\gamma})$ = (0.073, 3.0144) for same sizes. The combined group is generated with MLE of λ, γ as $(\tilde{\lambda}, \tilde{\gamma})$ = (0.080, 2.811) for the combined sample sizes as 20, 30, 40, 60 and 80. The log likelihood is computed for each size corresponding to each group, which is given in column 2, 4 and 6, respectively, in Table 6. The likelihood ratio test statistic has been obtained and simultaneously the significant value for each case has been computed as given in the last column in table 6. The cases where $p < 0.05$, for group sizes 30 and 40 likelihood ratio test has been found to be sufficient, but if $p > 0.05$, for group sizes 10, 15 and 20 we fail to reject the null hypothesis and the likelihood ratio test is not sufficient as likelihood ratio test is more appropriate for asymptotic cases.

Thoman and Bain two-sample test has been applied for cases with sample sizes 10, 15 and 20, respectively, where likelihood ratio test is not sufficient. According to this test, first the equality of the shape parameter is tested by finding $R = \frac{\hat{\gamma}_1}{\hat{\gamma}_2}$, under H_0. It has been found that in all the three sample sizes 10, 15 and

20, the test statistic R is less than ℓ_α at α =.05. Thus, we fail to reject the equality of shape parameters (Tables in Thoman and Bain, 1969). The respective values of R for the three sample sizes are given in Table 7. After accepting the equality of shape parameter, the equality of scale parameters is tested by finding the statistic G , for all the three sample sizes. Since the value of the test statistic G is found to be less than z_α at α =.05, we fail to reject H_0 (Tables in Thoman and Bain, 1969). Respective comparisons are given in Table 7. Thus, from the generated study we conclude that for sample sizes 10, 15 and 20, there is no significant difference between the two left truncated Weibull distributions, whereas for samples sizes 30 and 40, there is a significant difference between the two generated left truncated Weibull distributions.

Table 6. Generated left truncated Weibull distribution with log-likelihood values and test statistic values of samples with different sizes

Sample Size (group-1)	Generated Weibull	Sample Size (group-2)	Generated Weibull	Sample Size (combined group)	Generated Weibull	X_L	Significant Value
	$\hat{\lambda}_1$ =0.077 $\hat{\gamma}_1$ =3.221		$\hat{\lambda}_2$ =0.072 $\hat{\gamma}_2$ = 3.014		$\hat{\lambda}$ =0.080 $\tilde{\gamma}$ =2.811	$=2(\mathrm{L}_1(\hat{\lambda}_1,\hat{\gamma}_1) + \mathrm{L}_2(\hat{\lambda}_2,\hat{\gamma}_2) + L(\hat{\lambda},\tilde{\gamma},\hat{\lambda},\tilde{\gamma}))$	
	Log Likelihood		Log Likelihood		Log Likelihood		
10	-18.566	10	-27.945	20	-46.913	0.806	0.857*
15	-28.905	15	-25.655	30	-55.771	2.423	0.369*
20	-36.120	20	-41.717	40	-78.002	0.332	0.932*
30	-56.411	30	-57.292	60	-120.468	13.532	0.002
40	-78.775	40	-125.537	80	-214.058	19.492	<0.001

* In these cases Thoman and Bain test procedure has been used for comparing the two survival distributions.

Table 7. Thoman and Bain test procedure for comparing two survival distributions

Test	Statistic	Sample size	Inference
$H_0 : \gamma_1 = \gamma_2$ $H_1 : \gamma_1 > \gamma_2$	R^*=1.069	$n_1 = n_2$ =10	R <1.893, Fail to reject H_0
		$n_1 = n_2$ =15	R <1.688, Fail to reject H_0
		$n_1 = n_2$ =20	R <1.553, Fail to reject H_0
$H_0 : \lambda_1 = \lambda_2$ $H_1 : \lambda_1 > \lambda_2$	G^*=0.088	$n_1 = n_2$ =10	G <0.992, Fail to reject H_0
		$n_1 = n_2$ =15	G <0.704, Fail to reject H_0
		$n_1 = n_2$ =20	G <0.593, Fail to reject H_0

* $R = \frac{\hat{\gamma}_1}{\hat{\gamma}_2}$, $G = 0.5(\hat{\gamma}_2 + \hat{\gamma}_1)(Log(\hat{\lambda}_2) - Log(\hat{\lambda}_1))$

3.3. Weighted Kaplan-Meier method to compare DN onset time of two groups of patients, including censored cases

The KM method is useful to compare the survival function (with different weights) of two groups over different time. In this case, firstly, the 132 type-2 diabetic patients have been divided into two groups (on the basis of gender) of 59 males and 73 females. The duration of diabetes of 132 individuals is mutually independent. The DN onset times are known for 28 males and 32 females. The DN onset times of these 60 patients have been arranged as $t_1 < t_2 < ...., t_n$ in a pooled sample. The survival function has been estimated for both the groups and also for the pooled sample, using KM method. This survival function is redistributed according to the time interval as given in Table 8. To test the equality of the two survival distributions, we have found the weight function, as suggested by KM, at each t_i. The values of the weight function W_{KM}, estimate of the standard deviation $\hat{\sigma}_p$ and the test statistic Z came out to be 1.347, 0.204 and 6.599, respectively, as given in Table 8. The null hypothesis has been rejected since the p-value came out to be less than 0.001 (p < 0.001). Further, with the same objective but using a different weight function we have obtained the values of W_{AG} and $\hat{\sigma}_{AG}$, by replacing m, m_1 and m_2 by $n/m, n_1/m_1$ and n_2/m_2, respectively. The values of W_{AG}, $\hat{\sigma}_{AG}$ and test statistic Z, came out to be 0.998, 0.091 and 10.932, respectively. Again, the null hypothesis has been rejected since p-value came out to be less than 0.001 (p < 0.001). Thus, we conclude that the two survival functions are significantly different. We have also compared graphically the survival function (onset DN times) of male and female groups, as shown in Figure 3.

Now, secondly, the 132 diabetic patients have been divided into two groups (on the basis of age at diabetes diagnosis) of 81 patients whose age at diabetes diagnosis was less than or equal to 45 years and of 51 patients whose age at diabetes diagnosis was greater than 45 years. The DN onset times are known for 33 patients whose age at diabetes diagnosis was less than or equal to 45 years and for 27 patients whose age at with diabetes diagnosis was greater than 45 years. The DN onset times of these 60 patients have been arranged as $t_1 < t_2 < ...., t_n$ in a pooled sample. The survival function has been estimated for both the groups. This survival function is redistributed according to the time interval as given in Table 8. To test the equality of the two survival distributions, we have used the same procedure as given above, with two weight functions. The two test statistic, with different weight functions, came out to be 5.252 and 6.057, respectively, as given in Table 9. Thus, we reject the null hypothesis in both the cases, since p-value came out to be less than 0.001 (p < 0.001) and conclude that the two survival distributions are significantly different. We have also compared graphically the DN onset times of two groups, classified on the basis of age at diabetes diagnosis, as shown in Figure 4.

Table 8. Estimates of survival functions with group variable as gender and age at diabetes diagnosis, using KM method

Duration of diabetes	Survival function		Survival function		Pooled Survival Function
	Group1 =Male	Group2 =female	Group1=diabetes diagnosis age ≤45	Group1=diabetes diagnosis age >45	
6≤ t_i <7	0.983	1	1	0.979	0.992
7≤ t_i <8	0.958	0.966	0.982	0.937	0.963
8≤ t_i <9	0.958	0.885	0.915	0.914	0.918
9≤ t_i <10	0.958	0.863	0.894	0.888	0.906
10≤ t_i <11	0.906	0.841	0.851	0.888	0.870
11≤ t_i <12	0.906	0.794	0.808	0.859	0.845
12≤ t_i <13	0.850	0.794	0.786	0.859	0.819
13≤ t_i <14	0.793	0.768	0.764	0.802	0.778
14≤ t_i <15	0.765	0.743	0.743	0.773	0.751
15≤ t_i <16	0.708	0.666	0.653	0.716	0.684
16≤ t_i <17	0.680	0.559	0.572	0.656	0.616
17≤ t_i <18	0.566	0.536	0.497	0.567	0.533
18≤ t_i <19	0.507	0.512	0.429	0.504	0.459
19≤ t_i <20	0.507	0.442	0.304	0.504	0.398
20≤ t_i <21	0.439	0.421	0.243	0.465	0.352
21≤ t_i <22	0.439	0.301	0.243	0.465	0.352
22≤ t_i <23	0.256	0.271	0.195	0.310	0.248
23≤ t_i <24	0.192	0.271	0.130	0.310	0.124
24≤ t_i <25	0.128	0.090	0.130	0.078	0.083
t_i≥25	0.064	0	0.065	0	0.041

Table 9. Comparison of Diabetic Nephropathy onset time of two groups with different weight functions

Grouping variable	Weight function	Statistic	Sqrt(variance)	Z
Gender	$w(t)=\dfrac{m\hat{H}_1(t)\hat{H}_2(t)}{m_1\hat{H}_1(t)+m_2\hat{H}_2(t)}$	W_{KM} =1.347	$\hat{\sigma}_p$ =0.204	6.599
	$w_{AG}(t)=\dfrac{(n/m)\hat{H}_1(t)\hat{H}_2(t)}{(n_1/m_1)\hat{H}_1(t)+(n_2/m_2)\hat{H}_2(t)}$	W_{AG} =0.998	$\hat{\sigma}_{AG}$ =0.091	10.932
Diabetes diagnosis age	$w(t)=\dfrac{m\hat{H}_1(t)\hat{H}_2(t)}{m_1\hat{H}_1(t)+m_2\hat{H}_2(t)}$	W_{KM} =1.529	$\hat{\sigma}_p$ =0.291	5.252
	$w_{AG}(t)=\dfrac{(n/m)\hat{H}_1(t)\hat{H}_2(t)}{(n_1/m_1)\hat{H}_1(t)+(n_2/m_2)\hat{H}_2(t)}$	W_{AG} =1.085	$\hat{\sigma}_{AG}$ =0.179	6.057

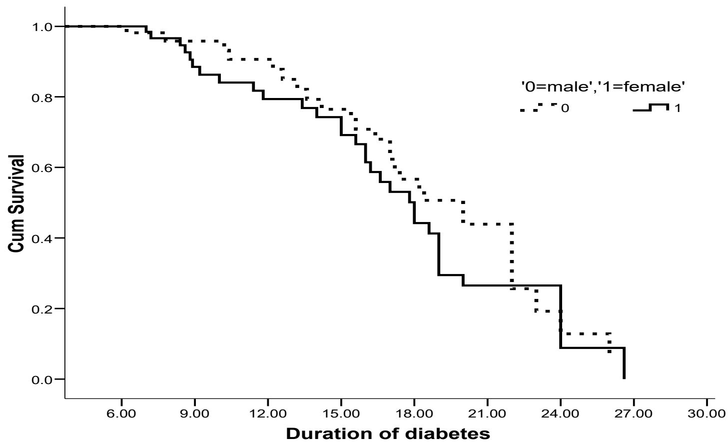


Figure 3 Comparison of DN onset time of male and female type-2 DM patients, using the KM estimator

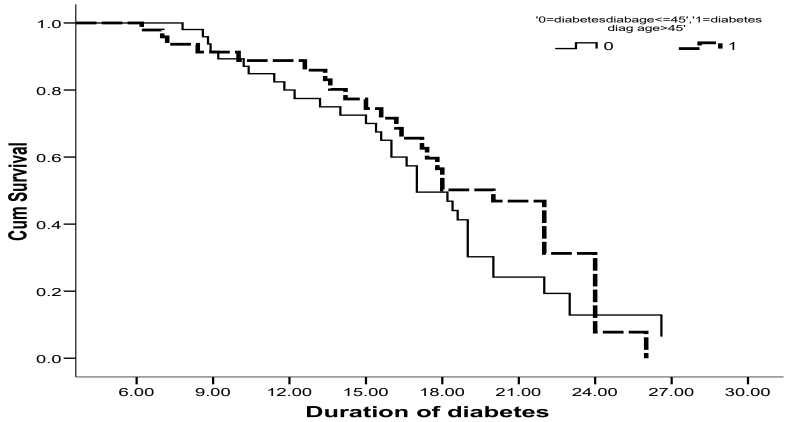


Figure 4 Comparison of DN onset time of diabetic patients with age at diabetes diagnosis ≤45years and diabetes diagnosis age>45years, using the KM estimator

5.3.4. Modified Brookmeyer and Crowley method to compare DN onset time of two groups of patients based on the median survival function of DN onset time

In this part, an alternative method has been used based on median survival time rather than comparing the difference in the survival functions over time. Again, firstly, the 132 type-2 diabetic patients have been divided into two groups of 59 males and 73 females. Using KM method survival function has been estimated at each event point for male, female and pooled sample. We have arranged the onset times in ascending order in a pooled sample and in this case it has been found that there exist no t_i, for which $\hat{S}_p(t_i) = 0.5$. Then, to find the estimated median survival time $\hat{M}$, we have computed M_L, which is the largest event time with $\hat{S}_p(M_L)$ > 0. 5 and M_U be the smallest event with $\hat{S}_p(M_U)$ < 0.5. Then, the median lies in the interval (M_L, M_U) and is obtained by linear interpolation, i.e. M_L and M_U as 18.0 and 18.2, respectively, and the corresponding values of $\hat{S}_p(M_L)$ and $\hat{S}_p(M_U)$ came out to be 0.502 and 0.488, respectively. Thus the median lies in the interval (18, 18.2) and by linear interpolation it came out to be 18.034. Using the survival time for both the samples, we have computed T_{Lj} and T_{Uj}, the death time in the j^{th} sample such that the estimated median survival time lies between these two death times. We have found that the onset time in the male sample satisfied the condition $T_{Lj} \le \hat{M} < T_{Uj}$ as the median lies in between 18 and 20 (18<18.034<20). The estimated probability of survival, $\hat{S}_1(\hat{M})$, by linear interpolation, beyond $\hat{M}$ in the male sample came out to be 0.508. To find the variance we have obtained the values of V_1 and V_2 as 0.016 and 0.0002, respectively. Finally, the variance and the test statistic came out to be 0.515 and 0.764, respectively, as given in Table 9. Thus, we fail to reject the null hypothesis as p= 0.447 and conclude that there is no significant difference in the median survival time of DN onset time of male and female groups.

The above test, which is based on median survival time, can also be applied when data is divided into two groups on the basis of age at diabetes diagnosis. Out of 132 individuals there were 81 patients whose age at diabetes diagnosis was less than or equal to 45 years and there were 51 patients whose age at diabetes diagnosis was greater than 45 years. The median survival time came out to be, i.e. 18.034, as it is based on the pooled sample. Using the survival time for both the samples, we have found that the onset time of patients whose age at diabetes diagnosis was greater than 45 years satisfy the condition $T_{Lj} \le \hat{M} < T_{Uj}$ as 18<18.034<20. The estimated probability of survival, $\hat{S}_1(\hat{M})$, by linear interpolation, beyond $\hat{M}$ for patients whose age at diabetes diagnosis was greater than 45 years, came out to be 0.503. To find the variance, we first obtained the values of V_1 and V_2 as 0.014 and 0.012, respectively. Finally, the variance and the test statistic came out to be 0.827 and 0.475, respectively, as given in Table 10. Thus, we fail to reject the null hypothesis as p=0.496 and conclude that there is no significant difference in the median survival time of DN onset times of age at diabetes diagnosis less than or equal to 45 years and age at diabetes diagnosis greater than 45 years.

Table 10. Comparison of DN onset time between two groups based on median survival time

Grouping variable	Median survival time based on weighted sample	Variance	Test Statistic
Gender	18.034	0.515	0.764
Age at diabetes diagnosis	18.034	0.827	0.475

4. Discussion

The aim of this study is to compare the survival time (onset time of nephropathy from the diagnosis of type-2 diabetes) of two groups of patients with type-2 diabetes. The major advantage of developing methods for comparing the DN onset time of type-2 DM patients is that the DN onset time pattern of new DM patients can be predicted depending on the patient group. Also, it can be used as a baseline for further studies.

Firstly, we have used parametric methods to compare the survival times of two groups, on the basis of gender and age at the time of diagnosis of diabetes, assuming that DN onset times follow Weibull distribution. The two parameters of the Weibull parametric distribution provide additional flexibility that potentially increases the accuracy of the description of collected survival data, since the shape parameter allows the hazard function to increase or decrease with increasing time (Collet, 2003). The likelihood ratio test is applied here on collected data of type-2 DM patients with minimum 5 years of duration of diabetes, to compare groups: male and female and of patients whose age at diabetes diagnosis is less than or equal to 45 years and greater than 45 years. This test is widely used in comparing two survival distributions for the cases where sample sizes are not small and the equality of two Weibull distributions is rejected (Lee, 2003). We have applied two sample test of Thoman and Bain (1969) to compare small samples of equal sizes, and cases where we fail to reject the equality of shape and scale parameter. Thus, in some cases limitation of likelihood ratio test can be handled with Thoman and Bain test. If the data comes from a Weibull distribution, the most accurate way in this case is to use a parametric test such as likelihood ratio test or the method of comparing maximum likelihood estimates proposed by Thoman and Bain (Lam and Spelt, 2007). In all the above tests we have considered only uncensored cases. Clearly, considering only uncensored cases will increase the mean of the survival time (Li and Lagakos, 1997).

To avoid the model validity issues for left truncated and right censored data (LTRC), we have used the nonparametric approach, supported by the well-developed Kaplan-Meier product limit estimator and related techniques, in the second part of the methods, to compare the survival functions of two groups, on the basis of gender and age at the time of diagnosis of diabetes. We have used the KM estimator to estimate the survival function for both the groups because this estimate is the most important and widely used method to estimate the survivor function and it is a generalization of the empirical survivor function, which accommodates censored observations also (Collet, 2003). The KM method with weights W_{KM}, a censored data analog of the two sample t-test, suggest a test which down weights differences late in time when there is heavy censoring (Klien and Moeschberger, 2003). The weight W_{AG} , suggested by us, using the KM method also includes the ratio in which censored observations appear in the data. This method helps us to compare the survivor function of two groups over different time periods, as the data includes patients with minimum 5.6 years and maximum 27-year duration of disease.

Another method to compare the survival function of two groups with LTRC data, based on median survival function, derived by Brookmeyer and Crowley, is where median survival time is based on the common survival function, $S_w(t) = \frac{m_1\hat{S}_1(t) + m_2\hat{S}_2(t)}{m}$. We have modified $S_w(t)$, by pooled survival function, which meets the situation when the number of events in $\hat{S}_1(t)$ and $\hat{S}_2(t)$ is different as well as censoring and event pattern of group is different. But this pooled survival function depends on the censoring patterns in the two samples. It is found that two groups are not significantly different when we have compared the median survival function of DN onset times. Since the distribution of survival time tends to be positively skewed, the median is the preferred summary measure of the location of the distribution (Collet, 2003).

From the parametric and nonparametric methods used in this study we conclude that the DN onset time of male group differs significantly from that of the female group. Also, the DN onset time of patients whose age at diabetes diagnosis is less than or equal to 45 years differs significantly from that of the group whose age at diabetes diagnosis is more than 45 years. For future studies, the procedures used in this paper can be used for the comparison of survival time of two independent groups in any biomedical study. The application of these models to other complex event history data can also be explored.

REFERENCES:

ASHFAQ, S., GHAZZAL, Z. DOUGLAS, J. S., MORRIS, D. C., VELEDAR, E., WEINTRAUB, W. S., (2006). Impact of diabetes on five-year outcomes after vein graft interventions performed prior to the drug-eluting stent era. The journal of invasive Cardiology, 16(3), pp. 100–105.

COLLETT, D., (2003). Modeling survival data in medical research. 2nd ed. London: Chapman and Hall/CRC.

CZEKALSKI, S., (2005). Diabetic nephropathy and cardiovascular diseases. Annales Academiae Medicae Bialostocensis, 50, pp. 122–125.

FELLINGER, E. K., LEAVITT, B. J., HEBERT, J. C., (2002). Serum levels of prophylactic cefazolin during cardiopulmonary bypass surgery. Ann. Thorac. Surg., 74, pp. 1187–1190.

GROSS, J. L., DE AZEVEDO, M. J., SILVEIRO, S. P., CANANI, L. H., CARAMORI, M. L., ZELMANOVITZ,T., (2005). Diabetic nephropathy: diagnosis, prevention, and treatment. Diabetes Care, 28, pp. 164–176.

JOSS, N., PATERSON, K. R., DEIGHAN, C. J., SIMPSON, K. AND BOULTON-JONES, J. M., (2002). Survival Analysis Techniques for Censored and Truncated data. 2nd ed. Springer Science Business Media Inc.

LAM, R. AND SPELT, J. K., (2007). Comparison of Weibull Small Samples using Monte Carlo Simulations. Quality and Reliability Engineering International, 23, pp. 503–513.

LEE, E. T., WANG, W. Y., (2003). Statistical Methods for Survival Data Analysis, 3rd ed. New Jersey: John Wiley & Sons.

LI, Q. H., LAGAKOS, S. W., (1995). Use of the Wei-Lin-Weissfield Method for the Analysis of a Recurring and a Terminating Event. Statistics in Medicine, 16, pp. 925–940.

PEPE, M. S., FLEMING,T. R., (1989). Weighted Kaplan-Meier Statistics: A Class of Distance Tests for Censored Survival Data. Biometrics; 45(2), pp. 497–507.

ROSSING, P., TARNOW, L., NIELSEN, F. S., BOELSKIFTE, S., BRENNER, B. M., PARVING, H-H., (1995). Short stature and diabetic nephropathy. BMJ, 310, pp. 525–528.

THOMAN, D. R. BAIN, L. J., (1969). Two sample tests in the Weibull Distribution. Technometrics, 11 (4), pp. 805–815.

VILLAR, E., CHANG, S. H., MCDONALD, S. P., (2007). Incidences, Treatments, Outcomes, and Sex Effect on Survival in Patients With End-Stage Renal Disease by Diabetes Status in Australia and New Zealand (1991–2005). Diabetes Care, 30 (12), pp. 3070–3076.

VISWANATHAN, V., (2004). Prevention of diabetic nephropathy: A diabetologist's perspective. Indian J. Nephrol, 14, pp. 157–162.

WILD, S., GREEN, G. A., SICREE, R., KING, H., (2004). Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030. Diabetes Care, 27, pp. 1047–1053.

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 551–561, DOI 10.21307/stattrans-2018-030

CHANNEL PERFORMANCE UNDER VENDOR MANAGED CONSIGNMENT INVENTORY CONTRACT WITH ADDITIVE STOCHASTIC DEMAND

Milena Bieniek[1],

ABSTRACT

Consignment as the shifting of the inventory ownership to a supplier is widely implemented in virtual market. In this form of business arrangement the supplier places goods at a retailer's location without receiving payment, until the goods are sold. We consider a single period supply chain model, where the supplier contracts with the retailer with some probability of return. Market demand is additive, linearly price–dependent and uncertain. We focus on vendor managed consignment inventory (VMCI) channel, in which the supplier decides the consignment price and his service level and the retailer chooses the retail price. We study channel performance under VMCI setting by analysing how the model parameters influence decision quantities, channel profit and risk function. We also illustrate the obtained results by a numerical example, which explains the overall solutions well.

Key words: stochastic demand, supply chain, consignment, operations management.

1. Introduction

Supply chain management has been one of the major tasks for management professionals. The top practice for reducing the inventory cost is using the consignment, which is shifting the inventory ownership to the suppliers. The consignment is the process of placing goods in the retailer's location and no payment is made to the supplier before the item is sold. Hence, the retailer faces lower risk associated with uncertain demand, since he has no money tied up in inventory. This arrangement is called vendor managed consignment inventory (VMCI). The pioneer of VMCI is Wal-Mart, but it is applied also by on-line market places such as Amazon.com or eBay.com (Li, Zhu and Huang (2009)).

The aim of our paper is to study the decisions and channel performance for VMCI contract with additive random demand. We ask the questions how the firms interact in the channel and how system parameters affect channel performance. We focus on the type of VMCI setting, which is introduced in Ru and Wang (2010) The authors of this paper build a newsvendor type game–theoretic model to capture the connections between the supplier and the retailer, when the supplier controls the supply channel inventory. Market demand for the product is random and price–sensitive. Both the supplier and the retailer incur a linear cost for producing and

[1]Maria Curie-Sklodowska University, Plac Marii Curie-Sklodowskiej 5, 20-031 Lublin, Poland. E-mail: milena.bieniek@umcs.lublin.pl

handling the product. The supplier offers the consignment price charged to the retailer before the demand is realized and decides on the inventory level. At the same time, the retailer chooses the retail price. In Ru and Wang (2010) the authors adopted multiplicative demand function, which let them give the precise solutions. The recent paper considering similar VMCI arrangement is Hu, Chen and Yu (2015), where the variability of the channel performance with respect to the changes of the model parameters is studied. The authors use known expressions obtained in Ru and Wang (2010) and consider the uncertainty in customer return. They also discuss how the probability of return influences the price, inventory decisions and profit function. Moreover, they calculate a semi-deviation profit as a risk measure. Our study is similar to those given in Hu, Chen and Yu (2015), but it is done for the additive demand. We continue the subject of Bieniek (2017), where the precise solutions for the optimal channel components are calculated. Now we use a slightly different model with the probability of return. The findings for the risk function give important insights for the retailer or the supplier to consider, when implementing the consignment contract can give some benefits in expected profit. Furthermore, we derive for additive demand the risk measure based on a semi-deviation of profit. Finally, we analyse the numerical example using uniform distribution.

In the following we provide a review of the papers that are closely related to our study. In the paper Lee and Chu (2005) the authors present the consignment contract with inventory ownership and try to answer to the question who should control the supply chain: the supplier or the retailer. In the consignment settings studied in that paper the wholesale and the consignment prices are exogenously given. In Wang, Jiang and Shen (2004) the authors consider the consignment contract with revenue sharing between the supplier and the retailer. In this consignment arrangement the supplier retains ownership of the inventory bearing all risk. The retailer specifies the percentage allocation of sales revenue and, at the same time, the supplier chooses the product quantity and the retail price. In their study they use an iso-elastic random and price dependent demand. In Hu, Li and Govindan (2014) the return policies are added to the VMCI contract. Lately, in Olah et al. (2017) VMCI arrangement and it's benefits for the supply chain participants are widely described.

In the summary our paper contributes to the literature by: 1. considering uncertain return behaviour in VMCI contract in an additive random demand framework 2. investigating the risk analysis 3. figuring out how the uncertainty of customer return and other model parameters influence the decision variables, profit function and risk function 4. deriving managerial insights for expected profit and risk aspect. The paper is organized as follows. In Section 2 we present general assumptions and recall the definitions of the notions used in the subsequent analysis. In Section 3 we consider the performance of the decentralized channel under VMCI contract. In this section we recall the main statements of Bieniek (2017) and then we study the variability of the obtained results analytically. Section 4 is devoted to a numerical example and finally Section 5 concludes the paper.

2. General assumptions

We consider a single–period supply chain, in which the supplier (vendor) produces and sells a product to the retailer. The supplier decides his consignment price w, charged to the retailer for each unit sold. The retailer chooses the retail price p for selling the product to consumers. Denote by c_s the supplier's unit production cost and by c_r the retailer's unit handling cost. Also define $C = c_s + c_r$ as the total unit cost for channel and $\alpha = c_r/C$ as the share of channel cost that is incurred by the retailer. We assume no product shortage costs. The market demand is defined by $D(p,\varepsilon) = a - bp + \varepsilon$, where $a,b > 0$. Here ε is a continuous random variable with the expected value μ, the cumulative distribution function $F(.)$ and the probability distribution function $f(.)$, which are defined on the support $[A,B]$, where $A < 0$ and $B > 0$. The general assumptions are as follows: $C < p < p_{max}$, where $p_{max} = max_{p:a-bp+A>0} = \frac{A+a}{b}$ (cf. Rubio-Herrero, Baykal-Gursoy and Jaskiewicz (2015)) and $A + a - bC > 0$. The assumptions guarantee that realization of the demand $D(p,\varepsilon)$ is positive. The linear additive demand has different features than the iso-price-elastic demand, studied in other papers cited in our study. The additive model does not preserve iso-price-elasticity property. Its price elasticity index of the expected demand is given by $bp/(\mu + a - bp)$. The price-elasticity index is for linear demand increasing in b at any price p, so one can consider the parameter b as a surrogate of the price elasticity index like for instance in Wang, Jiang and Shen (2004).

We add consumer returns to the model studied in Bieniek (2017) and assume that the product is returned with the probability k (cf. Mostard and Teunter (2006)). The customer gets a generous full refund. At the end of the selling season the retailer pays the supplier based on the net sale, which is equal to total sales minus total returns. It appears that customer returns are very often in the e-shops. It happens for several reasons. First of all when buying via Internet the customers do not see physically the product, which often turns out to be in the wrong size or colour. Additionally, in the virtual market customers make the decision very quickly without taking it under deep consideration. But according to the legislation in many countries it is allowed to return the purchased product without giving any reason in a given time (eg two weeks).

We use the following expressions. Let c be such that $c = \frac{C}{1-k}$. Also let $z = Q - (a - bp)$ and $\mu(z) = \mu + \int_z^B (z-u)f(u)du$. As indicated in Petruzzi and Dada (1999) the quantity z can be interpreted as a safety stock since for the selected value of z we face shortages if $z < \varepsilon$ or leftovers if $z > \varepsilon$. Also z corresponds to a customer service level given by $P(D(p,\varepsilon) \leq Q) = P(\varepsilon \leq Q - (a - bp) = z) = F(z)$. The variability of the function $\mu(z)$ is important for the following analysis. Note that $\frac{d\mu(z)}{dz} = 1 - F(z)$, $\mu(.)$ is increasing in $z \in [A,B]$, $\mu(A) = A < 0$ and $\mu(B) = \mu$. We need also the lost sales rate elasticity (LSR) concept. This notion is provided by Kocabiyikoglu and Popescu (2011) and it is the percentage change in the rate of lost sales with respect to the percentage change in the price for a given quantity. The LSR elasticity combines the relative sensitivity of the lost sales with respect to its underlying factors, the price and the inventory.

Definition 1. *The LSR elasticity for a given price $p(z)$ and a service level z for additive price–dependent demand is defined as*

$$\kappa(p(z),z) = \frac{bp(z)f(z)}{1-F(z)}.$$

3. Channel performance under VMCI contract

3.1. General solutions for VMCI

Some statements given in this subsection are proved in Bieniek (2017). The novelty is the introduction of the probability of return to the model and the derivation of expressions for the risk measure.

Under VMCI contract, the supplier makes the decisions and he is the owner of the goods until they are sold, but physically they are stored in the retailer's location. An illustrative example of VMCI is as follows. Company S (supplier) provides his product on consignment and it is responsible for manufacturing costs. Under consignment, company S is no longer responsible for material handling. From now the retailer has responsibility for any damage of goods on their properties. Company S makes the inventory decision and sets the consignment price. The retailer is able to choose the retailer price. The supplier is favoured in VMCI contract and he is the Stackelberg leader, and hence he obtains more profit. There are three steps in VMCI setting with return. Namely, in Step 1 the supplier chooses the order quantity and the consignment price charged for the retailer for each unit sold. In Step 2 the retailer decides the retail price. In Step 3 if some consumers return the product, then the retailer pays the supplier based on net sales.

The optimization problem is addressed following backward induction. First we focus on the retailer's optimization part (Step 2) and then on the supplier's optimization problem (Step 1). In Step 2 for a given consignment price and a given service level, chosen by the supplier in Step 1, the retailer determines the retail price, which maximizes his own expected profit given by

$$\Pi_R(p|w,z) = (p-w)(1-k)E(\min(D(p,\varepsilon),Q)) - C\alpha Q.$$

This expression is equivalent to

$$\Pi_R(p|w,z) = (1-k)\{(p-w)(\mu(z)+a-bp) - c\alpha(z+a-bp)\}.$$

The total sales of the returned items is equal to $(1-k)E(\min(D(p,\varepsilon),Q))$. Assuming that the retailer does not get profit from the returned items, he can obtain $p-w$ from the each net sale. Under the above assumptions in Step 2 for any given service level $z \in [A,B]$ and a consignment price $w > 0$, the retailer's unique optimal retail price $p_d^*(w,z)$ is given by

$$p^*(w,z) = \frac{\mu(z)+a+bc\alpha+bw}{2b}. \tag{1}$$

In Step 1, knowing that the retailer's optimal price p^* is determined by (1), the supplier's aim is to set the optimal consignment price w^* and the optimal service level z^*, which maximize his own expected profit. The supplier's expected profit is equal to

$$\Pi_S(w,z|p) = (1-k)\{w(\mu(z)+a-bp) - c(1-\alpha)(z+a-bp)\}.$$

It is proved, that for any given service level z, the supplier's unique optimal consignment price $w^*(z)$ maximizing $\Pi_S(w,z|p)$ has the form

$$w^*(z) = \frac{\mu(z)+a+bc(1-2\alpha)}{2b}. \tag{2}$$

If LSR elasticity satisfies

$$\kappa(w^*(z)+c(1-\alpha),z) \geq \frac{1}{2} \quad \text{for any} \quad z \in [A,B], \tag{3}$$

then the service level z^* is uniquely determined by

$$\frac{\mu(z^*)+a-4bc\alpha+3bc}{2b} = \frac{2c(1-\alpha)}{1-F(z^*)}.$$

Putting the formula (2) into (1) we get $p^*(z^*) = \frac{3\mu(z^*)+3a+bc}{4b}$. Thus the total channel expected profit is equal to

$$\Pi^* = \Pi_R^* + \Pi_S^* = (1-k)\left\{\frac{3\mu(z^*)+3a+bc}{16b}(\mu(z^*)+a+3bc) - c(z^*+a)\right\}.$$

We see that (3) is equivalent to $f(z^*)(\mu(z^*)+a+3bc-4bc\alpha)-(1-F(z^*)) > 0$. Moreover, under the general assumptions $p^*(z^*) \leq \frac{3\mu(B)+3a+bc}{4b} \leq \frac{A+a}{b}$, which gives $4A+a-bc-3\mu > 0$.

In our model a supply chain bears demand risk. We consider 1-st central-semi-deviation of profit as a risk measure defined in Ogryczak and Ruszczynski (2001) and given by $\delta = E(max(\Pi-\pi),0)$, where π is a random profit and Π is the expected profit. We figure out the risk sharing by looking at $\delta_R = Emax(\Pi_R-\pi_R,0)$. It is known that $\pi_R = (1-k)((p-w)(min(\varepsilon,z)+a-bp)-c\alpha(z+a-bp))$, which implies

$$\delta_R^* := \delta_R(p^*,w^*,z^*) = (1-k)(p^*-w^*)\int_A^{\mu(z^*)}(\mu(z^*)-\varepsilon)f(\varepsilon)d\varepsilon.$$

Similarly we obtain, $\delta^* = \delta_R^* + \delta_S^* = (1-k)p^*\int_A^{\mu(z^*)}(\mu(z^*)-\varepsilon)f(\varepsilon)d\varepsilon$. Then the retailer share of channel risk defined by $\beta = \frac{\delta_R^*}{\delta^*}$ for additive demand is equal to

$$\beta = \frac{p^*-w^*}{p^*} = \frac{\mu(z^*)+a-bc+4bc\alpha}{3\mu(z^*)+3a+bc}.$$

3.2. Sensitivity analysis of VMCI channel

Now we analyse how the model parameters influence the channel performance. All proofs of the propositions from this subsection are given in Appendix.

Proposition 1. *If* (3) *holds then* z^* *increases in* α *and for* $\alpha > \frac{1}{2}$ *it decreases in* b*,* k *and* C*.*

The optimal service level z^* increases if the share of channel cost increases and then the retailer shares larger part of channel cost than the supplier. If the price elasticity increases then the service level decreases. Moreover, when uncertainty increases, the amount of the returned items increases and the net sale decreases. Consequently, the decision maker has to stock less of an item.

Proposition 2. *If* (3) *holds then* p^* *increases in* α *and it decreases in* b*. The monotonicity of* p^* *in* k *and* C *depends on other parameters.*

The optimal retail price is increasing with the share of channel cost and decreasing with the price elasticity, which is the same as the conclusions for the service level. But the variability with respect to the probability of return and the total cost depends on other parameters.

Proposition 3. *If* (3) *holds then the monotonicity of* w^* *in* α *depends on other parameters and* w^* *decreases in* b*. Moreover, if* $\alpha > \frac{1}{2}$ *then* w^* *decreases in* C *and* k*, otherwise it's monotonicity depends on other parameters.*

The conclusion for the consignment price regarding the influence of the price sensitivity is similar to those obtained for the retail price.

Proposition 4. *If*

$$\kappa\left(w^*(z)+c\left(\alpha+\frac{1}{3}\right),z\right)\geq\frac{1}{2}\quad \text{for any}\quad z\in[A,B], \tag{4}$$

and (3) *holds, then* Π^* *is increasing-decreasing both in* z^* *and* α*.*

Remark 1. *The constraint* (4) *is equivalent to* $(\mu(z)+a+\frac{5}{3})f(z)\geq(1-F(z))$*, which is independent on* α*.*

The crucial part of the above proposition is that the optimal channel profit is first an increasing function and then a decreasing function of z^*. Other statements are the consequence of that property. This is different than for the multiplicative model, where the channel profit always increases in z. It should be underlined that for additive demand the propositions hold only for selected values of model parameters, which satisfy (3) and (4) due to the very demanding assumption of the profit concavity. Finding the milder assumptions can be the topic of a future research.

Proposition 5. *If* (3) *holds then* β *increases in* z^* *for* $\alpha < \frac{1}{3}$ *and it decreases otherwise. Moreover, it increases in* α *for* $\alpha < \frac{1}{3}$ *and it's monotonicity depends on other parameters otherwise.*

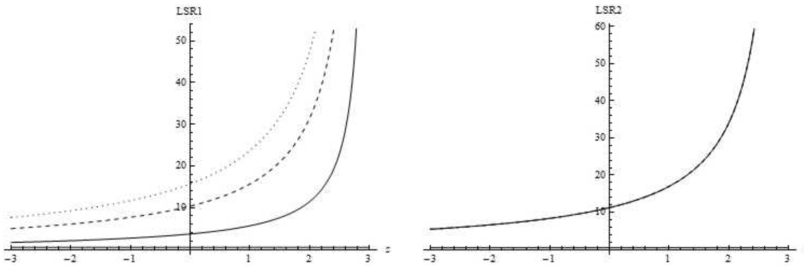


Figure 1: LSR elasticity $\kappa(w^*(z)+c(1-\alpha),z)$ for $\alpha=0.9$ (solid), $\alpha=0.4$ (dotted) and $\alpha=0.1$ (dashed) (left) and $\kappa(w^*(z)+c(\alpha+1/3),z)$ (right) with respect to z

From the above statement we see that a higher risk - higher expected profit phenomenon does not always apply to the retailer and it depends critically on his share of channel cost. For $\alpha<\frac{1}{3}$ the higher retailer share of risk corresponds to the higher share of channel cost or equivalently to the lower retailer's profit.

4. Numerical example

In order to illustrate the results previously obtained we proceed with a numerical example. We use uniformly distributed demand on the interval $[-3,3]$. Moreover, we set the base values equal to $a=35$, $b=1$, $C=10$, $k=0.5$ (then $c=20$). Using the figures we present how the model parameters influence the optimal solutions. We investigate the variability in one parameter with other base parameters. All figures justify the statements of the propositions. First we state that α should be such that (3) and (4) are satisfied. This can be confirmed by the fact, that the figures of the elasticities should lie above the line $y=\frac{1}{2}$ (Fig. 1).

Figure 2 shows the variability of the service level. In the figure on the left we can see that the service level increases in α. In the figure of the right we observe that if b changes from 0.1 to 1.2 then the service level decreases for different values of α. Figure 3 presents the variability of the retail price. We see that the retail price increases if α increases and decreases in b. Finally, on Figure 4 it is shown that the optimal expected profit is increasing-decreasing in α and it decreases in b.

5. Conclusions

In this paper we study the vendor managed inventory contract with consignment (VMCI), which is widely applied in many industries, including virtual market. In VMCI arrangement the upstream supplier owns the product until it is sold by the downstream retailer. The supplier decides on the consignment price and the inventory level. The retailer chooses the retail price.

Our study is mainly based on Ru and Wang (2010), Hu, Chen and Yu (2015) and Bieniek (2017). VMCI program studied by Ru and Wang (2010) uses multiplicative

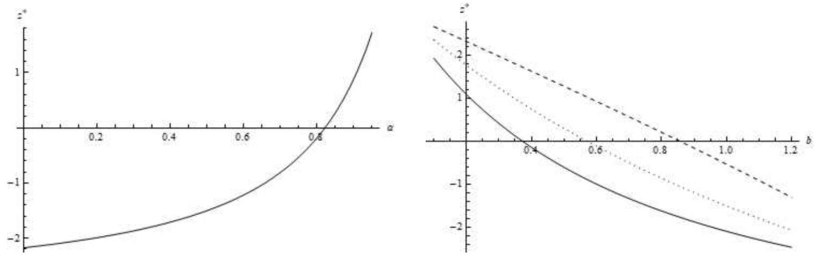


Figure 2: Service level with respect to α (left) and b for $\alpha = 0.1$ (solid), $\alpha = 0.5$ (dotted) and $\alpha = 0.75$ (dashed) (right)

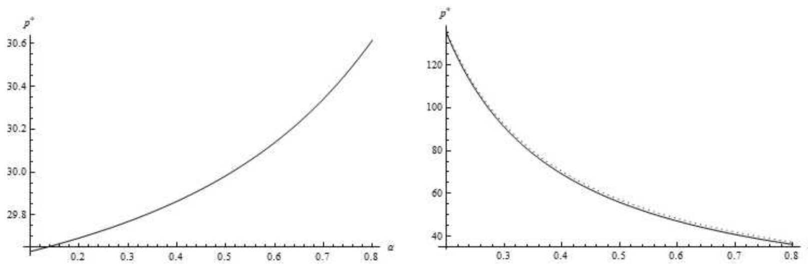


Figure 3: Price with respect to α (left) and b for $\alpha = 0.1$ (solid), $\alpha = 0.75$ (dotted) (right)

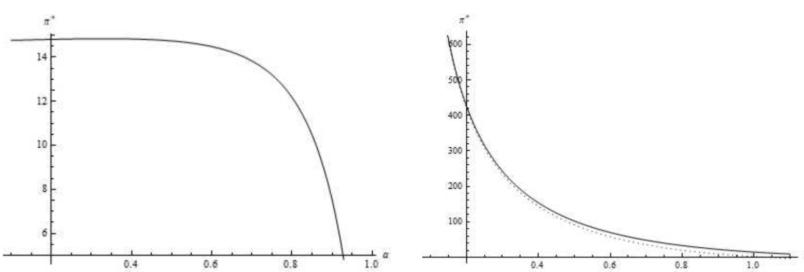


Figure 4: Profit with respect to α (left) and b for $\alpha = 0.05$ (solid), $\alpha = 0.9$ (dotted) (right)

random demand, which is exponentially dependent on the price. The additive random demand linearly dependent on the price is considered in Bieniek (2017). In that paper the precise formulas are given, but the sensitivity of the optimal solutions is not examined. We consider this sensitivity in our paper, where also an opportunity of return with a given return probability is added to the model. Our work is done in the same light as in Hu, Chen and Yu (2015), but for additive demand form. Precisely, we show how the share of channel cost, probability of return and other parameters influence the price, inventory decisions and profit function. We discuss also the retailer share of channel risk and it's dependence on the share of channel cost. We use the central-semi deviation risk measure discovered by Ogryczak and Ruszczynski (2001). The assumptions involve the lost sales rate elasticity, which has an economic interpretation. In the last part we consider the numerical example. We decide to present rather figures than tables since they illustrate the results more clearly. We use uniformly distributed demand with mean zero and the model parameters satisfying all of the assumptions. It can be seen that the presented figures justify the statements of the established propositions.

Our research proves that the form of the demand influences on the supply chain performance. Our findings for risk function give important insights for the retailer or the supplier to consider, when implementing consignment contract can give some benefits in the expected profit. As a future research the considerations on models with multiple suppliers or competing retailers could be done.

6. Appendix

Proof.[of Proposition 1] From (3.1) we have $G(\alpha,z^*)=0$, where $G(\alpha,z^*)=(\mu(z^*)+a-4bc\alpha+3bc)(1-F(z^*))-4bc(1-\alpha)$. By implicit function theorem we get $\frac{dG}{d\alpha}+\frac{dG}{dz^*}\frac{dz^*}{d\alpha}=0$, which gives

$$\frac{dz^*}{d\alpha}=\frac{4bcF(z^*)}{f(z^*)(\mu(z^*)+a+3bc-4bc\alpha)-(1-F(z^*))^2},$$

which by (3) is positive. Now let $m=bc=\frac{bC}{1-k}$. Then $G(m,z^*)=(\mu(z^*)+a+m(3-4\alpha)(1-F(z^*))-4m(1-\alpha)$ and by implicit function theorem we get

$$\frac{dz^*}{dm}=\frac{1+(3-4\alpha)F(z^*)}{-f(z^*)(\mu(z^*)+a+3m-4m\alpha)+(1-F(z^*))^2},$$

which by (3) is negative. The proof is complete. □

Proof.[of Propositions 2, 3, 5] The proofs can be conducted analogously to the proof of Proposition 1 using the standard calculus. □

Proof.[of Proposition 4] Let $\Pi_d=\frac{1}{1-k}\Pi^*$. Then $\frac{d\Pi_d}{dz^*}=\frac{1-F(z^*)}{8b}(3\mu(z^*)+3a+5bc)-c$. Moreover, $\frac{d\Pi_d}{dz^*}|_A=\frac{3}{8b}(A+a-bc)>0$ and $\frac{d\Pi_d}{dz^*}|_B=-c<0$, which implies that there exists z_d such that $\frac{d\Pi_d}{dz_d}=0$. A sufficient condition for uniqueness is that the second derivative of $\Pi_d(z)$ should be negative, which is true since by (4)

$$\frac{d^2\Pi_d}{dz^2} = -\frac{1-F(z)}{8b}\left\{\frac{f(z)}{1-F(z)}(3\mu(z)+3a+5bc)-3(1-F(z))\right\} < 0.$$

We state that Π_d depends on α only by z thus it is increasing-decreasing in α. The same conclusions hold for Π^*. The proof is complete.

□

REFERENCES

BIENIEK, M.,(2017). Consignment Contracting with Inventory Control with Additive Price-Dependent Demand. In: Yu. G. Evtushenko, M. Yu. Khachay, O. V. Khamisov, Yu. A. Kochetov, V. U. Malkova, M. A. Posypkin (eds.): Proceedings of the OPTIMA-2017 Conference, Petrovac, Montenegro, 2017, http://ceur-ws.org., 1987, pp. 88–94.

HU, W., CHEN, D., YU, H., (2017). Benefit and risk analysis of consignment contracts. Annals of Operations Research, 257, pp. 641–659.

HU, W., LI, Y., GOVINDAN, K., (2014). The impact of consumer returns policies on consignment contracts with inventory control. European Journal of Operational Research, 233, pp. 398–407.

KOCABIYIKOGLU, A., POPESCU, I., (2011). An elasticity approach to the newsvendor with price-sensitivity demand. Operations Research, 59, pp. 301–312.

LEE, C., CHU, W., (2005). Who should control the inventory in a supply chain? European Journal of Operational Research, 164, pp. 158–172.

LI, S., ZHU, Z., HUANG, L., (2009). Supply chain coordination and decision making under consignment contract with revenue sharing. International Journal of Production Economics, 120, pp. 88–99.

MOSTARD, J., TEUNTER, R., (2006). The newsboy problem with resalable returns: A single model and case study. European Journal of Operational Research, 16, pp. 81–96.

OGRYCZAK, W., RUSZCZYNSKI, J., (2001). On consistency of stochastic dominance and mean-semideviation models. Mathematical Programming, 89, pp. 217–232.

OLAH, J., LAKNER, Z., HOLLOSI, P., POPP, J., (2017). Inventory methods in order to minimize raw materials at the inventory level in the supply chain. LogForum, 13, pp. 439–454.

PETRUZZI, N. C., DADA, M., (1999). Pricing and newsvendor problem: a review with extensions. Operations Research, 47, pp. 183–194.

RU, J., WANG, Y., (2010). Consignment contracting: Who should control inventory in the supply chain? European Journal of Operational Research, 201, pp. 760–769.

RUBIO-HERRERO, J., BAYKAL-GURSOY, M., JASKIEWICZ, A., (2015). A price setting newsvendor problem under mean-variance criteria. European Journal of Operational Research, 247, pp. 575–587.

WANG, Y., JIANG, L., SHEN, Z.-J., (2004). Channel performance under consignment contract with revenue sharing. Management Science, 50, pp. 34–47.

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 563

Uniwersytet
Kardynała Stefana Wyszyńskiego
w Warszawie

An international scientific conference

Community Cohesion and Well-Being, and Innovative Development
will be held on December 4, 2018, in Warsaw

The conference is organized by Statistics Poland and Cardinal Stefan Wyszyński University in Warsaw (CSWU)

The conference is devoted to the broadly conceived issue of well-being of local communities and their members (inhabitants), with the intention of deepening the knowledge about the relationships in which it remains with the level and nature of local development. While the exploration of these dependencies was the subject of a similar conference, organized two years ago jointly by Statistics Poland and CSWU, Quality of life and spatial cohesion. Interactions of development and welfare in the local context, this event can be considered as its extension, in two interrelated dimensions – cohesion and innovation – treated at the same time as factors of well-being, both at the individual and community level.

Venue: Cardinal Stefan Wyszyński University in Warsaw (CSWU)

Call for Papers: http://sdsl2018.stat.gov.pl/en/call-for-papers-en

Registration form: http://sdsl2018.stat.gov.pl/en/registration

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 565–567

ABOUT THE AUTHORS

Al-Nasser Amjad D. is a Professor at the Department of Statistics in Yarmouk University Irbid, Jordan. His research interests are Generalized Maximum Entropy, Structural Equation Modelling, acceptance sampling and quality control. Professor Al-Nasser has published more than 125 research papers in international/national journals and conferences. He has also published six books/monographs. Professor Al-Nasser is the founding editor of the Journal of Applied Statistical Analysis.

Bieniek Milena is an Associate Professor at the Statistics and Econometric Department, Economic Faculty, Maria Curie-Sklodowska University in Lublin. She has PhD in mathematics, but lately her main interests include the operations research, mainly the inventory management and the facility location problem. Her earlier interests included statistics and probability, more precisely in the partition of the unit segment and the theory of the order statistical data. She is a reviewer in Mathematical Reviews of American Mathematical Society and several top journals in the operations research field including European Journal of Operational Research, Omega or International Journal of Production Economics.

Danish Faizan is a Lecturer in Statistics at the Department of Mathematics and Statistics, Government Degree College, Shopian, J&K India. He has pursued PhD in Statistics at the Division of Statistics and Computer Science, Faculty of Basic Sciences in Sher-e-Kashmir University of Agricultural sciences and Technology Jammu, J&K India. He has done his MSc from Department of Statistics, University of Kashmir Srinagar, J&K India. His research interests are sampling theory, mathematical programming and applied statistics. He has published research articles in reputed journals of statistics and mathematics. He has also published a book and several chapters in different areas of statistics. He has presented a number of papers in several national as well as international conferences.

Dwivedi Laxmi Kant , PhD, is an Assistant Professor in the Department of Mathematical Demography & Statistics at the International Institute for Population Sciences (IIPS), Mumbai, India. His main areas of teaching and research are applied multivariate analysis, survey research, econometrics, indirect techniques for demographic estimation, simulation and decomposition techniques and impact evaluation in the field of demography/population studies. He has contributed in designing large-scale surveys such as fourth and fifth rounds of National Family Health Survey and several intervention studies. Dr Dwivedi has published over 50 research papers. He is the recipient of Chandrasekaran Young Scientist and K B Pathak Memorial Awards of the Indian Association for the Study of Population (IASP), India for contributing the best research papers in the field of demography/population studies.

Grover Gurprit is a Professor of Bio-Statistics in the Department of Statistics, Faculty of Mathematical Sciences, University of Delhi. Her other areas of interest include demography, reliability theory and statistical quality control. She has published many research papers in various national and international journals of repute. She has also co-authored three books at advanced levels from Lap Lambert publishing house, Germany. She has guided 14 PhDs and 23 MPhils until now and at present many students are working under her guidance.

Jabłońska Katarzyna holds a Master's degree in Mathematics from Jagiellonian University in Krakow (2017). She has been working at Jagiellonian University Medical College, Chair of Epidemiology and Preventive Medicine. Her areas of interest include statistical modelling, structural equation modelling, meta-analysis methods and applications of mathematics in biology and medicine.

Krzyśko Mirosław is a Full Professor and Professor Emeritus at the Department of Probability and Mathematical Statistics in Adam Mickiewicz University, Poznań, Poland. His research interests are multivariate statistical analysis, analysis of multivariate functional data, statistical inference and data analysis in particular. Professor Krzyśko has published more than 150 research papers in international/national journals and conferences. He has also published five books/monographs. Professor Krzyśko is an active member of many scientific professional bodies.

Łukaszonek Wojciech is working as an assistant at President Wojciechowski Higher Vocational State School in Kalisz, Poland. In 2002 he received his Master's Degree in Mathematics from Technical University of Łódź. Currently he is working on PhD at Poznań University of Economics and Business in the field of analysis of higher educational market.

Majdzińska Anna is a PhD research fellow at the Unit of Demography and Social Gerontology, University of Lodz. Her main research areas include ageing process, regional demography and multivariable analysis. Her academic interests also lie in methods of demographic regionalisation.

Mustafa Ahmed Bani is an Associate Professor in applied statistics at the Department of Mathematics of the School of Engineering at the Australian College of Kuwait. His research interests are broad, more recently covering methodological developments in applied statistics, but for the most mainly in applications of statistics. His work in statistical methodology is focused primarily on sampling techniques and model selection techniques, modelling checking, recursive estimation and residuals, and their applications. More generally, his interest is in applied statistics, and collaborations with researchers from different fields including medicine, economics, finance, accounting, education, psychology, sports science. Applications include regression models such as: simple and multiple linear regression Generalized Linear Models (GLM), Linear Mixed Models (LMM), GLMM, GAM, GAMM, time series, multivariate analysis, and Structural Equation Modelling.

Sabharwal Alka is an Associate Professor at the Department of Statistics, Kirori Mal College, University of Delhi, India. She is an established statistician, with specialization in biostatistics. Her research interests are stochastic processes, biostatistics and data analysis using R. Associated with a number of national and international journals, she is a reviewer for research papers from similar fields.

Yaya OlaOluwa Simon is an Assistant Professor at the Department of Statistics, University of Ibadan, Nigeria. His research interests are economic, financial and computational time series with papers written in frictional integration, unit roots, volatility modelling and regime switching. He has over 60 research papers to his credit, published in reputable journals. He is an active member of professional bodies such as the Royal Statistical Society, Nigerian Statistical Society.

Wołyński Waldemar is an Assistant Professor at the Faculty of Mathematics and Computer Science in Adam Mickiewicz University in Poznań, Poland. His major research interests focus on various aspects of multivariate statistical analysis. He is an author or co-author of over 50 research papers.

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 italics) 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).