Cracow University of Economics

ARGUMENTA OECONOMICA CRACOVIENSIA

No 1(22)

2020

Cracow University of Economics

ARGUMENTA OECONOMICA CRACOVIENSIA

No 1 (22)

2020

ISSN 1642-168X
e-ISSN 2545-3866

Ministry of Science and Higher Education
Republic of Poland

Work related to the publication of Argumenta Oeconomica Cracoviensia as it applies to language editing is financed by the Polish Ministry of Science and Higher Education under contract No. 913/P-DUN/2019 from funds earmarked for the general promotion of science

ISSN 1642-168X
e-ISSN 2545-3866

Original version: printed publication
All the articles published in Argumenta Oeconomica Cracoviensia are available on the journal's website (www.aoc.uek.krakow.pl) and in the CEEOL database (http://www.ceeol.com)

Cracow University of Economics Press
Rakowicka 27, 31-510 Kraków, Poland
www.uek.krakow.pl

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 1(22) • 2020
ISSN 1642-168X
e-ISSN 2545-3866

CONTENTS

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 1(22) • 2020
ISSN 1642-168X
e-ISSN 2545-3866

Dear Readers,

We hereby commend the latest issue of *Argumenta Oeconomica Cracoviensia* to you. The current issue reflects the profile of the journal, which publishes original contributions in the field of economics and finance. This approach is a response to the global trend, especially pronounced since the 2008–09 financial crisis, towards the search for, and explanation of, the relationships and dependencies between phenomena and processes occurring in the financial sphere and in the real economy. The social context of these efforts is expressed through research into the causes and effects of income and wealth inequalities, in various aspects, and with different impacts on society and economic growth. Other important issues that have been tackled in *Argumenta Oeconomica Cracoviensia* include the effects of globalisation, the challenges of an ageing population, the economic impact of climate change, the digitalisation of economic and social life, and the associated opportunities for economic development and social progress. One can only hope that the crisis phenomena triggered by the COVID-19 pandemic that we are witnessing today will provide the inspiration for new, multifaceted research. We invite contributors to present the results of that research in this forum.

Among the contributions in the current issue are papers on economic theory, papers on methodological issues of key importance to the study of economic and financial phenomena, and papers presenting the results of empirical research. As such the journal's formula provides a platform for the publication of papers that address the achievements of economic and financial theory as well as proposals for its practical application. Economics is, after all, an applied science.

One example of a methodological work whose findings may be applied to an analysis of real phenomena is Czesław Domański and Robert Kubacki's paper entitled "Application of the Differential Evolution Algorithm to Group a Bank's Individual Clients". The approach presented by the authors differs from methods used to date and hence may stimulate discussion among academics and inspire further research.

Małgorzata Złotoś's article, "On the Use of Permutation Tests in the Significance Testing of Response Surface Function Parameters", is another work with a clear methodological profile. In economics, unlike in the natural or technical sciences, the possibilities of using experiments on a wider scale are limited. Unfortunately, macroeconomic policy is generally conducted by trial and error. Certain opportunities for experimentation arise at the microeconomic level – in the enterprise. These possibilities are discussed in the article, which considers the use of a permutation test that allows assessment of the significance of response surface function parameters when the quantity of experimental data is small. The article is noteworthy for the fact that the author presents not only the course and results of the experiment itself, but also the significance of the preparatory phase for the results achieved.

The economy is a system of dependencies. The accuracy of diagnoses depends on the degree to which these dependencies are recognised, especially in terms of cause and effect, and this in turn allows us to better predict, and possibly shape, economic and social processes. The formulation of hypotheses and their verification plays an important role in the analysis and identification of the dependencies that exist between phenomena. These issues are the subject of Grzegorz Kończak's contribution entitled "Applications of Permutation Methods in the Analysis of Associations". The author presents selected possibilities of applying permutation methods in testing hypotheses concerning the relationships between variables. The virtue of this article lies in the author's proposed method to test the significance of the association of two sets of variables. Moreover, the author also compares the size and power of the proposed test with tests known from canonical correlation analysis. This comparison leads the author to interesting conclusions that should be of interest to specialists in the use of quantitative methods in economics and finance.

In Monika Banaszewska's work, "Condemned to Success (Failure)? The Allocation of EU Funds among Polish Municipalities", the author successfully attempts to quantify the effects of EU fund allocation in Poland at the level of municipalities. The basic criterion for allocating EU funds at the local level is to mitigate differences in living standards across local communities as part of the European Union's cohesion policy. The basic tool of this policy is per capita expenditure. The results presented by the author, achieved using the rank-rank regression method, paint an interesting picture of the movement of Polish municipalities within the ranking of per capita EU fund expenditures over two election cycles in Poland. The author rightly points out that the results need to be treated with caution, however, as it was not possible to identify, within the remit of the article, the causes of this movement within the ranking. To identify these reasons would have involved taking into account changes in the level of social capital and the impact of organisational structures (local authorities) on the acquisition EU funds, which was beyond the scope of the article. The advantage of this work is likewise the fact that the results obtained using the selected method can be compared with the results obtained by other researchers, to which the author refers. The findings may inspire further in-depth research by the author herself, as well as encourage other scholars to undertake new qualitative analyses.

The article by Czesława Pilarska and Grzegorz Wałęga, "Born Globals in Poland: Development Factors and Research Overview", confirms the presence in our journal of issues related to the globalisation of the world economy. In this paper, the authors characterise the relatively new phenomenon of born globals. This concerns the creation of new (young) companies that have quickly entered international markets relatively recently. Globalisation, the elimination of barriers to free trade, and the digitalisation of the economy create opportunities for such companies to expand. The authors present the theoretical foundations for the development of born globals and then verify these theoretical assumptions on the example of Poland. The research shows that Polish companies can achieve a competitive advantage thanks to their well-educated managerial staff, but even greater expansion of born globals is limited, among others, by a lack of sufficient capital resources.

As mentioned earlier, one of the theoretical and practical challenges associated with an ageing population is to provide security in retirement. Theoreticians and practitioners alike have struggled with this problem, proposing, and sometimes implementing, specific pension systems. Observing the current state of knowledge on this subject and evaluating the various systems in operation around the world leads to the conclusion that it is very difficult to build an optimal pension system. For this reason, the article by Anna Magdalena Gierusz, "A Comparison of Risk-Sharing Approaches in Hybrid Occupational Pension Schemes", deserves attention. The main problem that requires resolution is that it is much easier to define the amount of contributions paid in to a pension system than to define the amount of retirement benefits paid out. These difficulties arise, in particular, from the fact that the period when contributions are paid begins and ends over several decades of an employee's working life. During this period there are a number of factors influencing the pension amount that the employee will receive. The most important factor is the performance of the invested contributions (savings) on which the pension amount depends, and this in turn is closely linked to the changing situation on financial markets. That is why there is an ongoing search for hybrid systems (models) in which the risk is shared between employer and employee. In this paper, the author presents risk-sharing in selected types of hybrid schemes – cash balance and self-annuitising schemes – using a formalised approach. The novelty of the work is the original type of hybrid scheme proposed by the author, which should inspire critical assessment of this concept and encourage further research.

While commending the present issue to our readers, we would also like to invite contributions in the form of original texts, information about important academic events, and reviews of outstanding books. Texts in the field of economics and finance will be treated as most relevant to the journal's profile.

Prof. Stanisław Owsiak
Editor-in-chief

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 1(22) • 2020
ISSN 1642-168X
e-ISSN 2545-3866
AOC, 2020; 1(22): 9–20
https://doi.org/10.15678/AOC.2020.2201

Czesław Domański
Robert Kubacki

APPLICATION OF THE DIFFERENTIAL EVOLUTION ALGORITHM TO GROUP A BANK'S INDIVIDUAL CLIENTS

Abstract

Objective: The aim of the article is to present the results of grouping individual clients of a bank with the differential evolution algorithm.
Research Design & Methods: The research offers conclusions based on analysis of the bank's customer base and deductive and inductive reasoning.
Findings: The results of the authors' research show that the differential evolution algorithm correctly groups bank customers and can be used for this purpose.
Implications/Recommendations: The differential evolution algorithm is an alternative to the commonly used k-means algorithm. The algorithm generates several competing solutions in one iteration. It enables independence from starting vectors and greater effectiveness in searching for an optimal solution. The differential evolution algorithm was itself enriched with a variable that allows the optimal number of clusters to be selected. Each iteration contained proposed solutions (chromosomes) that were evaluated by the target function built on the CS measure proposed by Chou.
Contribution: The article presents the application of the differential evolution algorithm to group a bank's clients.

Czesław Domański, University of Lodz, Faculty of Economics and Sociology, Department of Statistical Methods, POW 3/5, 90-255 Łódź, Poland, e-mail: czedoman@uni.lodz.pl, ORCID: https://orcid.org/0000-0001-6144-6231.

Robert Kubacki, University of Lodz, Faculty of Economics and Sociology, Department of Statistical Methods, POW 3/5, 90-255 Łódź, Poland, e-mail: robertkubacki@o2.pl, ORCID: https://orcid.org/0000-0003-0591-9529.

Keywords: CS measure, differential evolution algorithm, clustering methods, banking.
JEL Classification: C38, M31, G21.

1. Introduction

In today's world more and more companies are having problems with the effective management of available data. The gap between the amount of data that is generated and stored and the degree of companies' understanding is constantly widening. According to a survey conducted by IBM among representatives of the largest banks, over 40% of them have problems with an excess of information and the lack of appropriate tools for analysing it (Giridhar *et al.* 2011).

Grouping methods are effective in describing populations. Many authors have studied these methods (Everitt *et al.* 2011, Feoktistov & Janaqi 2004, Gan, Ma & Wu 2007).

Classical grouping algorithms have two major disadvantages:

1. They easily fall into local optima in multidimensional spaces that have multimodal objective functions,
2. The efficiency of searching for a solution depends very much on the start vectors.

Grouping methods are known as methods without a supervisor, while most traditional algorithms require a priori knowledge of the number of clusters, which means that this is not a method without the interference of an outsider. On the other hand, in many practical applications it is impossible to provide even an approximate number of groups for an unknown data set.

The limitations of classical grouping methods, including the k-means algorithm, led the researchers to search for new, more effective grouping methods. One of the development trajectories for grouping algorithms has been to treat them as an optimization problem. Over time, the paradigm of evolutionary computation – the relationship between optimization and biological evolution – has evolved. Evolutionary calculations use the power of natural selection and allow the computing power of computers to be used for automatic optimization (Das, Abraham & Konar 2009).

The first part of the article presents selected issues related to the differential evolution algorithm. Next, the assumptions of the study and the steps taken to prepare the data are described. Selected results for the study are then outlined. In the final part of the article, the authors indicate possible directions for further research in this area.

2. Differential Evolution Algorithm – Selected Issues

The differential evolution algorithm is part of heuristic methods because the goal of optimization is not to find the exact equation describing the studied phenomenon, but to search the available space for solutions. These solutions are constructed using random elements. What is more, in one iteration of the algorithm several competing solutions are created. Subsequent solutions are created using similarities to the evolutionary mechanisms occurring in nature. These are the ones that, according to the defined objective function, are the best. The characteristic feature of the differential evolution algorithm is that solutions are created on the basis of real variable vectors, not vectors coded to zero-one sequences (Das, Mullick & Suganthan 2016).

Since 1995 the differential evolution algorithm (Storn 1995, Storn & Price 1997) has drawn the attention of optimization practitioners due to the degree of resistance, the speed of convergence and the accuracy of solutions for real optimization problems. The differential evolution algorithm has defeated many algorithms, such as genetic algorithms, evolutionary strategies and memetic algorithms (Das, Mullick & Suganthan 2016).

Suppose we have a set of objects *Np* vectors, each of which has *D* dimensions. In addition, we mark $\boldsymbol{P}_{\boldsymbol{X}}$ as the current population of solutions to the optimization problem, which was created as an initial solution or at any subsequent stage of the algorithm's operation (Das, Abraham & Konar 2009).

$$\boldsymbol{P}_{\boldsymbol{X},g} = (\boldsymbol{X}_{i,g}), \quad i = 0, 1, \ldots, Np - 1, \quad g = 0, 1, \ldots, g_{\max}, \tag{1}$$

$$\boldsymbol{X}_{i,g} = (x_{j,i,g}), \quad j = 0, 1, \ldots, D - 1. \tag{2}$$

Index $g = 0, 1, \ldots, g_{\max}$ denotes the generation to which the vector belongs. Each vector is assigned to the corresponding population index $i = 0, 1, \ldots, Np - 1$. The dimensions of the vector are marked by $j = 0, 1, \ldots, D - 1$.

The differential evolution algorithm generates mutant vectors in the next step, which will be marked as follows (Das, Abraham & Konar 2009):

$$\boldsymbol{P}_{V,g} = (\boldsymbol{V}_{i,g}), \quad i = 0, 1, \ldots, Np - 1, \quad g = 0, 1, \ldots, g_{\max}, \tag{3}$$

$$\boldsymbol{V}_{i,g} = (v_{j,i,g}), \quad j = 0, 1, \ldots, D - 1. \tag{4}$$

However, the vectors after crossover will be marked as follows:

$$\boldsymbol{P}_{U,g} = (\boldsymbol{U}_{i,g}), \quad i = 0, 1, \ldots, Np - 1, \quad g = 0, 1, \ldots, g_{\max}, \tag{5}$$

$$U_{i,g} = (u_{j,i,g}), \quad j = 0, 1, \ldots, D-1. \tag{6}$$

The first stage, i.e. setting the initial vectors, consists in generating starting vectors. The initial parameters (for $g = 0$) are set within limits that correspond to a range that is acceptable for the intended solution. Therefore, if j-th the search task parameter has ranges marked as $x_{\min, j}$ and $x_{\max, j}$ and $rand_{i,j}(0, 1)$ means j-th realizations of a uniform distribution from the range from 0 to 1 for i-th vector, then j-th component i-th population element can be determined as (Das, Abraham & Konar 2009):

$$x_{i,j}(0) = x_{\min, j} + rand_{i,j}(0,1) \cdot (x_{\max, j} - x_{\min, j}). \tag{7}$$

The differential evolution algorithm searches for the global optimum in D-dimensional continuous hyperspace. It starts with a randomly selected population Np D-dimensional values of parameter vectors. Each vector, also known as a genome/chromosome, is a proposed solution in a multidimensional optimization issue. The next generations of solutions in the differential evolution are marked as $g = 0, 1, 2, \ldots, g, g + 1$.

The vector parameters may change with the appearance of new generations, therefore the notation for which it will be accepted, for which i-th population vector for the current generation over time ($g = g$) as (Das, Abraham & Konar 2009):

$$\vec{X}_i(g) = [x_{1,1}(g), x_{i,2}(g), \ldots, x_{i,D}(g)]^T, \tag{8}$$

where $i = 1, 2, \ldots, Np$.

Mutation means a sudden change in the characteristics of the chromosome gene. In the context of evolutionary computation, a mutation means a change or disorder of a random component. Most evolutionary algorithms simulate the effect of mutations through the additivity of the component generated with a given probability distribution. In the differential evolution algorithm, a uniform distribution of the vector of the form differences was used (Das, Abraham & Konar 2009):

$$\Delta \vec{X}_{r2,r3} = (\vec{X}_{r2} - \vec{X}_{r3}). \tag{9}$$

In the differential evolution algorithm, the mutation creates a successor vector $\vec{V}_i(g)$ for changing the population element $\vec{X}_i(g)$ in every generation or iteration of the algorithm.

To create a vector $\vec{V}_i(t)$ for each i-th element of the current population, the other three disjoint vectors $\vec{X}_{r_1^i}(g), \vec{X}_{r_2^i}(g), \vec{X}_{r_3^i}(g)$ are randomly selected from the current population. Indexes r_1^i, r_2^i, r_3^i are mutually exclusive integers selected from a range [1, *Np*], which are also different from the index and the base vector. Indexes are generated randomly for each mutated vector. Then, the difference of any two of the three vectors is scaled by the number *F* and added to the third vector. In this way, we get a vector $\vec{V}_i(g)$ expressed as (Das, Abraham & Konar 2009):

$$\vec{V}_i(g) = \vec{X}_{r_1^i}(g) + F \cdot (\vec{X}_{r_2^i}(g) - \vec{X}_{r_3^i}(g)). \tag{10}$$

The mutation scheme shows different ways of differentiating the proposed solutions.

The crossover operation is used to increase the diversity of the population of solutions. Crossing takes place after generating a donor vector through a mutation. The algorithms of the differential evolution family use two intersection schemes – exponential and binomial (zero-one). The donor vector lists the components with the target vector $\vec{X}_i(g)$ to create a trial vector:

$$\vec{U}_i(g) = [u_{1,1}(g), u_{i,2}(g), \ldots, u_{i,D}(g)]^T. \tag{11}$$

In exponential crossover, we first select a random integer *n* from range [0, *D* – 1]. The drawn number is the starting point for the target vector from which the components are crossed with the donor vector. An integer *L* is also selected from range [1, *D*]. *L* indicates the number of components in which the donor vector is involved. After selection *n* and *L*, the trial vector takes the form (Das, Abraham & Konar 2009):

$$u_{i,j}(g) = \begin{cases} v_{i,j}(g) & \text{for} \quad j = \langle n \rangle_D, \langle n+1 \rangle_D, \ldots, \langle n+L-1 \rangle_D \\ x_{i,j}(g), & \text{for other } j \in [0, D-1] \end{cases} \tag{12}$$

where the intervals denote the module modulo function *D*. Integer *L* is drawn from the sequence [1, 2, …, *D*] according to the following pseudocode:

L = 0;
Do
{
L = *L* + 1;
} *while (rand*(0, 1) < *CR*) *AND* (*L* < *D*));

As a result, the probability $(L \geq v) = (CR)^{v-1}$ for any $v > 0$. The crossover rate (*CR*) is a parameter the same as *F*. For each donor vector, a new set *n* and *L* must be drawn as described above.

On the other hand, binomial crossover is carried out for each *D* variable each time, when the number selected from 0 to 1 is less than or equal to the value *CR*. In this case, the number of parameters inherited from the donor has a very similar distribution to the binomial one. This scheme can be represented in the following way (Das, Abraham & Konar 2009):

$$u_{i,j,g} = \begin{cases} v_{i,j,g}, \textit{if } (rand_{i,j}(0,1) \leq CR \ \text{ or } \ j = j_{rand}) \\ x_{i,j,g} \ \text{otherwise} \end{cases} \quad (13)$$

where $rand_{i,j}(0,1) \in [0,1]$ is a randomly drawn number that is generated for every *j*-th of the *i*-th parameter of the vector. $j_{rand} \in [1, 2, \ldots, D]$ is a randomly selected index that ensures that $\vec{U}_{i,g}$ contains at least one component from the vector $\vec{V}_{i,g}$.

This is determined once for each vector in a given generation. *CR* is an estimate of true probability p_{CR} the event that the component of the sample vector will be inherited from the parent. It may also happen that in the two--dimensional search space, three possible test vectors can be the result of one-dimensional mating of the mutant/donor vector $\vec{V}_i(g)$ with the target vector $\vec{X}_i(g)$. Trial vectors:

a) $\vec{U}_i(g) = \vec{V}_i(g)$ both components $\vec{U}_i(g)$ inherited from the vector $\vec{V}_i(g)$,

b) $\vec{U}_i'(g) = \vec{V}_i(g)$ one component (*j* = 1) comes from vector $\vec{V}_i(g)$, the second (*j* = 2) from vector $X_i(t)$,

c) $\vec{U}_i''(g) = \vec{V}_i(g)$ one component (*j* = 1) comes from vector $X_i(g)$, the second (*j* = 2) from vector $\vec{V}_i(g)$.

The last stage of the differential evolution algorithm is selection, i.e. the choice between the vector $\vec{X}_i(g)$ and a newly designated test vector $\vec{U}_i(g)$. The decision which of the two vectors will survive in the next generation *g* + 1 depends on the value of the matching function. If the values of the matching function for the sample vector are better than the value of the target vector, the existing vector is replaced with the new vector (Das, Abraham & Konar 2009).

$$\vec{X}_i(g+1) = \begin{cases} \vec{U}_i(g) \ \text{ for } \ f(\vec{U}_i(g)) \leq f(\vec{X}_i(g)) \\ \vec{X}_i(g) \ \text{ for } \ f(\vec{U}_i(g)) > f(\vec{X}_i(g)) \end{cases} \quad (14)$$

where $f(\vec{X})$ is a minimized function. The selection process consists in selecting one of two variants. The adjustment of population members improves in subsequent generations or remains unchanged, but never deteriorates.

The *CS* (Candidate Solution) measure proposed by Chou (Chou, Su & Lai 2004) is an objective function in this study. Group centroids are determined as the average vectors belonging to a given cluster

$$\overline{m}_i = \frac{1}{N_i} \sum_{\overline{Z}_j \in C_i} \overline{Z}_j. \tag{15}$$

The distance between two points $\overline{Z}_p$ and $\overline{Z}_y$ is marked as $d(\overline{Z}_p, \overline{Z}_y)$. Then the *CS* measure can be defined as:

$$CS(k) = \frac{\frac{1}{k}\sum_{i=1}^{k}\left[\frac{1}{|C_i|}\sum_{\overline{Z}_y \in C_i} \max_{\overline{Z}_y \in C_i}\{d(\overline{Z}_p, \overline{Z}_y)\}\right]}{\frac{1}{k}\sum_{i=1}^{k}\left[\min_{j \in k, j \neq i} d(\overline{m}_i, \overline{m}_j)\right]} = \frac{\sum_{i=1}^{k}\left[\frac{1}{|C_i|}\sum_{\overline{Z}_y \in C_i} \max_{\overline{Z}_y \in C_i}\{d(\overline{Z}_p, \overline{Z}_y)\}\right]}{\sum_{i=1}^{k}\left[\min_{j \in k, j \neq i} d(\overline{m}_i, \overline{m}_j)\right]}. \tag{16}$$

The measure is a function of the ratio of the amount of intra-group dispersion and the separation between groups. The *CS* measure is more effective at clusters with different density and/or different sizes than other measures.

3. Design of the Study

A database of a commercial bank's clients from Europe was used for the study. It was limited to that part of the population for which the actions taken will translate in the maximum way into business benefits. In particular, clients meet the following criteria: individual clients with active products as on 1 March 2017, aged between 18 and 75 years, who are not bank employees, with positive marketing consent, and without delays in the repayment of loan products[1].

As for the variables used for the study, the choice was not accidental. The variables selected for the study can be evaluated for each customer regardless of whether they have deposit, credit or investment products. By pre-processing data it was possible to eliminate outliers from the studied population. Due to the strong right-side skewness of the variables, a transformation was made by adding the constant 0.001, and then their

[1] The authors are not permitted to disclose the exact name of the bank which supported the data for this study.

logarithmisation. The resulting distributions of variables are thus more symmetrical.

The final set of variables used in the study is presented below:

– ZM1 (DEPOZYTY) – total funds on accounts and deposits in thousand PLN,

– ZM2 (INWESTYCJE) – total funds in investment products in thousand PLN,

– ZM3 (LUDNOSC) – number of inhabitants, based on the city from the correspondence address and data published by Statistics Poland,

– ZM4 (KREDYTY) – amount of bank loans taken in thousand PLN,

– ZM5 (SALDO_BIK) – balance for repayment on credit products outside the bank, based on inquiries from BIK in thousand PLN,

– ZM6 (AVG_TRN_INCOMING_ALL_3M) – average monthly income on customer accounts in the last 3 months in thousand PLN,

– ZM7 (AVG_TRN_INCOMING_CLEAN_3M) – cleaned average monthly income on customer accounts in the last 3 months in thousand PLN,

– ZM8 (AVG_TRN_OUTGOING_ALL_3M) – average monthly outflows from customer accounts in the last 3 months in thousand PLN,

– ZM9 (AVG_TRN_OUTGOING_CLEAN_3M) – cleaned monthly average outflows from customer accounts in the last 3 months in thousand PLN,

– ZM10 (AVG_TRN_OUT_DEBIT_3M) – average monthly transaction amount on the debit card from the last 3 months in thousand PLN,

– ZM11 (AVG_TRN_OUT_CREDIT_3M) – average monthly amount of credit card transactions from the last 3 months in thousand PLN,

– ZM12 (WIEK_LATA) – customer age in years,

– ZM13 (STAZ_LATA) – customer experience in years.

Table 1 shows the constants used in the algorithm.

Table 1. Constants Used in the Study

Constant	Value	Description of Constant
LZ	13	number of variables describing the client
LC	13	number of chromosomes
LK	15	maximum number of clusters
SA	0.2	constant activation of the vector
F	0.7	mutation operator
Iterations	15	number of iterations
CR	1	crossover rate

Source: authors' own calculations.

For the purpose of optimizing the number of centroids, a dimensional matrix $MR_{c,k,z}$ is created, where c means the number of chromosomes, k means the number of clusters, and z means the number of variables. The number of variables is increased by 1. An additional variable is used to store information on whether the cluster is active or inactive in the given iteration (Das, Abraham & Konar 2008). The values for individual matrix elements are generated according to formula (7). An additional variable indicating focus activation is determined on the basis of the following rule: if the randomly generated number from the range 0 to 1 is smaller than the activation constant (SA), then the variable takes the value 0, otherwise it takes the value 1.

4. Results of Empirical Analyses

The smallest value of the *CS* function in the fifteenth iteration was obtained for chromosome number 3. This solution was chosen as the optimal solution.

Table 2 presents the characteristics of chromosome 3, which divided the surveyed population of the bank's clients into 9 groups (the maximum number of groups into which the population could be divided was 15).

Table 2. Numbers and Share of Groups for Chromosome 3

Group	Number of Clients	% of Total
8	92,109	45.71
4	44,545	22.11
6	29,047	14.41
3	20,003	9.93
5	5,476	2.72
14	3,582	1.78
1	2,839	1.41
15	2,075	1.03
12	1,832	0.91
Sum	201,508	100

Source: authors' own calculations.

The results of the grouping in Table 2 indicate that the distinguished groups are characterized by nonequal distribution of the number of clients in groups. Group 8 is more selective and includes 45.71% of clients, group 4

Table 3. Average Values of Variables ZM1–ZM13 for Clusters Obtained by the Differential Evolution Algorithm

Cluster	ZM1	ZM2	ZM3	ZM4	ZM5	ZM6	ZM7	ZM8	ZM9	ZM10	ZM11	ZM12	ZM13
8	8	0	452	137	88	5	4	5	3	0	0	43	5
4	2	5	273	5	1	5	4	5	4	0	0	43	6
6	20	10	500	182	146	22	17	22	17	1	0	42	6
3	28	4	453	7	45	1	1	1	0	0	0	47	6
5	60	61	627	325	0	26	20	27	19	0	1	41	6
14	113	97	758	414	144	113	84	106	73	2	1	42	6
1	20	67	473	223	114	4	3	4	2	0	0	43	6
15	24	48	321	264	10	10	8	7	5	0	0	41	5
12	0	2	131	7	117	15	11	18	15	0	0	41	3

Source: authors' own calculations.

contains 22.11% of clients, and group 6, the third group – 14.41% of clients. The three mentioned groups account for over 80% of the surveyed population.

More detailed characteristics of the distinguished groups of clients are presented in Table 3, which contains the average values of features in individual groups. The data presented in Table 3 indicate that individual groups differ from each other. Thanks to knowing the average values for particular groups, it is possible to indicate groups of transactionally active customers (groups 14, 5 and 6) and customers who use accounts less frequently (groups 3, 8, 4 and 1). The most affluent group of customers with very high means is without a doubt group 14.

Thanks to the use of the differential evolution algorithm to group the bank's clients, we can, in a relatively short period of time, get information on how many natural groups exist. Moreover, the number of groups is calculated by the algorithm, not imposed in advance. The algorithm evaluated and compared the obtained results for other candidate solutions in subsequent iterations, recognizing, according to the values of the objective function, that the optimal division of this group of customers contains 9 clusters.

5. Conclusions

The differential evolution algorithm is a promising approach to optimization because it generates a whole set of solutions that can be easily adapted to carry out the optimization again. The fact that a set of solutions is retained, and not just the best solution, allows faster adaptation to new conditions using the previously made calculations. It is resistant in terms of the choice of parameters as well as the regularity in which it finds the global optimum. The algorithm is a direct search solution method, versatile enough to solve problems whose objective function lacks the analytical description needed to determine the gradient. The algorithm is also very simple to use and modify.

Evolutionary algorithms, in particular the differential evolution algorithm, do well with continuous variables when grouping clients. Customers from particular groups can be synthetically described by the mean vector for variables used in clustering. Customers with the same basket of products, but differing in the level of individual variables, can be effectively separated.

The results of the study show that the differential evolution algorithm can be successfully applied to group retail banks' clients. Further research might be conducted with an extended list of variables i.e. those with a wider window of observation (maximum balance in a deposit product in the last 6 months, maximum balance in an investment product in the last 6 months). It would also be advisable to analyse exclusively deposit clients or credit clients with specific variables calculated and populated for those groups.

Bibliography

Chou, C. H., Su, M. C. and Lai, E. (2004) “A New Cluster Validity Measure and Its Application to Image Compression”. *Pattern Analysis and Applications* 7 (2): 205–20, https://doi.org/10.1007/s10044-004-0218-1.

Das, S., Abraham, A. and Konar, A. (2008) “Automatic Clustering Using an Improved Differential Evolution Algorithm”. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 38 (1): 218–37, https://doi.org/10.1109/TSMCA.2007.909595.

Das, S., Abraham, A. and Konar, A. (2009) *Metaheuristic Clustering*, vol. 178. Springer.

Das, S., Mullick, S. S. and Suganthan, P. N. (2016) “Recent Advances in Differential Evolution – an Updated Survey”. *Swarm and Evolutionary Computation* 27: 1–30, https://doi.org/10.1016/j.swevo.2016.01.004.

Everitt, B. S., Landau, S., Leese, M. and Stahl, D. (2011) *Cluster Analysis: Wiley Series in Probability and Statistics*. Wiley.

Feoktistov, V. and Janaqi, S. (2004) “New Strategies in Differential Evolution” in *Adaptive Computing in Design and Manufacture* VI. London: Springer, pp. 335–46.

Gan, G., Ma, C. and Wu, J. (2007) *Data Clustering: Theory, Algorithms, and Applications*, vol. 20. Siam.

Giridhar, S., Notestein, D., Ramamurthy, S. and Wagle, L. (2011) “Od złożoności do orientacji na klienta”. *Executive Report, IBM Global Business Services*.

Storn, R. (1995) “Differential Evolution – a Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces”. *Technical Report* 11. International Computer Science Institute.

Storn, R. and Price, K. (1997) “Differential Evolution – a Simple and Efficient Heuristic for Global Optimization over Continuous Spaces”. *Journal of Global Optimization* 11 (4): 341–59.

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 1 (22) • 2020
ISSN 1642-168X
e-ISSN 2545-3866
AOC, 2020; 1 (22): 21–29
https://doi.org/10.15678/AOC.2020.2202

Małgorzata Złotoś

ON THE USE OF PERMUTATION TESTS IN THE SIGNIFICANCE TESTING OF RESPONSE SURFACE FUNCTION PARAMETERS

Abstract

Objective: The methods of experimental design were first used in agricultural experiments performed by R. A. Fisher. The development of experimental design methods took place along with their effective use in production companies. The most frequently used designs of experiments are the factorial designs. One of the stages in the factorial design of experiments is the estimation of the response surface function formula which describes the influence of factors on the response variable values. The aim of this article is to propose a method to indicate the factors which have a significant influence on the response variable.
Research Design & Methods: In this case, in the classical approach, the *t*-test of the significance of particular parameters of the response surface function is used. The *t*-test requires fulfilment of the assumptions about the distribution and independence of the model errors. If the assumptions are not fulfilled, or the sample size is not sufficient, the use of the *t*-test is unjustified. An alternative approach to verify the significance of response surface parameters is a permutation test. Permutation tests use simulation methods and do not entail the fulfilment of strict assumptions relating to the distribution of errors and the sample size of experimental data.
Findings: The paper deals with the use of a permutation test that allows us to assess the significance of response surface function parameters when the quantity of experimental data is small. These results were obtained using parametric tests and permutation tests.
Implications/Recommendations: Based on the performed calculations, it was found that it is possible to use permutation tests to analyse the response surface function,

Małgorzata Złotoś, University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, 1 Maja 50, 40-287 Katowice, Poland, e-mail: malgorzata.zlotos@ue.katowice.pl, ORCID: https://orcid.org/0000-0002-8680-4848.

especially when the assumptions about the residuals of the model are not fulfilled or the number of considered experimental trials is small.
Contribution: A proper analysis of the response surface function is an important stage in the design of experiments. In the case of a small quantity of experimental data, assessment of the significance of the model and the parameters of the response surface function using parametric tests may lead to incorrect conclusions. Therefore, the use of permutation tests was indicated as an alternative approach in the analysis of the response surface function.

Keywords: design of experiments, permutation tests, response surface function, model significance.
JEL Classification: C99, C12, C15.

1. Introduction

Experimental design methods are among those statistical quality control tools which are used effectively in practice. Their implementation leads to some improvement in the technological parameters of the manufacturing process, which enhances the quality of products and decreases financial losses related to the production process in question. The proper use of experimental design methods requires adequate practical knowledge about the process and knowledge of statistical methods (Kończak 2007).

An experiment is a sequence of experimental trials. An individual experimental trial is an obtainment of the response variable Y with the fixed values of factors $X_1, X_2, \ldots, X_m$. Then the design of the experiment is defined as the layout of factor levels in further experimental trials. The dependence of the response variable Y and of the values of factors is defined as the statistical model (Wawrzynek 1993):

$$Y(X_1, X_2, \ldots, X_m) = y(X_1, X_2, \ldots, X_m) + \varepsilon, \tag{1}$$

where $EY(X_1, X_2, \ldots, X_m) = y(X_1, X_2, \ldots, X_m)$, $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2$ and σ^2 is a fixed value. The model (1) can be presented as the formula of the general linear model as follows (Wawrzynek 2009):

$$Y^T = (Y_1\ Y_2\ \ldots\ Y_m), \tag{2}$$

$$\varepsilon^T = (\varepsilon_1\ \varepsilon_2\ \ldots\ \varepsilon_n), \tag{3}$$

$$\beta^T = (\beta_1\ \beta_2\ \ldots\ \beta_k), \tag{4}$$

$$f^T(x) = (f_1(x)\ f_2(x) \ldots f_k(x)), \tag{5}$$

$$F = \begin{bmatrix} f_1(x_1) & \dots & f_k(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_n) & \dots & f_k(x_n) \end{bmatrix}, \tag{6}$$

where $f_i(x_j) \equiv x_{ij}$, for $i = 1, 2, \dots, k, j = 1, 2, \dots, n$. Then the response surface function is defined as $y = F\beta$.

Usually, the response surface functions which do not include any interaction between the factors are considered. Their formula is as follows (Montgomery 2001, Wawrzynek 2009):

$$y(x_1, x_2, \dots, x_m) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m. \tag{7}$$

The response surface functions which take into account the interactions of the factors are also considered:

$$y(x_1, x_2, \dots, x_m) = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \beta_{12} x_1 x_2 + \dots + \beta_{m-1m} x_{m-1} x_m. \tag{8}$$

In the classical approach, in order to estimate the parameters of vector β of the response surface function, the least square method is used (Aczel 2000, Elandt 1964, Montgomery 1997, Wawrzynek 1993).

2. The Significance of Response Surface Function

The response surface function is a mathematical description of the dependence of factors on the response variable. In particular, the analysis of the response surface function allows us to verify the model significance and the significance of the individual variable impact on the response variable. Then, in order to use proper parametric tests, it is important to assume that the distribution of model residuals is a normal distribution with the expected value 0 and with a standard deviation σ^2 (Montgomery 2001).

In order to verify the significance of the response surface function model, the following hypotheses were formulated (Aczel 2000):

$$\begin{aligned} &H_0\colon \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ &H_1\colon \beta_j \neq 0, \ \text{for some } j. \end{aligned} \tag{9}$$

Assuming that the null-hypothesis is true, the test statistic

$$F = \frac{(\hat{\beta}' X' y - n\bar{y}^2)/k}{(y'y - \hat{\beta}' X' y)/(n - k - 1)} \tag{10}$$

has a F-Snedecor distribution with k and $n - k - 1$ degrees of freedom. The null-hypothesis should be rejected when $F > F_{\alpha, k, n-k-1}$.

The indication of factors which should be included or omitted in the considered model of response surface function is possible because of the verification of the hypotheses formulated below:

$$H_0\colon \beta_j = 0 \qquad H_1\colon \beta_j \neq 0, \tag{11}$$

where β_j is an established parameter of the response surface function. Then the value of the test statistic (Aczel 2000):

$$t = \frac{\hat{\beta}_j}{\sqrt{\sigma^2 C_{jj}}}, \tag{12}$$

is calculated, where C_{jj} is a diagonal element of matrix $(X'X)^{-1}$. If the value of the test statistic (12) satisfies the inequality $|t| > t_{\alpha/2, n-k-1}$ the null--hypothesis should be rejected. Then the variable X_j should not be included in the considered model of response surface function.

3. Permutation Tests

Permutation tests, like experimental design, were proposed by R.A. Fisher in the 1920s. However, due to computational difficulties, they did not find application as early as experimental design did. It was not until the beginning of the 21st century that permutation methods were developed (Kończak 2016).

A permutation test is described as a general method of estimating the probability of an event occurring. Permutation tests are an alternative to parametric tests which use only sample data; do not require assumptions regarding the distribution form in the population; are resistant to the occurrence of outliers; and can be used for the sample with a small number of observations (Berry, Johnston & Mielke 2014). Moreover, the power of permutation tests is comparable to the power of traditional parametric tests (O'Gorman 2012).

Good (2005) presents the procedure for permutation tests in the following stages:

1. Define the null-hypothesis and the alternative hypothesis.
2. Choose the formula of testing statistic T.

3. Determine the value of the test statistic (T_0) for the sample data.

4. Count the value of the test statistic T for a sufficiently large number (N) of permutations of a data set to obtain the set $\{T_1, T_2, \ldots, T_N\}$.

5. Determine the ASL value and make your decision.

If the alternative hypothesis is right-sided, then the value of ASL (Achieving Significance Level) is described as follows:

$$ASL = P(T_i \geq T_0). \tag{13}$$

Then the estimation of ASL value is determined on the basis of the following formula:

$$\widehat{ASL} = \frac{card\{i: T_i \geq T_0\}}{N}. \tag{14}$$

When the two-sided alternative hypothesis is considered, then the ASL value can be rewritten as follows:

$$ASL = P(|T_i| \geq |T_0|), \tag{15}$$

and the approximate value of ASL is calculated using the following formula:

$$\widehat{ASL} = \frac{card\{i: |T_i| \geq |T_0|\}}{N}. \tag{16}$$

The null-hypothesis should be rejected when the value $\widehat{ASL}$ is smaller than the assumed significance level.

The use of permutation tests to verify the significance of the model (7) or (8) and the significance of its parameters is connected with the description of the permutation rules of multidimensional data. O'Gorman (2012) gives four appropriate methods for these permutations: permutation of errors, permutation of residuals, permutation of independent variables, and permutation of the response variable. Because of the fixed values of factors (independent variables) in particular experimental trials, the permutation of the response variable is taken into account for the considered data in experimental design. In order to verify the significance of the response surface model and the significance of the parameters of the response surface function, in the permutation test procedure the test statistics (10) and (12) can be accordingly used (Kończak 2012, 2016).

4. Example

It is assumed that the brake horsepower (response variable Y) developed by an automobile engine depends on the engine speed in revolutions per minute (factor X_1), the road octane number of the fuel (factor X_2), and the engine compression (factor X_3). The experimental data for $n = 12$ experimental trials are presented in Table 1.

Table 1. The Experimental Data

No.	X_1	X_2	X_3	Y
1	2,000	90	100	225
2	1,800	94	95	212
3	2,400	88	110	229
4	1,900	91	96	222
5	1,600	86	100	219
6	2,500	96	110	278
7	3,000	94	98	246
8	3,200	90	100	237
9	2,800	88	105	233
10	3,400	86	97	224
11	1,800	90	100	223
12	2,500	89	104	230

Source: Montgomery (2001).

The response surface function which does not include interactions between factors is taken into consideration and has the following form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 . \tag{17}$$

Using the least mean square method, the values of the response surface function parameters were estimated:

$$y = -266.031 + 0.011 x_1 + 3.135 x_2 + 1.867 x_3 . \tag{18}$$

For the estimated values of the model residuals, Shapiro-Wilk's test was prepared. The obtained p-value $= 0.3706$, so the sample does not provide sufficient evidence to reject the null-hypothesis saying that the distribution of residuals is the normal distribution (Figure 1).

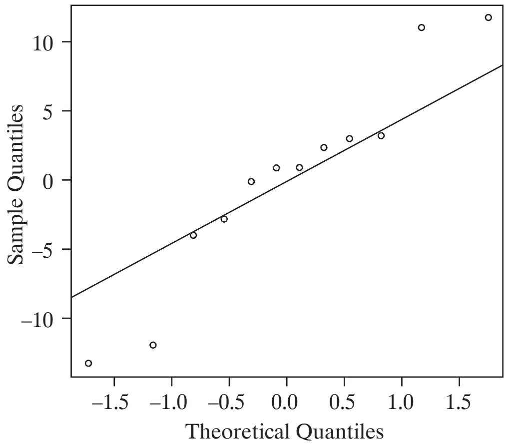


Fig. 1. Normal Q-Q Plot for Residuals of Model (18)

Source: the author's own elaboration.

The significance of the response surface model was verified and the value of statistic $F = 11.12$ was obtained. The calculated value of statistic F is bigger than the critical value $F_{0.05,3.8} = 4.07$, so the null-hypothesis should be rejected, which means that the response surface model (18) is significant. Moreover, the significance of the response surface parameters was investigated with parametric test t. The results are presented in Table 2.

Table 2. Results of Testing Significance of Response Surface Function (18)

Parameter	Estimate	Standard Error	t-value	p-value
β_0	–266.03	92.674	–2.871	0.021
β_1	0.017	0.004	2.390	0.044
β_2	3.135	0.844	3.712	0.006
β_3	1.867	0.535	3.494	0.008

Source: the author's own elaboration.

The considered experimental data include a small number of experimental trials. Therefore, because of the assumptions about the distribution of residuals, the use of the parametric test of model significance or the t test can be unfounded. Thus, in order to confirm or deny the obtained results, the proper permutation tests were carried out.

The permutation test of model significance uses the test statistic (10). For N = 1000 permutations of the response variable, formulas of response surface functions were estimated, with the appropriate values of the test statistic F. According to (14), the value $\widehat{ASL} = 0.003$ was estimated. Therefore, on the significance level $\alpha = 0.05$ the null-hypothesis should be rejected, which confirms that the response surface function model (18) is significant.

The significance of the response surface function parameters was verified with the test statistic (12). For every parameter of response surface function N = 1000 permutations of the response variable were performed and response surface functions were estimated with t-statistic values. Then, according to (16), the values of $\widehat{ASL}$ were estimated. The results are presented in Table 3.

Table 3. The Values of $\widehat{ASL}$

Parameter	$\widehat{ASL}$-value
β_0	0.078
β_1	0.030
β_2	0.005
β_3	0.008

Source: the author's own elaboration.

On the basis of the performed calculations, it can be seen that the conclusions for parametric tests and for appropriate permutation tests are similar. It should be emphasised that the permutation tests did not require fulfilment of the assumptions regarding the distribution of residuals of the considered model.

5. Conclusions

The methods of experimental design are used primarily in statistical quality control procedures. One of the fundamental stages of experimental design is to estimate the response surface function model and its analysis, which allows proper recommendations for the production process in question to be formulated. In particular, analysis of the response surface function relies on an assessment of the significance of the estimated model and its parameters, where classical parametric tests are used. The present study analyses the response surface function for an experiment involving a small number of experimental trials, which may lead to incorrect conclusions in the case of parametric tests. Then, it was proposed to use permutation tests, which do not require fulfilment of the restrictive assumptions regarding the distribution of model residuals and can be used for a data set with a small number of observations.

Bibliography

Aczel, A. (2000) *Statystyka w zarządzaniu.* Warszawa: Wydawnictwo Naukowe PWN.

Berry, K. J., Johnston, J. E. and Mielke Jr., P. W. (2014) *A Chronicle of Permutation Statistical Methods*. New York: Springer International Publishing.

Elandt, R. (1964) *Statystyka matematyczna w zastosowaniu do doświadczalnictwa rolniczego.* Warszawa: Państwowe Wydawnictwo Naukowe.

Good, P. (2005) *Permutation, Parametric and Bootstrap Tests of Hypotheses.* New York: Springer Science Business Media.

Kończak, G. (2007) *Metody statystyczne w sterowaniu jakością produkcji*, Katowice: Wydawnictwo Akademii Ekonomicznej w Katowicach.

Kończak, G. (2012) "On Testing the Significance of the Coefficients in the Multiple Regression Analysis". *Acta Universitatis Lodziensis. Folia Oeconomica* 269: 61–71.

Kończak, G. (2016) *Testy permutacyjne. Teoria i zastosowania.,* Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.

Montgomery, D. C. (1997) *Introduction to Statistical Quality Control*, New York: John Wiley & Sons.

Montgomery, D. C. (2001) *Design and Analysis of Experiments.* New York: John Wiley & Sons.

O'Gorman, T. W. (2012) *Adaptive Tests of Significance Using Permutations of Residuals with R and SAS*. New Jersey: John Wiley and Sons.

Wawrzynek, J. (1993) *Statystyczne planowanie eksperymentów w zagadnieniach regresji w warunkach małej próby*, Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.

Wawrzynek, J. (2009) *Planowanie eksperymentów zorientowane na doskonalenie jakości produktu*, Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 1(22) • 2020
ISSN 1642-168X
e-ISSN 2545-3866
AOC, 2020; 1(22): 31–45
https://doi.org/10.15678/AOC.2020.2203

Grzegorz Kończak

APPLICATIONS OF PERMUTATION METHODS IN THE ANALYSIS OF ASSOCIATIONS

Abstract

Objective: The permutation model in hypothesis testing was introduced by R. A. Fisher in 1925. These methods permit us to test hypotheses with as minimal assumptions as possible. The tests require high computing power and therefore have found greater application in recent years. However, the concept of permutation methods is much wider than the issue of permutation testing. In 1923 J. Spława-Neyman introduced a permutation model for the analysis of field experiments. The purpose of the article is to present the possibilities of applying permutation methods in the analysis of dependencies. The article presents selected possibilities of data rearranging in dependency analysis.

Research Design & Methods: The study considered the analysis of multivariate data. The paper presents theoretical considerations and refers to the Monte Carlo simulation.

Findings: A proposed method to allow investigation of the significance of the relationship between two data sets is presented. The considerations are supplemented by comparing the size and power of the proposed test with tests known from canonical correlation analysis.

Implications/Recommendations: The proposal is most powerful for non-normally distributed variables and small samples.

Contribution: The proposed test can be used in the analysis of multidimensional economic and social phenomena.

Grzegorz Kończak, University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, ul. 1 Maja 50, 40-287 Katowice, Poland, e-mail: grzegorz.konczak@ue.katowice.pl, ORCID: https://orcid.org/0000-0002-4696-8215.

Keywords: permutation methods, data labels permutation, association analysis, canonical correlations.
JEL Classification: C12, C15, C18.

1. Introduction

The permutation model in hypothesis testing was introduced by R. A. Fisher in 1925. These methods permit us to test hypotheses with as minimal assumptions as possible. The tests require high computing power and therefore have found greater application in recent years. However, the concept of permutation methods is much wider than the issue of permutation testing. In 1923, J. Spława-Neyman introduced a permutation model for the analysis of field experiments (Spława-Neyman 1923, Berry, Johnston & Mielke 2014). This paper was published in Polish. Neyman's work was translated into English in 1990 and published in *Statistical Science* (Ledwina 2012), which allowed it to reach the international statistician community (Lehmann & Romano 2005, p. 210). The paper was recognized as a pioneering achievement in the field of statistical methodology for the analysis of causal relationships.

The purpose of this article is to present the possibilities of applying permutation methods in the analysis of dependencies. Berry, Johnston and Mielke Jr. (2018) present a permutation approach for generating resampling probability values for various measures of association. They show many examples of the practical use of permutation methods in association analysis. The article presents selected possibilities of applying permutation methods in the analysis of association between variables. The main focus is on various possibilities of using permutation methods and the possibilities of data labels rearranging. A proposed method to allow investigation of the significance of the relationship between two data sets is presented. The presented example illustrates the use of permutation methods for testing the significance of the relationship between two sets of variables. The proposed method is compared to the well-known canonical correlation analysis.

2. Permutation Methods

In mathematics, the term "permutation" denotes the act of rearranging objects or values in an ordered fashion. In statistics, the idea of permutation objects is used in several methods, especially in testing hypotheses (permutation tests).

Permutation tests permit us to choose the form of the test statistic. Through sample size reduction, permutation tests can reduce the costs of experiments and surveys. Permutation tests are the most powerful of statistical procedures. There are five steps in the process of permutation testing (Good 2005):

1. Identify the null and the alternative hypothesis.
2. Choose the form of the test statistic.
3. Calculate the test statistic for the sample data.
4. Determine the frequency distribution of the test statistic using data permutations.
5. Make a decision using this empirical distribution as a guide.

The basic idea in permutation testing is to generate a reference distribution by recalculating a test statistic for many permutations of the data. The term "permutation methods" should not, however, be limited to the problem of testing hypotheses. In recent years, various statistical methods have been proposed that refer to the idea of the permutation model introduced by J. Spława-Neyman (1923).

Corain, Arboretti and Bonnini (2016) present a novel permutation-based nonparametric approach for ranking several multivariate populations. Using data collected from both experimental and observational studies, it covers some of the most useful designs widely applied in research and industry investigations, such as the multivariate analysis of variance and multivariate randomized complete block designs.

Berry, Johnston and Mielke Jr. (2018) use rearranging data to generate probability values and measures of effect size for various measures of association. They define association for two interval-level variables, measures of association for two nominal-level variables or two ordinal-level variables, and measures of agreement for two nominal-level or two ordinal-level variables.

Berry, Mielke Jr. and Johnston (2016) provide a synthesis of many statistical tests and measures, which, at first sight, appear disjointed and unrelated. Numerous comparisons of permutation and classical statistical methods are presented, and the two classes are compared via probability values and, where appropriate, measures of effect size.

Mielke and Berry Jr. (2007) offer a broad treatment of statistical inference using permutation techniques. Its purpose is to make available to practitioners a variety of useful and powerful data analysis tools that rely on very few distributional assumptions. Although many of these procedures have appeared in journal articles, they are not readily available to practitioners.

3. Permutational Methods in Hypothesis Testing

3.1. General Remarks

The testing of statistical hypotheses is a major branch of study in classical statistical inference. Based on a relatively small sample, one can infer the characteristics of a much larger population. There are two basic kinds of statistical hypotheses: parametric and non-parametric. According to the parametric or non-parametric hypothesis, a parametric statistical test or a nonparametric test is used. Parametric tests require that the sample is taken from a specified distribution, usually the normal distribution. Permutation tests such as nonparametric tests do not require specific population distributions of the variables such as the normal distribution. These tests use data labels rearranging. A typical application of permutation tests is to compare the distributions of two or more populations based on two samples taken independently.

3.2. Methods of Data Permutations

The permutation model is based on the data labels permutation. O'Gorman (2012, p. 78) lists some permutation methods which can be used for testing the significance of the coefficient in the linear regression model. In the linear regression model there is one dependent variable Y and k independent variables $X_1, X_2, \ldots, X_k$. Some methods of rearranging data labels include:

– permute the dependent variable,
– permute the independent variable for the considered variable,
– permute the residuals from the reduced model,
– permute the residuals from the complete model.

To test the significance of a parameter in the linear regression model, various methods of rearranging data labels can be used. The results of testing the significance of the coefficient for different permutation methods are not equivalent. O'Gorman (2012) points out that it is not clear which method is superior.

Let us assume that $\boldsymbol{y} = [y_1, y_2, \ldots, y_k]^T$ and we want to shuffle this vector. To permute this vector we can use the square matrix of dimension $k \times k$

$\mathbf{P}_k = [a_{ij}]$ where a_{ij} are only zeros and ones for $i, j = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} a_{ij} = 1$ for $j = 1, 2, \ldots, k$ and $\sum_{j=1}^{k} a_{ij} = 1$ for $i = 1, 2, \ldots, k$.

The examples of a permutation matrix in the case $k = 4$ are:

$$\mathbf{P}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ \mathbf{P}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \ \mathbf{P}_3 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

It is easy to notice that for the first matrix we have

$$\mathbf{P}_1 \boldsymbol{y} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} y_2 \\ y_3 \\ y_1 \\ y_4 \end{bmatrix}.$$

So vector $[y_2, y_3, y_1, y_4]^T$ is a rearrangement of vector $[y_1, y_2, y_3, y_4]^T$.
For the above-given permutation matrices $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ we have

$$\mathbf{P}_1 \boldsymbol{y} = [y_2, y_3, y_1, y_4]^T,$$

$$\mathbf{P}_2 \boldsymbol{y} = [y_1, y_4, y_2, y_3]^T,$$

$$\mathbf{P}_3 \boldsymbol{y} = [y_2, y_3, y_4, y_1]^T.$$

3.3. Properties of Permutations

Let $\mathbf{A}$ be the vector given by $\mathbf{A} = [y_1, y_2, \ldots, y_k]^T$. The permutation of vector $\mathbf{A}$ is function α: $\mathbf{A} \rightarrow \mathbf{A}$ that is bijective (i.e. both one-to-one and onto).

Identity permutation. Let $\mathbf{D} = [d_{ij}]$ be the identity matrix such that $d_{ii} = 1$, and $d_{ij} = 0$ for $i \neq j$ then matrix $\mathbf{D}$ leads to the identity permutation. If $\mathbf{D}$ is the matrix of permutation α, then for each vector $\mathbf{A}$: $\alpha\mathbf{A} = \mathbf{A}$.

Composition of permutations. Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be two matrices of permutations α and β, then $\mathbf{S}_2 \mathbf{S}_1$ is the matrix of the composition of permutations $\gamma = \beta\alpha$.

Inverse permutation. Let $\mathbf{S}$ be the matrix of permutation α, then $\mathbf{S}^{\mathbf{T}}$ is the matrix of inverse permutation β, that $\beta\alpha\mathbf{A} = \alpha\beta\mathbf{A} = \mathbf{A}$ for each vector $\mathbf{A}$.

Let $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$ be the matrices of permutation α, β and γ. Permutation composition is associative, e.g. $\mathbf{S}_3\,\mathbf{S}_2\,\mathbf{S}_1$ is the matrix of permutation $\gamma(\beta\alpha) = (\gamma\beta)\alpha$.

4. Permutation Methods for Regression Models

Pearson's correlation coefficient ρ tells us about the strength of the linear relationship between two variables *X* and *Y*. To perform a test of the significance of the correlation coefficient a random sample of size *n* should be taken. The sample data is used to calculate *r*, the correlation coefficient for the sample. Permutation tests use data labels rearranging to obtain the empirical distribution of the coefficient. The data should be permuted *N* times (*N* is usually greater or equal to 1000). The empirical distribution of *r* is obtained from *N* permuted sets of data. The decision of hypothesis H_0 is based on the value of *r* and its empirical distribution. The original two-dimensional data and three examples of randomly permuted data are presented in Figure 1.

X	*Y*		*X*	*Y*		*X*	*Y*		*X*	*Y*
*x*1	*y*1		*x*1	*y*2		*x*1	*y*6		*x*1	*y*5
*x*2	*y*2		*x*2	*y*1		*x*2	*y*5		*x*2	*y*1
*x*3	*y*3		*x*3	*y*9		*x*3	*y*8		*x*3	*y*9
*x*4	*y*4		*x*4	*y*6		*x*4	*y*7		*x*4	*y*7
*x*5	*y*5		*x*5	*y*10		*x*5	*y*9		*x*5	*y*6
*x*6	*y*6		*x*6	*y*4		*x*6	*y*4		*x*6	*y*8
*x*7	*y*7		*x*7	*y*7		*x*7	*y*1		*x*7	*y*3
*x*8	*y*8		*x*8	*y*8		*x*8	*y*10		*x*8	*y*2
*x*9	*y*9		*x*9	*y*5		*x*9	*y*2		*x*9	*y*4
*x*10	*y*10		*x*10	*y*3		*x*10	*y*3		*x*10	*y*10

Fig. 1. Example of a Two-dimensional Data Set (*X*, *Y*) and Three Random Data Sets (*X*, *Y*) with Randomly Rearranged Variable *Y*

Source: author's own elaboration.

There are more possibilities of rearranging data labels in the case of a multiple regression model than in the case of testing the significance of the correlation coefficient. Let us consider the multiple regression linear model:

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_k X_k + e. \tag{1}$$

There are many ways of data labels rearranging for one dependent variable Y and k independent variables X_1, X_2, …, X_k. There are some examples of possible ways of permutation of one dependent variable and k ($k > 2$) independent variables:

– permute the dependent variable Y (it is the same result as permute all independent variables $X_1, X_2, \ldots, X_k$ in the same way),

– permute only one independent variable X_i ($i = 1, 2, \ldots, k$),

– permute two independent variables X_i and X_j ($i, j = 1, 2, \ldots, k, i \neq j$) in the same way,

– permute two independent variables X_i and X_j ($i, j = 1, 2, \ldots, k, i \neq j$) independently,

– permute two independent variables X_i and X_j ($i, j = 1, 2, \ldots, k, i \neq j$) in the same way and the variable X_s ($s = 1, 2, \ldots, k$) independently.

Labels of original data should be rearranged to test the significance of the parameters of the linear model. The empirical distribution of the parameter is obtained based on the results for the N permuted models. Let S, S_1, S_2,…, S_k are the permutation matrices. Some typical methods of permutation for the multiple regression model (1) for $k = 3$ are:

$$Y = a_1 SX_1 + a_2 SX_2 + \cdots + a_k SX_k + e$$

$$Y = a_1 S_1 X_1 + a_2 XS_2 X_2 + \cdots + a_k S_k X_k + e$$

$$Y = a_1 X_1 + a_2 SX_2 + \cdots + a_k SX_k + e.$$

5. The Significance of the Association of Two Sets of Variables

In addition to the methods of studying the relationship between two variables or a dependent variable and many dependent variables, there are statistical methods that measure the association between two sets of variables **X** and **Y**. One of these methods is the canonical correlations analysis. Let **X** = (X_1, …, X_p) and **Y** = (Y_1, …, Y_q) be two sets of variables. To determine the strength of the association the permutation method could be used. The use of permutation methods requires data labels to be rearranged. These two sets can be rearranged in many ways. The variables in one set or two sets can be permuted independently or dependently.

The variables in one set (or two sets) could be grouped and permuted within these groups independently or dependently. The original set of variables for $p = q = 3$ and two examples of permuted sets are presented in Figure 2. In the first case (on the left), the variables X_1, X_2 and X_3 are permuted dependently. In the second case (on the right), all variables are permuted independently.

*Y*1	*Y*2	*Y*3	*X*1	*X*2	*X*3
*y*11	*y*21	*y*31	*x*11	*x*21	*x*31
*y*12	*y*22	*y*32	*x*12	*x*22	*x*32
*y*13	*y*23	*y*33	*x*13	*x*23	*x*33
*y*14	*y*24	*y*34	*x*14	*x*24	*x*34
*y*15	*y*25	*y*35	*x*15	*x*25	*x*35
*y*16	*y*26	*y*36	*x*16	*x*26	*x*36
*y*17	*y*27	*y*37	*x*17	*x*27	*x*37
*y*18	*y*28	*y*38	*x*18	*x*28	*x*38
*y*19	*y*29	*y*39	*x*19	*x*29	*x*39

*Y*1	*Y*2	*Y*3	*X*1	*X*2	*X*3
*y*11	*y*21	*y*31	*x*12	*x*22	*x*32
*y*12	*y*22	*y*32	*x*14	*x*24	*x*34
*y*13	*y*23	*y*33	*x*16	*x*26	*x*36
*y*14	*y*24	*y*34	*x*15	*x*25	*x*35
*y*15	*y*25	*y*35	*x*11	*x*21	*x*31
*y*16	*y*26	*y*36	*x*18	*x*28	*x*38
*y*17	*y*27	*y*37	*x*19	*x*29	*x*39
*y*18	*y*28	*y*38	*x*13	*x*23	*x*33
*y*19	*y*29	*y*39	*x*17	*x*27	*x*37

*Y*1	*Y*2	*Y*3	*X*1	*X*2	*X*3
*y*11	*y*23	*y*35	*x*12	*x*21	*x*31
*y*12	*y*22	*y*34	*x*14	*x*27	*x*32
*y*13	*y*21	*y*39	*x*13	*x*28	*x*36
*y*14	*y*25	*y*37	*x*18	*x*24	*x*33
*y*15	*y*24	*y*31	*x*15	*x*26	*x*34
*y*16	*y*26	*y*33	*x*17	*x*22	*x*38
*y*17	*y*29	*y*36	*x*16	*x*29	*x*39
*y*18	*y*28	*y*32	*x*11	*x*25	*x*37
*y*19	*y*27	*y*38	*x*19	*x*23	*x*35

Fig. 2. The Original Set of Two Sets of Variables (Top) and Two Examples of Possible Permutations of These Sets

Source: author's own elaboration.

6. Canonical Correlations

Pearson's correlation coefficient measures the strength of the linear dependency for two variables *X* and *Y*. The multiple linear regression model could be used to describe the dependency between the dependent variable *Y* and a set of independent variables $X_1, X_2, \ldots, X_k$. Sometimes the association between the two sets of variables **X** and **Y** should be considered. Canonical correlation analysis could be used in this case. This method was proposed by H. Hotelling in 1935–36. Canonical correlation analysis is employed to study the relationships between two variable sets when each variable set consists of at least two variables. The main objectives of canonical correlations are as follows (Thompson 1984):

– determining the strength of the relationships that may exist between the two sets,

– deriving a set of weights for each set of dependent and independent variables so that the linear combinations of each set are maximally correlated,

– explaining the nature of any relationships existing between the sets of dependent and independent variables.

Canonical correlation analysis develops several independent canonical functions that maximize the correlation between the linear composites, also known as canonical variates, which are sets of dependent and independent variables.

The form of Pearson's correlation coefficient:

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}. \tag{2}$$

The form of the canonical correlation coefficient:

$$\rho = \max_{a \in R^p, b \in R^q} \frac{Cov(\boldsymbol{a}'\mathbf{X}, \boldsymbol{b}'\mathbf{Y})}{\sqrt{Var(\boldsymbol{a}'\mathbf{X})Var(\boldsymbol{b}'\mathbf{Y})}}. \tag{3}$$

Canonical correlation analysis constructs such vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ based on the following criteria:

1. The first canonical variate $\boldsymbol{U}_1 = \boldsymbol{a}_1'\boldsymbol{X}, \boldsymbol{V}_1 = \boldsymbol{b}_1'\boldsymbol{Y}$ is constructed from the maximization of (3).

2. The second canonical variate pair $\boldsymbol{U}_2 = \boldsymbol{a}_2'\boldsymbol{X}, \boldsymbol{V}_2 = \boldsymbol{b}_2'\boldsymbol{Y}$ is constructed from the maximization of (3) with the restriction that $\boldsymbol{D}^2(\boldsymbol{U}_2) = \boldsymbol{D}^2(\boldsymbol{V}_2) = 1$ and pairs $(\boldsymbol{U}_1, \boldsymbol{V}_1)$ and $(\boldsymbol{U}_2, \boldsymbol{V}_2)$ are uncorrelated.

3. At the k step, the k-th canonical variate pair $\boldsymbol{U_k} = \boldsymbol{a}'_k\boldsymbol{X}, \boldsymbol{V_k} = \boldsymbol{b}'_k\boldsymbol{Y}$ is obtained from the maximization of (3) with the restriction that $\boldsymbol{D}^2(\boldsymbol{U_k}) = \boldsymbol{D}^2(\boldsymbol{V_k}) = 1$ and $(\boldsymbol{U_k}, \boldsymbol{V_k})$ are uncorrelated with the previous $(k - 1)$ canonical variate pairs.

4. Repeat step 3 until the number of canonical variates $s = \min(p, q)$.

The first canonical correlation coefficient is denoted by r_1, the second by r_2, and so on. The number of calculated CCA coefficients is equal to the minimum of the number of variables in two considered sets $s = \min(p, q)$. For testing the significance of the first canonical correlation coefficient, the following statistics can be used: Wilks' Lambda, the Hotelling-Lawley Trace, the Pillai-Bartlett Trace, and Roy's Largest Root.

7. Testing the Significance of the Association of Two Sets of Variables

One problem to be considered in the canonical correlation analysis is to test the hypothesis that none of the canonical correlation coefficients r_1, r_2, … , r_s is significant. To test such a hypothesis we usually use Wilks' lambda statistic, the Hotelling-Lawley trace, the Pillai trace or Roy's largest root. The form of these statistics is as follows:

Wilk's lambda statistic:

$$\Lambda_1 = \prod_{i=1}^{m} (1 - r_i^2). \tag{4}$$

This statistic is distributed as the Wilks Λ-distribution. Rejection of the null hypothesis is for small values of Λ_1.

Pillai trace statistic:

$$V^{(m)} = \sum_{i=1}^{m} r_i^2. \tag{5}$$

Rencher and Christensen (2012, pp. 391–95) has tables providing critical values for this statistic.

Hotelling-Lawley trace statistic:

$$U^{(m)} = \sum_{i=1}^{m} \frac{r_i^2}{1 - r_i^2}. \tag{6}$$

Rencher and Christensen (2012, pp. 391–95) has tables providing critical values for this statistic.

Roy's largest root statistic:

$$\theta = r_1^2. \tag{7}$$

Rencher and Christensen (2012, pp. 391–95) present tables providing critical values for Roy's largest root statistic. This statistic is based only on one largest canonical correlation coefficient and the other statistics use all the canonical correlation coefficients. For this reason, Roy's largest root statistic will not be included in the computer simulation studies.

8. Monte Carlo Study

The proposed method of testing the significance of dependency between two sets of variables is based on the permutation method. In the proposed method, Wilk's lambda statistic (4) is used and the dependent rearrangement of one of the variable set labels is used (see Figure 2). The proposal was compared to the three well-known methods of testing the significance of correlations in the canonical analysis in the Monte Carlo study. The permutation had $N = 1000$ random rearrangements of data labels. The significance level $\alpha = 0.05$ was assumed in computer simulations.

In the first part of a computer simulation, two sets of variables **X** and **Y** were considered. All variables were normally distributed. The sample size was $n = 100$. The above assumptions are typical for the canonical correlation analysis.

The first set of variables is given by $\mathbf{X} = (X_1, X_2, X_3)$, where variables X_1, X_2, X_3 are independent and normally distributed with mean 10 and variance 1. The second set of variables is given by $\mathbf{Y} = (1 - \delta)\mathbf{Z} + \delta\mathbf{X}$, where $\mathbf{Z} = (Z_1, Z_2, Z_3)$, and variables Z_1, Z_2, Z_3 are independent and normally distributed with mean 10 and variance 1. The parameter δ ($\delta = 0.00, 0.05, \ldots, 0.25$) describes the strength of the association of two sets of variables. For $\delta = 0$ the sets **X** and **Y** are not associated.

The estimated probabilities of rejection H_0 (there is no association between **X** and **Y**) are presented in Table 1. The probabilities were estimated based on 1000 random samples. In the case of normally distributed variables with the sample of size $n = 100$, the size of all considered tests is close to the significance level α. The power of all tests is quite similar. The proposal has no advantage in this case.

Table 1. Estimated Probability of Rejection H_0 – Normal Distribution, $n = 100$

δ	Method			
	Permutation	Wilk's Lambda	Hotelling--Lawley trace	Pillai trace
0.00	0.051	0.053	0.054	0.049
0.05	0.081	0.080	0.082	0.075
0.10	0.199	0.202	0.204	0.193
0.15	0.497	0.495	0.499	0.487
0.20	0.866	0.861	0.861	0.862
0.25	0.993	0.991	0.990	0.989

Source: author's own elaboration.

The advantages of permutation tests are for the application of small samples and non-normally distributed variables. The next three simulations are based on normal, beta and gamma-distributed variables for small samples where $n = 20$.

The association of two sets **X** and **Y** of the three-dimensional variables were analysed in this part of the study. The considered sets are $\mathbf{X} = (X_1, X_2, X_3)$ and $\mathbf{Y} = (1 - \delta)\mathbf{Z} + \delta\mathbf{X}$ where $\mathbf{Z} = (Z_1, Z_2, Z_3)$, variables Z_1, Z_2, Z_3 are independently distributed and the parameter δ (δ = 0.00, 0.05, …, 0.40) describes the strength of association of two sets.

The parameters of random variables X_1, X_2, X_3 and Z_1, Z_2, Z_3 are as follows:

– normal distribution with expected value 10 and variance 1,
– beta distribution with shape parameters $s_1 = 2$ and $s_2 = 2$,
– gamma distribution with shape parameter $s = 2$.

The probabilities of rejection of H_0 are presented in Tables 2, 3, and 4. The probabilities were estimated based on 1000 random samples. In the permutation method for each sample, the data labels were randomly rearranged $N = 1000$ times. In the first case, samples were taken from the normal distribution but the sample size was $n = 20$. The results are presented in Table 2. In the case of normally distributed variables and small samples (Table 2), the test based on the permutation method has the greatest power. The three other tests have similar power. The results of the computer study for beta distributed variables are presented in Table 3.

Table 2. Estimated Probability of Rejection H_0 – Normal Distribution, $n = 20$

δ	Method			
	Permutation	Wilk's Lambda	Hotelling--Lawley trace	Pillai trace
0.00	0.046	0.041	0.052	0.033
0.05	0.069	0.064	0.078	0.047
0.10	0.070	0.074	0.087	0.046
0.15	0.088	0.090	0.110	0.056
0.20	0.186	0.173	0.187	0.144
0.25	0.285	0.263	0.278	0.226
0.30	0.496	0.468	0.479	0.438
0.35	0.726	0.688	0.682	0.664
0.40	0.907	0.873	0.858	0.859

Source: author's own elaboration.

Table 3. Estimated Probability of Rejection H_0 – Beta Distribution, $n = 20$

δ	Method			
	Permutation	Wilk's Lambda	Hotelling--Lawley trace	Pillai trace
0.00	0.056	0.058	0.074	0.042
0.05	0.057	0.055	0.071	0.041
0.10	0.069	0.062	0.073	0.045
0.15	0.101	0.095	0.111	0.066
0.20	0.155	0.149	0.172	0.114
0.25	0.252	0.242	0.253	0.202
0.30	0.495	0.449	0.457	0.411
0.35	0.701	0.653	0.649	0.639
0.40	0.904	0.887	0.870	0.883

Source: author's own elaboration.

The size of the Hotelling-Lawley test is greater than the significance level α in the case of beta distributed variables. The most powerful test is the test based on the permutation method. The results of the computer study for gamma-distributed variables are presented in Table 4.

Table 4. The Estimated Probability of Rejection H_0 – Gamma Distribution, $n = 20$

δ	Method			
	Permutation	Wilk's Lambda	Hotelling--Lawley trace	Pillai trace
0.00	0.055	0.061	0.068	0.039
0.05	0.059	0.074	0.083	0.043
0.10	0.075	0.080	0.091	0.057
0.15	0.122	0.127	0.141	0.089
0.20	0.170	0.173	0.193	0.136
0.25	0.313	0.291	0.316	0.258
0.30	0.498	0.474	0.480	0.436
0.35	0.725	0.701	0.695	0.681
0.40	0.903	0.887	0.876	0.879

Source: author's own elaboration.

The size of the Hotelling-Lawley test is greater than the significance level α in the case of gamma-distributed variables. The most powerful test is the test based on the permutation method.

9. Conclusions

Permutation tests are one type of permutation method. They are very useful for non-normally distributed data and small samples. However, permutation methods can be used not just for testing hypotheses. They can be used for constructing rankings of multivariate data or measuring the association between variables and sets of variables. In these methods, the form of data labels rearrangement is very important. The data labels can be permuted in many ways. Some of these methods of rearranging labels which can be used in the association analysis are described in the paper.

As an example of the use of the permutation method in association analysis, the method of testing the significance of the association between two sets of variables is presented in the paper. This method uses Wilk's lambda statistic and is based on the dependent rearranging of data labels in variables in one of the considered sets. The properties of the proposal were compared to the well-known tests in canonical correlation analysis: Wilk's lambda, the Lawley-Hotelling trace, and Pillai trace. The Monte Carlo study shows that for normally distributed variables and big samples the power of all tests is quite similar. The proposal is most powerful for non-normally distributed variables and small samples.

Bibliography

Berry, K. J., Johnston, J. E. and Mielke Jr., P. W. (2014) *A Chronicle of Permutation Statistical Methods*. New York: Springer International Publishing.

Berry, K. J., Johnston, J. E. and Mielke Jr., P. W. (2018) *The Measurement of Association. A Permutation Statistical Approach.* Cham: Springer Nature Switzerland.

Berry, K. J., Mielke Jr., P. W. and Johnston, J. E. (2016) *Permutation Statistical Methods. An Integrated Approach.* Cham: Springer Nature Switzerland.

Corain, L., Arboretti, R. and Bonnini, S. (2016) *Ranking of Multivariate Populations. A Paermutation Approach with Applications.* Boca Raton: CRC Press.

Good, P. (2005) *Permutation, Parametric and Bootstrap Tests of Hypotheses*. New York: Springer Science Business Media.

Ledwina, T. (2012) "Neyman Jerzy (1894–1981)" in: *Statystycy polscy.* Warszawa: Główny Urząd Statystyczny, Polskie Towarzystwo Statystyczne.

Lehmann, E. L. and Romano, J. P. (2005) *Testing Statistical Hypotheses*. New York: Springer-Verlag.

Mielke, P. W. and Berry K. J. Jr. (2007) *Permutation Methods. A Distance Function Approach*. New York: Springer Science+Business Media.

O'Gorman, T. W. (2012), *Adaptive Tests of Significance Using Permutations of Residuals with R and SAS.* John Wiley and Sons.

Rencher, A. C. and Christensen, W. F. (2012) *Methods of Multivariate Analysis*. Hoboken: Wiley.

Spława-Neyman, J. (1923) "Próba uzasadnienia zastosowań rachunku prawdopodobieństwa do doświadczeń polowych". *Rocznik Nauk Rolniczych* 10: 1–51.

Thompson, B. (1984) *Canonical Correlation Analysis: Uses and Interpretation*. London: Sage Publications.

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 1(22) • 2020
ISSN 1642-168X
e-ISSN 2545-3866
AOC, 2020; 1(22): 47–58
https://doi.org/10.15678/AOC.2020.2204

Monika Banaszewska

CONDEMNED TO SUCCESS (FAILURE)? THE ALLOCATION OF EU FUNDS AMONG POLISH MUNICIPALITIES

Abstract

Objective: The aim of the paper is twofold. First, we investigate the mobility of Polish municipalities within the distribution of *per capita* EU fund expenditures over two terms of office: 2007–2010 and 2011–2014. The resulting joint distribution serves as the basis for some empirical analyses. Second, we consider the relationship between mobility and social capital in 16 Polish regions.
Research Design & Methods: We make use of a transition probability matrix and rank-rank regression. Additionally, we employ Spearman's rank correlation and Kendall's rank correlation. Both nation-wide and region-specific analyses are conducted.
Findings: The municipalities most (least) successful in attracting EU funds in the 2007–2010 period tended to maintain their positions in the 2011–2014 period. The relative and absolute mobility of municipalities – EU funds beneficiaries – differ considerably across regions. There is a significant negative correlation between within-region absolute mobility and the level of bridging social capital within regions.
Implications/Recommendations: The persistence of municipalities at the ends of the EU fund absorption ranking can facilitate forecasts of the spatial allocation of EU funds and, consequently, its effects.
Contribution: To the author's best knowledge, the study constitutes the first empirical analysis of within-distribution mobility of municipalities – EU funds beneficiaries. We also provide an initial study on the nexus between social capital and mobility within the ranking of EU funds beneficiaries. Adding to an earlier study by Swianiewicz *et al.* (2008), we show that the allocation of EU funds in Poland is affected by the level and structure of social capital.

Monika Banaszewska, Poznań University of Economics and Business, Department of Public Finance, al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: monika.banaszewska@ue.poznan.pl, ORCID: https://orcid.org/0000-0003-2787-9986.

Keywords: Cohesion Policy, relative mobility, absolute mobility, transition probability matrix.
JEL Classification: H72, H77.

1. Introduction

The primary focus of EU cohesion policy is on NUTS level 2 regions. It is justified by the convergence of productivity and GDP *per capita* across countries combined with persistent disparities among regions within countries (Giannetti 2002; Baldwin & Wyplosz 2009, pp. 386–389; Bachtler *et al.* 2017). In Poland NUTS level 2 regions are equivalent to administrative regions (voivodeships). In the 2007–2013 programming period, Polish regional self-governments played a crucial role in allocating EU funds. They performed the role of managing authorities for 16 regional operational programs. They also took an active part in implementing some priority axes in central programs as well as funds for rural areas and the fisheries sector. One should bear in mind that the allocation of EU funds is further disaggregated within regional borders. Lower-level governments (districts and especially municipalities) accounted for prominent beneficiaries of EU funds. In this paper we consider EU fund expenditures by municipality. Municipalities represent the lowest self-government tier (NUTS level 5). In relation to the EU cohesion policy administrative framework in Poland, municipalities are grouped by region.

2. Related Literature

Our paper builds on two strands of literature regarding, respectively: the allocation of EU funds and social capital measures. There is an increasing number of studies that focus on the determinants of the distribution of EU funds at the municipal level. A few examples are: Veiga (2012) for Portugal, Muraközy and Telegdy (2016) for Hungary, and Banaszewska and Bischoff (2017) for Poland. These studies report average effects within the analysed samples, which aids the assessment of whether managing authorities meet allocation objectives. They also allow the political economy aspect of EU fund allocation to be identified. Nevertheless, previous analyses documenting average effects should be supplemented with detailed analyses of the distribution of variables of interest. As an illustrative example, in their assessment of the effects of EU funds in Poland, Misiąg, Misiąg and Tomalak (2013) obtain a general inverse relationship between the size of municipality and the amount of attracted EU funds *per capita*. At the same time, the authors stress that the general relationship was driven by three regions (voivodeships): Małopolskie, Opolskie, and Podlaskie.

The existing literature says little about the mobility of EU funds beneficiaries within the distribution, which is equally crucial from a social welfare perspective. Kyriacou and Roca-Sagalés (2012) show that above a specific level of concentration of EU funds (estimated at 1.6% of a country's GDP), EU structural funds are likely to increase regional disparities. In a similar vein, Becker, Egger and von Ehrlich (2012) reveal that the reduction in EU fund support for NUTS 3 regions that already receive grants exceeding 1.3% of their GDP would not hamper their growth prospects. Given this result, they opt for a reallocation of EU funds across regions as a welfare-enhancing measure. This argument is relevant for Poland, as the concentration of EU funds there is among the highest in the EU.

Social capital is believed to be a fundamental factor driving economic growth and development (see, for instance: Knack & Keefer 1997, Woolcock 1998, Whiteley 2000). Besides this general consensus, there is an ongoing scholarly debate on how to properly measure social capital. Bednarek-Szczepańska (2013) offers a critical review of composite indices of social capital in Poland. Among the publications reviewed, only the one by Swianiewicz *et al.* (2008) provides indices of different types of social capital at the regional (voivodeship) level. The others report either a single measure of social capital and/or relate to lower-level statistical units (subregions, counties).

Swianiewicz *et al.* (2008) offer a unique study that incorporates both the perspective of EU funding policy and the area of social capital. Building on Putnam (1993), they construct composite indices of bridging and bonding social capital in Polish regions (voivodeships). The areas encompassed by these indices are listed in Figure 1. In regard to the 2004–2006 financial perspective, Swianiewicz *et al.* (2008) provide tentative evidence that bonding capital fosters EU fund absorption whereas bridging capital hinders this process.

Bonding capital	Bridging capital
– family and friendship ties – neighbourly ties – religious practices	– involvement in activities for the benefit of the local community – volunteering

Fig. 1. Components of the Indices of Bonding and Bridging Capital

Source: (Swianiewicz *et al.* 2008, pp. 83–102).

These considerations motivate the following research questions:

1) What are the levels of absolute and relative upward mobility in the EU fund utilization ranking in Poland overall and within 16 Polish regions?

2) Is the within-region mobility of municipalities – EU funds beneficiaries – correlated with the levels of bridging and bonding social capital?

We add to the empirical literature on EU fund distribution in Poland in two ways. To the author's best knowledge, we offer the first empirical analysis of within-distribution mobility of EU funds beneficiaries. To this end, we employ methods typical of studies on intergenerational income (wealth) mobility, such as Chetty *et al.* (2014b) and Chetty *et al.* (2014a). Second, we provide an initial study on the nexus between social capital and mobility within the ranking of EU funds beneficiaries. What makes the Polish case instructive is both the amount and the scope of EU support for municipal governments.

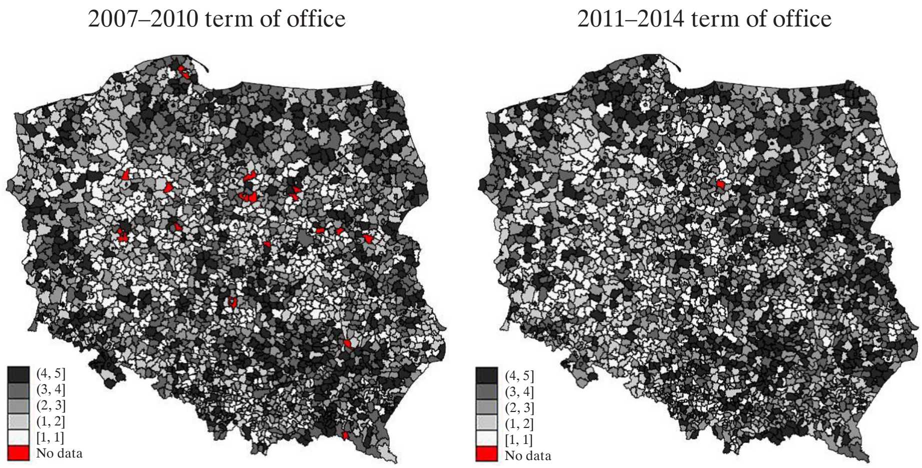
3. Data and Methods

In this paper, we investigate whether the set of municipalities (un)successful in attracting EU funds is persistent over time. The variable under investigation is *per capita* EU fund expenditures excluding co-financing, i.e. expenditures recorded in municipal budgets with 4th paragraph digit 1, 5, 7, and 8. The data encompass not only structural and cohesion funds but also funds for agriculture and fisheries, which are of particular importance in rural areas. The drawback of our approach is that we refer to the total amount of EU funds spent by municipalities. Unfortunately, the data reported to the Ministry of Finance cannot be broken down into funds allocated at the central and local level. Still, we argue that these types of funds are closely related (as complements as well as substitutes) so that one should focus on their joint allocation.

In order to ensure over-time comparability, the expenditures are expressed in real terms with the use of the GDP deflator. The data are from Ministry of Finance database and National Bank of Poland website. To relax the disturbance from short-term hikes (drops) in spent funds, expenditures are cumulated over four-year periods, consistent with two local government terms of office: 2007–2010 and 2011–2014. Because in the analysed period local elections took place in November and December, the election years (i.e. 2010 and 2014) are considered to be the last year of a given term-of-office. As a result, we obtain two observations for each municipality.

We use this joint distribution of (Y_i, X_i) (copula) to describe the mobility of EU funds beneficiaries.

Figure 2 shows the position of municipalities within quintiles according to their *per capita* EU fund expenditures over the 2007–2010 and 2011–2014 terms of office (where 1 represents the bottom quintile, 2 represents the second quintile and so on up to 5, which is the top quintile). A visual inspection shows that the ranks are spatially correlated. It is also evident that only few municipalities received no EU funds in the analysed period. Therefore, our sample includes virtually all Polish municipalities. Specifically, we use data for 2,459 out of 2,479 of the lowest-tier entities in the 2007–2014 period.



1 – bottom quintile, 2 – second quintile, 3 – middle quintile, 4 – fourth quintile, 5 – top quintile, no data – no EU funds spent.

Fig. 2. EU Fund Expenditures *per capita* in Polish Municipalities over the 2007–2010 and 2011–2014 Terms of Office

Source: author's own calculations based on Ministry of Finance data.

First, we calculate a transition probability matrix. We divide municipalities into five groups (quintiles). Then, we calculate the probabilities for a municipality reaching a specific quintile in the 2011–2014 term of office, conditional on its quantile rank in the 2007–2010 term of office.

Next, for the purpose of further analysis of the persistence of municipalities (un)successful in attracting EU funds over time, we employ rank-rank regressions. This time, for the 2007–2010 and 2011–2014 periods, the numbers from 1 (bottom percentile) to 100 (top percentile) are assigned to each municipality. The regression is as follows:

$$\text{EU\_funds\_pc\_percentile\_2011\_2014}_i = \alpha + \beta \, \text{EU\_funds\_pc\_percentile\_2007\_2010}_i,$$

where: i – municipality.

The slope represents relative mobility, i.e. the difference between the 2011–2014 mean percentile rank of municipalities most successful in attracting EU funds in the years 2007–2010 versus municipalities least successful in that respect. Putting it differently, it shows the association between ranks in the 1st and 2nd election cycle. Importantly, the slope does not depend on the standard deviation of Y_i and X_i. The relative mobility is inversely related to the slope. The above regression can be also used to determine absolute upward mobility at a specific percentile. For example, at the 25th percentile the formula is $\alpha + 25\beta$. Absolute mobility increases with the obtained result (Chetty *et al.* 2014a, 2014b).

Regional (voivodeship) governments played a critical role in the allocation of EU funds in the 2007–2013 financial perspective (Banaszewska & Bischoff 2017). Taking into account this possible geographical variation, we extend our research by region-specific analyses. We do not reassign municipalities, however, but use the same ranking for country-wide as well as region-specific considerations[1].

We also test whether the region-specific mobility of municipalities – EU funds beneficiaries – is correlated with regional social capital. Since the number of observations is very low ($n = 16$), we calculate Spearman's rank correlation coefficients. We use the levels of bridging and bonding capital reported by Swianiewicz *et al.* (2008) and represent them with the use of four-level ordinal scale, where 1 – the lowest level of social capital, 4 – the highest level of social capital. Because of this, we supplement our correlation analysis with the Kendall rank correlation coefficient. Strictly speaking, we use tau-b, which makes adjustments for ties.

4. Results

We analyse the mobility of Polish municipalities – EU funds beneficiaries – within the distribution in the 2007–2010 period *versus* the 2011–2014 period. Table 1 presents the joint distribution of EU fund expenditures *per capita* in the 2007–2010 period (X_i) and EU fund expenditures in the

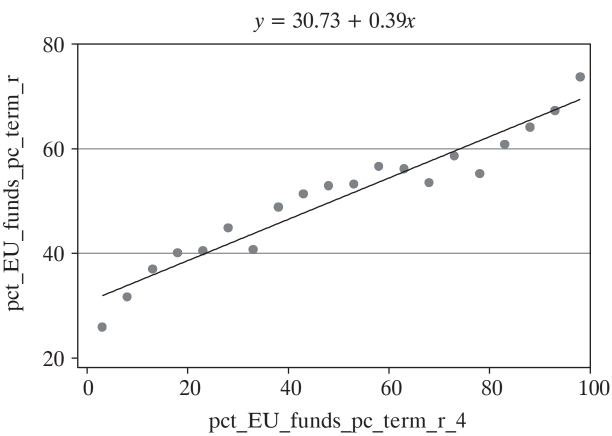
[1] An analogous approach is taken in, for instance, Chetty *et al.* (2014a).

2011–2014 period (Y_i). There is an indication of persistence at the ends of the distribution as 42.7% of the municipalities that received the lowest EU fund support in 2007–2010 also remained in the bottom quintile in the years 2011–2014. A similar percentage (41%) of municipalities most successful in attracting EU funds in the first term of office kept their top positions in the second four-year period. At the same time, dramatic movements within the distribution (from the bottom quintile to the top quintile and *vice versa*) were relatively rare. Their probability is lower than 10%.

Table 1. Transition Probability Matrix on EU Fund Expenditures *per capita* by Polish Municipalities over the Years 2007–2010 and 2011–2014

EU Fund Expenditures *per capita*		2011–2014 Term of Office				
		bottom quintile	second quintile	middle quintile	fourth quintile	top quintile
2007–2010 term of office	bottom quintile	42.7	22.4	14.2	12.4	8.3
	second quintile	23.8	25.0	23.6	16.3	11.4
	middle quintile	13.6	20.3	23.4	24.4	18.3
	fourth quintile	10.8	19.1	24.6	24.0	21.5
	top quintile	8.0	13.1	14.9	23.1	41.0

Source: author's own calculations based on Ministry of Finance data.



The diagram presents a binned scatter plot (20 bins in total) with a fitted line.
The equation of the fitted line is reported at the top of the diagram.

Fig. 3. Rank-rank Regression on Municipalities – EU Funds Beneficiaries – in the 2011–2014 Period *vs* the 2007–2010 Period in Poland

Source: author's own calculations based on Ministry of Finance data.

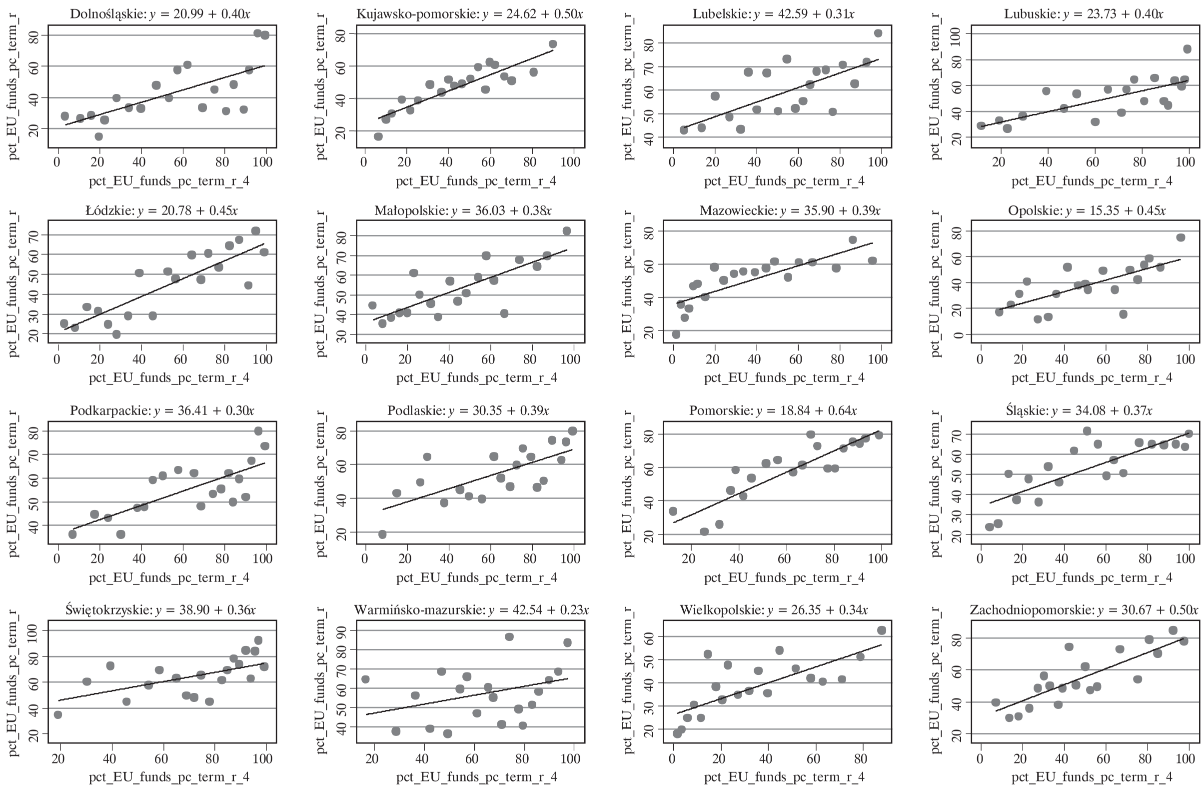
Figure 3 shows that the linear function accounts for a good approximation of rank-rank relationship in our sample. On average, municipalities initially most successful in EU fund acquisition (top percentile) end in a percentile that is higher by 39 positions than municipalities least successful in attracting EU funds. As for absolute mobility, municipalities located in the 25th percentile of the distribution in the 2007–2010 period are expected to move up to the 40th percentile in the second term of office.

In the next step we divide our sample into 16 administrative regions (voivodeships). The results are displayed in Figure 4 and Table 2. Relative upward mobility differs considerably across regions. It is highest for the Warmińsko-mazurskie region (slope equals 0.23) and lowest for Pomorskie (slope equals 0.64). We find that absolute mobility at the 25th percentile is highest for Lubelskie (mean end-period percentile is 50) and lowest for Opolskie (average end-period percentile is 27).

Table 2. Absolute Upward Mobility at the 25th Percentile of Municipalities – EU Funds Beneficiaries – in the 2011–2014 Period *vs* the 2007–2010 Period in Polish Regions

Region	Absolute Upward Mobility at the 25th Percentile
Dolnośląskie	30.99
Kujawsko-pomorskie	37.12
Lubelskie	50.34
Lubuskie	33.73
Łódzkie	32.03
Małopolskie	45.53
Mazowieckie	45.65
Opolskie	26.60
Podkarpackie	43.91
Podlaskie	40.10
Pomorskie	34.84
Śląskie	43.33
Świętokrzyskie	47.90
Warmińsko-mazurskie	48.29
Wielkopolskie	34.85
Zachodniopomorskie	43.17

Source: author's own calculations based on Ministry of Finance data.



The diagrams present binned scatter plots (20 bins in total) with fitted lines. The equations of fitted lines are reported at the top of each diagram.

Fig. 4. Rank-rank Regression on Municipalities – EU Funds Beneficiaries – in the 2011–2014 Period *vs* the 2007–2010 Period in Polish Regions

Source: author's own calculations based on Ministry of Finance data.

Table 3. Spearman's Rank Correlation Coefficients and Kendall's Rank Correlation Coefficients for Polish Regions in the 2007–2014 Period

Pair	Spearman's rho	Kendall's tau-b
Relative mobility of municipalities[a] – EU funds beneficiaries – and bridging social capital[b]	–0.433* (0.094)	–0.318 (0.130)
Absolute mobility of municipalities – EU funds beneficiaries – at the 25th percentile and bridging social capital	–0.595** (0.015)	–0.521** (0.011)
Relative mobility of municipalities – EU funds beneficiaries – and bonding social capital[b]	0.379 (0.148)	0.285 (0.186)
Absolute mobility of municipalities – EU funds beneficiaries – at the 25th percentile and bonding social capital	0.267 (0.317)	0.210 (0.329)

[a] Relative mobility is represented by the slope of rank-rank regression with a negative sign.
[b] Social capital indices are expressed with the use of a four-level scale, where 1 is the lowest level of social capital and 4 is the highest level of social capital.
$N = 16$; p-values in parentheses. Significance levels denoted as follows: * – significant at 10% level, ** – significant at 5% level.

Source: author's own calculations based on Ministry of Finance data.

Table 3 reports Spearman's rank correlation coefficients for relationships between mobility and social capital in Polish regions. Both Spearman's rho and Kendall's tau-b point to a negative and significant correlation between absolute mobility at the 25th percentile and bridging social capital. Similarly, the coefficient on the pair: relative mobility and bridging social capital is negative but loses significance once we use Kendall's rank correlation. Conversely, the correlations between bonding social capital and mobility are positive, However, they are not found to be significant at any conventional levels.

5. Conclusions

This paper offers an empirical analysis of movements within the ranking of *per capita* EU fund expenditures among Polish municipalities over two election cycles: 2007–2010 and 2011–2014. We find that the joint distribution of municipalities – EU funds beneficiaries – is characterized by "stockiness at the ends", i.e. municipalities most (least) successful in EU fund absorption rarely leave their top (bottom) positions. At the same time, using rank--rank regression, we document both the relative and absolute mobility of municipalities – EU beneficiaries, which varies considerably across regions.

We also find that within-region absolute mobility at the 25th percentile is negatively correlated with the level of bridging social capital. Comparing this result with the earlier study by Swianiewicz *et al.* (2008), we conclude that bridging social capital not only inhibits EU fund absorption but also preserves spatial patterns in EU expenditures. This result should be treated with caution as a correlation analysis does not allow any causal inferences to be made. In future studies it would be worth investigating which types of EU funds' regional administrative structures foster (hinder) mobility within the ranking of beneficiaries.

Bibliography

Bachtler, J., Martins, J. O., Wostner, P. and Zuber, P. (2017) *Towards Cohesion Policy 4.0: Structural Transformation and Inclusive Growth RSA Europe.* Regional Studies Association Brussels.

Baldwin, R. E. and Wyplosz, C. (2009) *The Economics of European Integration*. London: McGraw-Hill Education.

Banaszewska, M. and Bischoff, I. (2017) "The Political Economy of EU-funds: Evidence from Poland". *Jahrbücher für Nationalökonomie und Statistik* 237 (3), https://doi.org/10.1515/jbnst-2017-1105.

Becker, S. O., Egger, P. H. and von Ehrlich, M. (2012) "Too Much of a Good Thing? On the Growth Effects of the EU's Regional Policy". *European Economic Review* 56 (4), https://doi.org/10.1016/j.euroecorev.2012.03.001.

Bednarek-Szczepańska, M. (2013) "Zróżnicowanie przestrzenne kapitału społecznego w Polsce – ujęcie przeglądowe" [Spatial differentiation of social capital in Poland – a review]. *Przegląd Geograficzny* 85 (4), https://doi.org/10.7163/przg.2013.4.4.

Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014a) "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States". *The Quarterly Journal of Economics* 129 (4), https://doi.org/10.1093/qje/qju022.

Chetty, R., Hendren, N., Kline, P., Saez, E. and Turner, N. (2014b) "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility". *American Economic Review* 104 (5), https://doi.org/10.1257/aer.104.5.141.

Giannetti, M. (2002) "The Effects of Integration on Regional Disparities: Convergence, Divergence or Both?". *European Economic Review* 46(3), https://doi.org/10.1016/s0014-2921(01)00166-0.

Knack, S. and Keefer, P. (1997) "Does Social Capital Have an Economic Payoff? A Cross--country Investigation". *The Quarterly Journal of Economics* 112 (4), https://doi.org/10.1162/003355300555475.

Kyriacou, A. P. and Roca-Sagalés, O. (2012) "The Impact of EU Structural Funds on Regional Disparities within Member States". *Environment and Planning C: Government and Policy* 30 (2), https://doi.org/10.1068/c11140r.

Misiąg, J., Misiąg, W. and Tomalak, M. (2013) *Ocena efektywności wykorzystania pomocy finansowej Unii Europejskiej jako instrumentu polityki spójności społeczno--gospodarczej oraz poprawy warunków* życia [The assessment of European Union financial aid utilization efficiency as the instrument of socio-economic cohesion and the improvement in living conditions]. Rzeszów: Wyższa Szkoła Informatyki i Zarządzania.
Muraközy, B. and Telegdy, Á. (2016) "Political incentives and state subsidy allocation: Evidence from Hungarian municipalities". *European Economic Review* 89, https://doi.org/10.1016/j.euroecorev.2016.07.003.
Putnam, R. (1993) *Making Democracy Work.* Princeton: Princeton University Press.
Swianiewicz, P., Herbst, J., Lackowska, M. and Mielczarek, A. (2008) *Szafarze darów europejskich. Kapitał społeczny a realizacja polityki regionalnej w polskich województwach* [The dispensers of European donations. Social capital and the implementation of regional policy in Polish regions]. Warszawa: Scholar.
Veiga, L. G. (2012) "Determinants of the Assignment of EU Funds to Portuguese Municipalities". *Public Choice* 153 (1–2), https://doi.org/10.1007/s11127-011-9786-y.
Whiteley, P. F. (2000) "*Economic Growth and Social Capital". Political Studies* 48 (3), https://doi.org/10.1111/1467-9248.00269.
Woolcock, M. (1998) "Social Capital and Economic Development: Toward a Theoretical Synthesis and Policy Framework". *Theory and Society* 27 (2).

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 1 (22) • 2020
ISSN 1642-168X
e-ISSN 2545-3866
AOC, 2020; 1 (22): 59–72
https://doi.org/10.15678/AOC.2020.2205

Czesława Pilarska
Grzegorz Wałęga

BORN GLOBALS IN POLAND: DEVELOPMENT FACTORS AND RESEARCH OVERVIEW

Abstract

Objective: This paper investigates the concept of the born global, indicates the main attributes and sources of success for enterprises that internationalised early, and demonstrates the scale of the born global phenomenon in Poland.
Research Design & Methods: A review of empirical studies is the research method used in the paper. Secondary data from research studies is collected. The article outlines the theoretical background of born globals. A brief overview of the existing research (since 2004) is also provided. The results are compiled to demonstrate the scale of the born global phenomenon in Poland.
Findings: A review of quantitative research indicates that the percentage of born globals in Poland can be estimated at 30–50% of all SMEs which are engaged in international activities. Polish managers have the necessary knowledge and traits for early internationalisation; the primary factors limiting the development of born global are limited financial and organisational resources.
Implications/Recommendations: The existing quantitative analyses of born globals do not sufficiently enable the factors that constitute the main barriers to the development of this sector in Poland to be identified. There is still a room to plan the appropriate support policies for SMEs on international markets.
Contribution: This paper addresses mainstream research on born globals, which are part of broader trends in the modern economy. This new generation of small and

Czesława Pilarska, Cracow University of Economics, Department of Microeconomics, Rakowicka 27, 31-510 Kraków, Poland, e-mail: pilarskc@uek.krakow.pl, ORCID: https://orcid.org/0000-0003-4224-8517.

Grzegorz Wałęga, Cracow University of Economics, Department of Microeconomics, Rakowicka 27, 31-510 Kraków, Poland, e-mail: grzegorz.walega@uek.krakow.pl, ORCID: https://orcid.org/0000-0002-4355-5204.

medium-sized enterprises, just as economic globalisation, the opening of markets and acceleration in the field of ICT development, is particularly relevant to both developing countries and middle-income economies such as Poland.

Keywords: born globals, early internationalisation, SMEs, small and medium-sized enterprises.
JEL Classification: F21, F23, L26, M13, M16.

1. Introduction

Advancing globalisation as well as access to the internet and modern technologies are factors that have created new development opportunities for small and medium-sized enterprises related to expansion onto foreign markets. While the main actors of globalisation and internationalisation processes have primarily been large enterprises (transnational corporations), nowadays we are increasingly dealing with a new generation of small and medium-sized enterprises that have been globally active since launching their business activities. These firms, known as born globals or early internationalisation enterprises in the literature, are also gradually developing in Poland. The purpose of this article is to promote familiarisation with the born globals concept, indicate the main attributes and sources of success for these enterprises. The method used in the paper is a review of empirical research. We collect secondary data from research studies and synthesise their findings. The article begins with an outline of the theoretical background of born globals. We also give a brief overview of the existing research (since 2004). The results have been compiled to demonstrate the scale of the born global phenomenon in Poland.

2. Theoretical Background of Born Globals

At the outset, it should be noted that born globals are a special case with respect to the internationalisation of small and medium-sized enterprises and a challenge for traditional theories of the internationalisation of firms (Hollensen 2017). The internationalisation of born-global firms does not proceed in stages, sequentially, as in the Uppsala model (Johanson & Wiedersheim-Paul 1975, Johanson & Vahlne 1977). Instead, these firms are global "from inception". Gustafsson and Zasada (2011) emphasise that born-global firms usually create products with high added value that were not previously developed in a strong domestic market.

The concept of born global was introduced relatively recently in economics literature. It first appeared in a McKinsey report (1993) concerning Australian production enterprises. The report claimed that born globals have always perceived the world as a marketplace and the domestic market merely as support for their international activities (Rennie 1993). Australian born globals include technologically advanced firms as well as typical enterprises using well-known technological solutions in their daily work. According to Rasmussen and Madsen (2002), the primary factor in explaining the born global phenomenon is management engagement in the internationalisation process of Australian firms as well as their ability to standardise their product and marketing solutions and concentrate on market niches. Another characteristic trait of these firms was achieving a fast rate of growth compared to other firms in Australia and a high share of production intended for export relative to domestic sales. Despite the traits of born globals being mentioned in the McKinsey report (1993), the first working definition of this concept in an academic publication was presented by Knight and Cavusgil (1996), who described born-global firms as small and technologically oriented entities operating on international markets from their inception[1]. According to these authors, born-global firms usually operate in the small business sector. They employ fewer than 500 employees, earn sales revenues of less than USD 100 million, and use *cutting-edge technologies* to develop a relatively unique product or innovation process.

As time passed and as discussion continued in the academic community about this phenomenon and further research was pursued[2], new definitions appeared. Madsen and Servais (1997), for instance, in an attempt to find the most accurate definition of born globals, indicated seven attributes of such firms: 1) firms launched by strong entrepreneurs with strong international experience and perhaps, in addition, a strong product, 2) the extension of the born global phenomenon is positively associated with the degree of market internationalisation, 3) in comparison with other export firms, born globals are more specialised and niche-oriented, 4) the geographical location of activities in born globals is determined by the past experience of founders and partners as well as economic and capability or customer-related factors, 5) in comparison with other export firms, born globals are more likely to

[1] In later publications, the authors defined born globals as business organisations which, from the outset, achieved a significant competitive advantage from the use of existing resources and product sales in many countries (Cavusgil & Knight 2009).

[2] Moen (2002), for example, identified born globals on the basis of research carried out among French and Norwegian exporters, claiming that the differences between born globals and other enterprises were based on competitive advantage, a global orientation, export strategy, and market conditions.

rely on supplemental resources provided by other firms; in their distribution channels, they rely more often on hybrid structures, 6) the growth of a born global is positively associated with high innovative skills, including an ability to access effective R&D as well as distributions channels, often in partnerships involving close collaboration via international relationships, 7) firms in countries with small domestic markets have a higher propensity to become born globals than those in countries with large domestic markets. Moreover, countries with a high number of immigrants may have a higher proportion of born globals.

Differentiating born globals from other enterprises was the focus adopted by Cavusgil and Knight (2009). These authors identified the following differentiators in the case of born globals:

– born globals are characterised by extensive activity on many foreign markets, immediately or shortly after they are founded – most of these enterprises enter foreign markets via exports, which represents their major path to internationalisation. Born globals export their products or services a few years after their establishment and may export one fourth or more of their production;

– they have at their disposal limited financial and material resources – born globals are young and relatively small firms. Being small enterprises, they dispose of considerably fewer financial, human, and material resources in comparison with large international enterprises that have functioned for a long time on the market and dominate in global trade and investments;

– they engage in activities in many industries – many researchers suggest that born globals concentrate exclusively on high-tech industries. However, there are also many studies showing that these firms are also created in traditional branches of the economy (e.g. the metal, furniture or grocery industries);

– managers in these industries demonstrate a strong international outlook as well as entrepreneurship directed to the international market – many born-global firms are created by active managers with strong entrepreneurial attitudes. An entrepreneurial approach on the part of management manifests in a dynamic and aggressive attitude, which is the driving force for acquiring new foreign markets. This attitude is associated with a specific vision in terms of management, risk appetite, and competition;

– born globals use a differentiation strategy – many born globals fill a market niche by supplying products of varying quality oriented to a specific client segment, the production of which large enterprises show little interest in. The use of a differentiation strategy inspires consumer loyalty to the firm because it supplies them unique products adapted to their needs. Consumers as well as firms that communicate demand for certain specialised products create an important opportunity for the growth of small and medium-sized enterprises;

– they emphasise the highest product quality – many born globals supply products of more efficient construction and higher quality compared to their competitors. Born globals are often leaders in terms of technology in the industry. They target niches and supply unique products of the highest quality. Indeed, the founding of born-global firms is often associated with the growth of new products or services;

– they use a wide range of advanced communication and IT technologies to communicate with partners and consumers around the world at a marginal cost that is almost equal to zero. Advances in IT technologies in principle eliminate borders and expand the scope of business activities around the world;

– born globals usually take advantage of intermediaries when distributing goods on foreign markets – firms engage in direct sales of products on a foreign market (export) or use independent intermediaries for this purpose. They also often use external solutions (forwarders) that frequently organise foreign shipments. Export and the use of services provided by independent intermediaries means that born globals conduct their activities on foreign markets in a way that is very flexible. They can enter and withdraw from foreign markets quickly and relatively easily. Born globals with greater experience combine exports with other forms of foreign expansion such as joint ventures or direct foreign investments. The low cost and risk-free nature of exports makes it the most suitable form for small and young firms to enter foreign markets.

Cavusgil and Knight (2009) also enumerated five factors that foster the creation of born globals. These include: 1) market globalisation, 2) level of advancement in the field of communication and information technologies, 3) level of advancement in the field of technological production, 4) the existence of market niches, 5) the occurrence of global network connections. Moreover, the authors listed seven determining factors in the launch of an internationalisation process: 1) a mechanism that encourages exports in the form of external factors (export pull), 2) a mechanism that drives exports via the operation of an external agent (export push), 3) a monopolistic position on a global scale, 4) the product and market conditions necessary to undertake international engagement, 5) an advantage in terms of the product offered, 6) global network connections, 7) global market niches.

Hollensen (2017), in turn, defined a born-global firm as one that from its inception pursues a vision of becoming global and globalises rapidly without any preceding long-term domestic or internationalisation period. According to this author, born globals represent a type of firm that operates at the intersection (compression) of time and space, which allows them to achieve global reach from the very beginning. This “time-space” compression phenomenon means that geographic processes can be reduced and limited to the “here and now” of information exchange and trade around the world as long as the necessary infrastructure, communication, and IT facilities are introduced along with qualified personnel. The most significant distinguishing trait of born globals is that they strive to be managed by the most entrepreneurial visionaries, who perceive the world as one, without any boundaries, from the firm’s inception. Similarly, Chetty and Campbell-Hunt (2004) associate born globals with a group of firms that “from inception” treat the world as one big global market.

Andersson, Danilovic and Huang (2015), however, draw attention to the fact that research into born globals concerns first and foremost developed countries located in North America and Europe as well as enterprises in Australia and to a lesser extent emerging markets. In searching for the factors that are decisive in the success of this type of entity, they claim that despite the differences identified, there are also common elements such as: a specific number of entrepreneurs with international vision, experience, knowledge, level of education, cognitive abilities and processes, local and international networks, financial conditions, innovative culture, unique resources, level of market competition, strategies for entering foreign markets, geographical location, and government policy. In the case of less developed countries, the last of the above-mentioned factors has particular significance, namely, an active state policy supporting the activity of young exporters, including tax relief and assistance in organising promotions (e.g. exhibitions) abroad, which can accelerate the internationalisation process of domestic firms.

The level of development of native clusters is also important because they are a hotbed of entrepreneurship and innovation, and such an environment favours the emergence of firms implementing a global strategy of operation from the beginning (Andersson, Danilovic & Huang 2015). The developed model contains the essential success factors of born globals in the process of internationalisation. They have been classified and presented from four different perspectives: entrepreneurship, organisational, implemented strategy, and external environment (Figure 1).

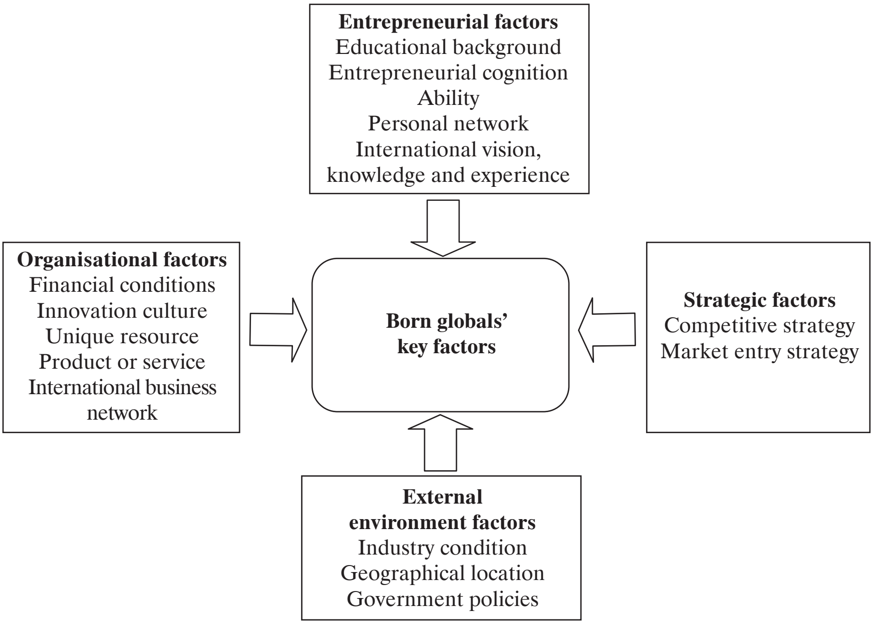


Fig. 1. Success Factors of Born Globals in the Process of Internationalisation

Source: Andersson, Danilovic & Huang (2015).

Of the listed factors, Andersson, Danilovic and Huang (2015) assign particular importance to a culture of innovation. The authors believe that without a unique, innovative product, resulting from the use of unique knowledge, it will be difficult for born-global firms to achieve success on the international market.

Issues related to born globals are also discussed in the Polish literature on the subject (Pietrasieński 2005, 2014, Nowiński 2006, Morawczyński 2007, 2008, Przybylska 2010, 2013, Duliniec 2011, Jarosiński 2012, Oczkowska 2013, Wach 2015, Limański & Drabik 2017). The definition of born globals proposed by Pietrasieński (2005), for example, referred to enterprises that begin to internationalise their activity from the very beginning of, or shortly after, their inception. He pointed out that in most cases the issue is the movement of products, and less often production factors (processes), to a large number of countries simultaneously or within a short time period. According to Morawczyński (2008), born globals are firms that adopt an international or global strategy after their establishment. Jarosiński (2012) claimed that the term born globals concerns only those enterprises that begin the internationalisation of their activity within three years of inception and earn at least 25% of their revenues from international markets. Limański and Drabik (2017), in turn, emphasised that the international competitive advantage of born globals is based on innovative and flexible activities (associated mainly with new technologies, but not excluding the use of traditional technologies) and offering unique products, despite having generally limited resources.

3. Born Globals in the Polish Economy – the Scale of the Phenomenon and Determinant Factors

During the centrally planned economy, the internationalisation of enterprises in Poland was a sporadic phenomenon, reserved primarily for large state-owned enterprises. Before 1989, internationalisation was primarily limited to exports to neighbouring countries with similar ideologies. The processes associated with the transformation of the socio-economic system followed by accession to the European Union have led Polish enterprises to gradually include expansion onto foreign markets in their growth strategies. It should be noted, however, that since Poland is a large country with a large market compared to other Central and Eastern European countries, the pressure on foreign expansion has been considerably lower.

Research into the operation and growth of born-global firms in Poland is carried out by both individuals and certain institutions. One of the first analyses for Poland was undertaken by Morawczyński. In 2004, he conducted a survey among 117 small and medium-sized enterprises operating in the Małopolska province. Based on these responses, he claimed that the pace of internationalisation among SMEs in the Małopolska province was faster than the theory of gradual internationalisation envisaged. More than 25% of firms began exporting within a year of inception. Half of the firms he researched entered a foreign market no later than three years from inception, which means that early internationalisation was observed in the case of approximately 50 of the enterprises included in the research. The research also indicated a relationship between a firm's period of operation and the share of exports in its total sales revenues: in enterprises existing on the market for up to five years, this indicator stood at 55%, while in the case of firms with thirteen or more years of experience, it stood at 75% (Morawczyński 2007, 2008).

Other studies on Polish born globals were carried out between 2007 and 2010 by a research team under the direction of Cieślik (2010). A total of

18,896 Polish exporters operating from 1994 to 2003 were included in the research. The study led the researchers to conclude that the phenomenon of early internationalisation among these firms was quite common. As many as 75% began exporting within three years of registration. The researchers distinguished three groups of enterprises: immediate exporters, which began exporting at inception or within the first full year of operation; rapid exporters, which began exporting products during the second or third year after establishment; and delayed exporters, which launched export activities after four years or more. The analysis carried out by Cieślik's team indicated that 46.1% of exporters could be described as immediate exporters, 28.8% as rapid exporters, and only 25.1% as delayed exporters.

Research conducted in 2008 by Kraśnicka (*Przedsiębiorczość*... 2008) also indicates the existence of born-global firms in Poland. One hundred small and medium-sized enterprises operating in the Silesia province were randomly selected for the study. Almost one third of them were already operating on the international market. However, in the case of 12% of these enterprises, the internationalisation process began from the moment of their inception.

Empirical research aimed at confirming the existence of born globals in Poland was also carried out by Przybylska (2010). The author surveyed 53 firms, 18 of which (34%) fulfilled the criteria of being a global enterprise from inception, in other words, firms that were established after 1989, employed fewer than 249 people, entered foreign markets within three years of launching business activity, and gained at least 30% of their revenues from exports.

Similar results were obtained by Nowiński and Nowara (2011), who attempted to identify the born globals in Poland. The research sample included 50 small and medium-sized private Polish firms that are engaged in exports. The research showed that the period required for a firm to decide on foreign expansion was 3.9 years on average. Each firm served six foreign markets on average, and most exported their goods to the European Union. The authors classified 30% of the firms surveyed as born globals or enterprises that within three years of inception earn at least 25% of their revenue from export sales.

In 2010, Jarosiński (2012) also carried out research on this subject. Among the 47 small and medium-sized enterprises included in the study, the author found that 32% satisfied the criteria of born globals. Most born-global firms pursued internationalisation rapidly: 73% of firms within the first year of operations, 13% within the second year, and another 13%

within the third year. The vast majority (close to 90%) operated only on European markets. On average, a firm operated on 3.8 markets (a median of 3). In conclusion, the author pointed out that although the researched enterprises met the definition of born global with respect to operational criteria, their characteristics diverged from the theoretical definitions.

Danik, Kowalik and Král (2016) also analysed born globals. For this research they selected 233 SMEs of which 45% were found to satisfy the criteria of born-global firms. Among Polish born globals, the main areas of activity were: production of food, metal goods, machinery and tools, rubber and artificial fibre goods as well as furniture. The research revealed that the dominant means of entry onto foreign markets was direct exports (as high as 98.1% of responses) and that most of the revenue (70.5%) earned by Polish born globals originated from exports of goods to EU markets.

In addition to individuals and research groups, the Polish Agency for Enterprise Development (PARP) has also carried out research into the operation of born-global firms in Poland. Among enterprises that declared an engagement in international activity, more than half (51.8%) of SMEs stated that they had been operating on foreign markets since the start of their business activity. This also applies to almost half of small and medium-sized exporters (49.8%). Quantitative research carried out by PARP only among SMEs, in turn, indicated that more than 33% of the firms surveyed (35%) in the sectors that dominate exports were present on foreign markets since their inception. These results led the researchers to claim that the scale of operation of born-global firms in Poland ranges between 35 and 50% of SMEs engaged in international activity, which amounts to an estimated 58,000 to 83,000 firms (*Ewaluacja*... 2014).

Research carried out by PARP has also shown that the decision by small and medium-sized Polish enterprises to internationalise is motivated by various factors. These can be divided into two groups: internal (related to the firm) and external (related to the environment). The major internal factors that influenced the decision to expand onto foreign markets included: the prospect of cooperation with a foreign partner, the possibility to sell products abroad at a higher price than within the domestic country, the desire to avoid dependence on domestic sales, ownership-based relationships with a foreign contractor, and activities to improve the firm's image. In the case of enterprises engaged in exports, these factors were mentioned by 84.4% of entities. Therefore, we can conclude that the most important internal determinant in the internationalisation of Polish enterprises belonging to the SME sector was a strategic drive by firms to increase profitability and achieve market diversification, which, as shown by PARP research, translates into better economic results compared to companies operating only on the domestic market. In the case of external determinants, the most important factors among the researched enterprises were related to the foreign market: high demand for products and fewer administrative and regulatory restrictions; however, with respect to the domestic market the most important factors were high competition and an under-developed market (*Ewaluacja*... 2014).

The quantitative research carried out by PARP is also supported by qualitative studies. These showed that the owner's experiences on foreign markets played a significant role in the decision to expand onto foreign markets. Internationalisation was also fostered by factors such as: knowledge of foreign languages and cultures as well as personal contacts on the part of management. Exporters from the SME sector believed that the most important advantages allowing them to place products on foreign markets included the high quality and competitive prices of their goods. Exporters assigned less importance to factors such as the unique character, modernity or design of their goods.

The results obtained by PARP were in line with those obtained by other authors. For example, Kowalik and Baranowska-Prokop (2013), in analysing the determinants of the decision and motives behind the expansion of small and medium-sized Polish enterprises that internationalised rapidly, identified the following main external determinants: the owner's personality, unique knowledge of the foreign market or experience of the board in international business. Other characteristics enumerated in the research included: a network of personal contacts, sales motivations and the firm's organisational culture, a unique advantage in the field of technology, and the board's focus on the development of human resources. In turn, the most important determinants of an internal nature included: the appearance of new business opportunities at the moment of Poland's accession to the EU, the chance to access a network of suppliers and partners among international corporations, geographical proximity, the impact of systemic transformation, greater market potential, and higher prices abroad. The research indicated that the personal characteristics of their founders played a key role in the launch and internationalisation of small and medium--sized Polish firms. Global vision on the part of management from the firm's inception and the use of personal and professional relationships with foreign partners distinguishes born-global firms from traditional exporters.

4. Conclusions

The results presented above, obtained by certain authors, indicate that the born global phenomenon is also visible in Poland. Enterprises undergoing early internationalisation began to appear in the Polish economy as early as the beginning of the 1990s. This is the same moment that other firms of the same type were developing in other countries. However, research into this phenomenon began considerably later in Poland.

Based on a review of selected analyses of the activities of Polish born-global firms, it can be concluded that there are SMEs that actively participate in the process of early internationalisation. A review of the research indicates that the percentage of these types of companies ranges from 30 to 50% of the population. However, these results should be treated with caution due to the different definitions of born globals adopted in individual studies and their methodologies. It is worth pointing out that the population of Polish born globals reflects the structure of enterprises in Poland – the number of enterprises from high technology industries on international markets is small. The research results point to the need for the state to be more involved in supporting the early internationalisation of enterprises.

It is apparent that Polish managers have the right entrepreneurial traits and strategic vision of enterprise development, which determine the formation of born globals; however, there are other factors that hinder this process. The main barriers hindering the expansion of Polish enterprises onto foreign markets include a lack of financial and human resources, the low prices foreign contractors want to pay for goods produced by Polish SMEs, the low recognition of Polish products and services abroad, a lack of free production capacity allowing for increases in export production, and the high expectations of foreign customers regarding quality, environmental standards innovation, and the design of products for export (Chilimoniuk-Przeździecka & Klimczak 2014).

In addition, weak pressure on the early internationalisation of enterprises may result from a belief that being a born global does not directly or necessarily translate into higher profits or improved business performance. Having a broad global presence also makes born globals particularly vulnerable to financial losses, and these entities suffer disproportionately from internationalisation (Oyson 2018).

The role of born-global firms in the Polish economy requires further research. Existing quantitative analyses of born globals do not fully enable the factors constituting the main barriers to the development of this sector in Poland to be identified. Qualitative studies are also necessary, especially case studies. This is necessary in order to plan the appropriate support policies for SMEs on international markets.

Bibliography

Andersson, S., Danilovic, M. and Huang, H. J. (2015) "Success Factors in Western and Chinese Born Global Companies". *iBusiness* 7(1): 25–38, https://doi.org/10.4236/ib.2015.71004.

Cavusgil, S. T. and Knight, G. (2009) *Born Global Firms. A New International Enterprise.* New York: Business Expert Press.

Chetty, S. and Campbell-Hunt, C. (2004) "A Strategic Approach to Internationalization: A Traditional versus a «Born-Global» Approach". *Journal of International Marketing* 12(1): 57–81, https://doi.org/10.1509/jimk.12.1.57.25651.

Chilimoniuk-Przeździecka, E. and Klimczak, T. (2014) *Umiędzynarodowienie polskich przedsiębiorstw – wyniki badania PARP.* Warszawa, http://porpw.parp.gov.pl/files/74/75/76/21621.pdf (accessed: 7 July 2018).

Cieślik, J. (2010) *Internacjonalizacja polskich przedsiębiorstw. Aktualne tendencje – implikacje dla polityki gospodarczej.* Warszawa: Akademia Leona Koźmińskiego.

Danik, L., Kowalik, I. and Král, P. (2016) "A Comparative Analysis of Polish and Czech International New Ventures". *Central European Business Review* 5(2): 57–73, https://doi.org/10.18267/j.cebr.152.

Duliniec, E. (2011) "Koncepcje przedsiębiorstw wcześnie umiędzynarodowionych. Rozważania terminologiczne". *Gospodarka Narodowa* 1–2: 63–80.

Ewaluacja potencjału eksportowego w Polsce (2014) Warszawa: PARP.

Gurău, C. (2007) *The „Born Global Purchasers": Internationalisation through Outsourcing in the UK Biotechnology Sector.* Oxford: Oxford Business & Economics Conference, 24–27 June.

Gustafsson, T. and Zasada, D. (2011) *Explanatory Power of Internationalization Theories: a Case Study of Norwegian SMEs Targeting Developing Countries.* Oslo: University of Oslo.

Halldin, T. (2012) "Born Global Firms – Do They Perform Differently?". *CESIS Electronic Working Paper Series* 269.

Hollensen, S. (2017) *Global Marketing.* 7th ed. Edinburgh Gate: Pearson Education Limited.

Jarosiński, M. (2012) "Early Internationalization of Polish Companies – the Results of Research". *Journal of Management and Financial Sciences* 5(7): 34–49.

Johanson, J. and Vahlne, J. E. (1977) "The Internationalization Process of the Firm – A Model of Knowledge Development and Increasing Foreign Market Commitments". *Journal of International Business Studies* 8(1): 23–32.

Johanson, J. and Wiedersheim-Paul, F. (1975) "The Internationalization of the Firm – Four Swedish Cases". *Journal of Management Studies* 1(3): 305–323, https://doi.org/10.1111/j.1467-6486.1975.tb.00514.x.

Knight, G. A. and Cavusgil, S. T. (1996) "The Born Global Firm: A Challenge to Traditional Internationalization Theory". *Advanced in International Marketing*, vol. 8: 11–26.

Kowalik, I. and Baranowska-Prokop, E. (2013) “Determinanty powstania i motywy ekspansji przedsiębiorstw wcześnie umiędzynarodowionych”. *Gospodarka Narodowa* 4: 41–64.

Limański, A. and Drabik, I. (2017) “Koncepcja born globals a etapowa internacjonalizacja przedsiębiorstwa”. *Zeszyty Naukowe Politechniki Poznańskiej* 73: 181–195, https://doi.org/10.21008/j.0239-9415.2017.074.12.

Madsen, T. K. and Servais, P. (1997) “The Internationalization of Born Blobals: An Evolutionary Process?”. *International Business Review* 6 (6): 561–583, https://doi.org/10.1016/S0969-5931(97)00032-2.

McKinsey (1993) *Emerging Exporters. Australia’s High Value-Added Manufacturing Exporters (1993).* McKinsey & Company. Melbourne: Australian Manufacturing Council.

Moen, Ø. (2002) “The Born Globals. A New Generation of Small European Exporters”. *International Marketing Review* 19 (2): 156–175, https://doi.org/10.1108/02651330210425015.

Morawczyński, R. (2007) “Zachowania małopolskich małych i średnich przedsiębiorstw eksportujących w świetle teorii stopniowego umiędzynarodowienia”. *Gospodarka Narodowa* 7–8: 65–82.

Morawczyński, R. (2008) “Born Globals and Traditionally Internationalised SMEs in the Polish Economy. Some Empirical Findings from Comparative Analysis” in: R. Borowiecki, P. Jaki (eds) *Enterprises in the Face of 21st Century Challenges. Development – Management – Entrepreneurship.* Cracow: Cracow University of Economics: 451–463.

Nowiński, W. (2006) “Wczesna internacjonalizacja przedsiębiorstw – nowy obszar badań”. *Gospodarka Narodowa* 11–12: 87–99.

Nowiński, W. and Nowara, W. (2011) “Stopień i uwarunkowania internacjonalizacji małych i średnich przedsiębiorstw”. *Gospodarka Narodowa* 3: 29–45.

Oczkowska, R. (2013) “Wczesna internacjonalizacja małych i średnich przedsiębiorstw”. *Zarządzanie Zasobami Ludzkimi* 3–4: 77–87.

Oyson, M. J. (2018) “The Rise of Small Globals: an Empirical Macro Analysis”. *International Journal of Globalisation and Small Business* 10 (1): 21–48, https://doi.org/10.1504/IJGSB.2018.091824.

Pietrasieński, P. (2005) *Międzynarodowe strategie marketingowe.* Warszawa: PWE.

Pietrasieński, P. (2014) *Aktywizowanie internacjonalizacji przedsiębiorstw.* Warszawa: PWE.

“Przedsiębiorczość międzynarodowa. Aspekty teoretyczne i praktyczne” (2008) ed. T. Kraśnicka. *Prace Naukowe Akademii Ekonomicznej w Katowicach.* Katowice.

Przybylska, K. (2010) “Born global – nowa generacja małych polskich przedsiębiorstw”. *Gospodarka Narodowa* 7–8: 63–84.

Przybylska, K. (2013) *Born global. Nowa generacja małych przedsiębiorstw.* Kraków: Fundacja Uniwersytetu Ekonomicznego w Krakowie.

Rasmussen, E. S. and Madsen, T. K. (2002) *The Born Global Concept*, 28th EIBA Conference, Special session „SME Internationalization and Born Globals – Different European Views and Evidence”. Athens.

Rennie, M. W. (1993) “Born Global”. *The McKinsey Quarterly* 4: 45–52.

Wach, K. (2015) “Incremental versus Rapid Internationalisation of Firms: Results of Exploratory Investigation form Poland”. *Entrepreneurial Business and Economics Review* 3 (4): 37–48, https://doi.org/10.15678/EBER.2015.030403.

ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 1 (22) • 2020
ISSN 1642-168X
e-ISSN 2545-3866
AOC, 2020; 1 (22): 73–88
https://doi.org/10.15678/AOC.2020.2206

Anna Magdalena Gierusz

A COMPARISON OF RISK-SHARING APPROACHES IN HYBRID OCCUPATIONAL PENSION SCHEMES

Abstract

Objective: Hybrid pension schemes provide a mixture of features of DB and DC schemes. They allow for the risk to be shared between employer and member. The aim of this article is to present the risk sharing between employer and member within selected forms of hybrid pension schemes – cash balance and self-annuitizing schemes – and in a proposed type of hybrid scheme.
Research Design & Methods: The variability of contributions required to provide a fixed level of benefit is chosen as a measure of risk within the schemes. Investment and longevity risk is introduced via changes in the investment rate of return and life table probabilities used to price annuities. The variability of member and employer contributions required in each scheme is compared.
Findings: In cash balance and self-annuitizing schemes risk sharing is achieved by allocating a given type of risk (investment or longevity risk) to either the employer or the member. In the proposed scheme, risk is shared irrespective of its type. This allows for better financial planning for the two parties involved by setting a limit on the employer's contributions and requiring an adjustment to the member's contributions only in certain instances.
Implications/Recommendations: A hybrid scheme which allows for risk to be shared irrespective of its type should be considered. Additional safeguards, such as setting an upper limit for contributions, should be employed.
Contribution: Further development of hybrid pension schemes and a comparison of the proposed solution with existing ones.

Keywords: occupational pension schemes, risk sharing, hybrid schemes, contributions.
JEL Classification: G22, J26.

Anna Magdalena Gierusz, University of Gdansk, Faculty of Management, Armii Krajowej 101, 81-824 Sopot, e-mail: anna.gierusz@ug.edu.pl, ORCID: https://orcid.org/0000-0002-6548-5016.

1. Introduction

The aim of occupational pension schemes is to allow members (employees of a sponsoring employer) to provide extra income for retirement. The actual cost of benefit provision is unknown (it is only known after all the benefits have been paid out or bought out through an insurance company), so there is a risk that the funds accumulated for the payment of retirement benefits could be inadequate. While there are several types of risk affecting retirement benefit provision (Blake 2006, p. 174; Turner 2014, pp. 5–6; Cooper 2005, pp. 6–7), this article focuses on two types of risk which have a big impact on the funding of a pension scheme: investment risk and longevity risk.

Investment risk is the risk that the rate of return which the scheme earns on its investments is higher or lower than expected. Longevity risk is the risk that a member's future life in retirement (which determines how long the benefits will be paid for) will be shorter or longer than assumed. As a result, the funds accumulated for the payment of retirement benefits may be lower or higher than required.

Defined contribution (DC) and defined benefit (DB) schemes are the two main forms of occupational pension scheme (Pugh & Yermo 2008, p. 6). In a DC scheme, the benefit amount is unknown and depends on contributions paid and the rate of the scheme's investment returns. The member bears the investment and longevity risk, although they can protect themselves against the latter by purchasing a life annuity from an insurance company. In a DB scheme, the benefit amount is set in advance and the employer has the responsibility to fund the scheme in such a way that the promised benefits can be paid. As such, it is the employer who bears the risk in this scheme. DB schemes have recently been in decline, with some employers choosing to close their DB schemes in favour of a DC scheme in order to decrease the level of risk to which they are exposed (Petelczyc 2016, p. 62). This, however, leads to members bearing all the risk, despite often not having adequate means or knowledge to manage and protect themselves against that risk (Clark & Monk 2006, pp. 43–44; Davis 2013, p. 685; Sweeting 2007, p. 2).

Hybrid pension schemes are schemes that are neither fully DB or DC, but are a mixture of both. They allow for the risk to be shared between the employer and the member.

Occupational pension schemes are common in countries such as Switzerland and the Netherlands, where participation in these schemes is compulsory (nationwide or for certain industries or professions), with the participation rate reaching almost 90% (European Commission 2017, p. 7).

In countries where participation is voluntary, the participation rate is usually lower, for example around 55% in the US (Bureau of Labor Statistics 2018). In Poland, only 2.6% of employees participated in an occupational pension plan in 2018 (KNF 2019). However, this figure can change after the introduction of new employee capital plans with automatic enrolment.

The aim of this article is, firstly, to illustrate risk sharing within selected forms of hybrid pension schemes, and then to illustrate it in a scheme which takes a different approach to risk sharing between employer and member. To this end, the variability of the employer's and member's contribution needed to achieve a target level of benefit is investigated. The effectiveness of occupational pension schemes could be increased in this way.

2. Risk Measures in Pension Schemes

Risk in a pension scheme can be measured in several ways. From the member's point of view, an important measure is the variability in benefit amount which can be received from the scheme (Blommestein *et al.* 2009, Cooper 2005, Davis 2013, Davis & Madland 2013). In a DB scheme, the member is guaranteed to receive a pension equal to a proportion of their salary, hence there is no variability of benefit amount. By contrast, in a DC scheme, the benefit amount is not known until retirement, as it depends on the amount of contributions paid, the rate of investment returns which the scheme has earned net of costs, and annuity conversion rates.

Another measure used to illustrate risk in pension schemes is the variability of the funding level, i.e. the ratio of the scheme's assets to its liabilities (Blommestein *et al.* 2009). This measure applies to DB schemes and schemes with some benefit guarantee. Depending on the financial performance of its assets, such a scheme may be under- or over funded. By contrast, a DC scheme is always fully funded. This is an important measure for the employer, as any changes in the funding level will need to be rectified, typically via an increase in the employer's contributions. From the member's point of view, this measure shows the security of benefits, as a very low funding level may indicate financial difficulty and the possibility of benefits not being paid.

The third measure, and the one that will be used in this article, is the variability of contributions needed to provide a fixed level of benefit (see also Gierusz 2019). The contributions of employer and member will be considered separately in order to illustrate risk sharing within a scheme. In a DB scheme, a member's contributions are fixed, whereas the employer's contributions are adjusted to ensure that funding remains at the required level. In a DC scheme, the situation is the opposite: the employer's contributions are fixed and the member's contributions may need to be adjusted in order to provide the required level of benefit. In a hybrid scheme, either the employer's or the member's contributions can change depending on who bears the given type of risk.

3. Hybrid Schemes Considered

For the purpose of this article, two forms of hybrid scheme have been chosen due to their risk-sharing characteristics: the cash balance scheme and the self-annuitising scheme.

In a cash balance scheme, a member's account is credited every year with a certain fixed percentage of their salary (a so-called salary or pay credit). Every year, the amount accumulated within the account is increased by a fixed interest rate guaranteed by the employer (interest credit). At the point of retirement, the member is entitled to the amount accumulated within the account, which can be used to purchase a life annuity from an insurance company (Mackenzie 2010, pp. 4–5; Szczepański & Brzęczek 2016, p. 114; Takayama 2013, p. 11). This scheme resembles a DB scheme during the pre-retirement phase, since the interest rate used to accumulate funds is guaranteed by the employer and is independent of the scheme's actual investment returns, hence the employer bears the investment risk. However, after retirement the risk is passed on to the member, who has to bear the longevity risk or purchase an annuity from an insurer.

The second hybrid scheme considered is a self-annuitising scheme. Before retirement this scheme resembles a DC scheme – contributions are fixed and accumulated at the actual rate of scheme's investment returns. However, at the point of retirement the amount accumulated within the member's account is used to "buy" a life benefit within the scheme, according to a fixed, pre-determined annuity conversion rate (Wesbroom & Reay 2005, pp. 13–14). Hence, the member bears the investment risk, and the employer bears the longevity risk.

4. Modelling Assumptions

4.1. General Remarks

All the schemes under consideration (DB, DC, cash balance and self-annuitising schemes) provide or aim to provide the same level of benefit.

A replacement rate of 60% of final salary was chosen as that target. This was based on the World Bank recommendation of a replacement rate equal to 50% of final salary (World Bank 1994, p. 294) and a replacement rate of around 75–95% as proposed by Palmer (2008, p. 24). Within each scheme there is only one member, who joins it aged 25 and remains an active member of the scheme until retirement at age 65. Assumptions about the rate of the scheme's investment returns, salary increases, interest rates, and survival probabilities used to calculate annuity conversion rates were made, and an annual contribution rate required to achieve a 60% replacement rate was calculated (this is referred to as the base contribution rate). Next, investment and longevity risks were introduced by varying the rate of investment returns and survival probabilities in the life table. A corresponding increase or decrease in the contribution rate required to maintain the target replacement rate was calculated.

4.2. Base Contribution Rate

All the calculations were made in relation to real values and amounts. The real rate of the scheme's investment returns was set at 4% p.a. This was based on the long-term assumptions of J.P. Morgan (2018), i.e. a 5.25% nominal rate of return and 1.5% inflation in Europe. In reality, this assumption will vary from scheme to scheme and possibly from member to member depending on the chosen investment strategy, risk attitudes, and time to retirement, but for the purpose of this article it was assumed that all schemes follow the same investment strategy. The real rate of salary increase was set at 1.5% p.a. based on J.P. Morgan's assumption (2018) of real GDP growth in Europe. In order to calculate the annuity conversion rate, the real interest rate was set at 0.5% p.a., and survival probabilities from the 2017 unisex life tables published by the Central Statistical Office in Poland were used (GUS 2018).

The benefit, equal to 60% of final salary, takes the form of a life annuity payable yearly in arrears. Using an interest rate of 0.5% p.a. and the 2017 unisex life tables, an annuity conversion rate of 16.67 was derived. Thus, for every monetary unit of the required pension amount, 16.67 monetary units have to be accumulated at retirement age 65. It is assumed that this conversion rate is used by an insurer to price annuities. For the chosen financial assumptions this means an amount equal to approximately 10 times final salary needs to be accumulated within the scheme at retirement in order to purchase the target benefit.

The base contribution rate, assuming a rate of investment return of 4% p.a. and salary increases of 1.5% p.a., was calculated to be 15% of salary. This was divided equally between member and employer, i.e. each party contributes 7.5% of salary into the scheme every year. It is assumed that the contribution is paid annually at the end of the year.

4.3. Specific Scheme Assumptions

In a DC scheme, the target benefit of 60% of final salary is set and annual contribution rates of 7.5% of salary for both the employer and member are agreed. If the assumptions are borne out, an annuity of the desired amount will be purchased at retirement. If the actual experience differs from that assumed, member contributions will need to change in order to achieve the desired level of benefit (employer contributions are fixed at 7.5% of salary). If the investment returns are lower than expected, or increases in longevity cause the insurer to increase the price of the annuity, the member's contribution rate will need to increase; if investment returns are higher or longevity lower than expected, the member's contribution rate will decrease. This means that while bearing the investment risk and longevity risk (via the risk of a change in annuity prices), the member bears the upside as well as the downside of risk. In practice, it is more common for members to pay a fixed contribution rate and bear the risk through the variability of benefits, but for the purpose of this article it is assumed that the member chooses (and is allowed) to vary their contribution rate.

In a DB scheme, the employer promises the member a benefit of the target amount. The employer bears the investment risk and the longevity risk, just as the member does in the DC scheme described above. For the purpose of comparison, it is assumed that the employer will buy out the liability when the member retires, i.e. will purchase an annuity similarly to a member in a DC scheme, rather than pay out the pension from the scheme. The employer's contribution rate is calculated in the same way in which a member's contribution is calculated in a DC scheme[1]. Any impact of regulation, which could require the sponsor of a DB scheme to use more pessimistic or prudent assumptions to fund the scheme is not considered.

[1] Often in a DB scheme, the contribution rate is set every year (or every couple of years) based on the benefits accruing over the next year (couple of years) of service (GCAE 2001). In this article, in order to compare DB and hybrid schemes with a DC scheme, it is assumed that in all schemes the contributions are calculated every year as constant rates of salary over the time remaining to retirement, taking into account the projected value of contributions already accumulated.

In a cash balance scheme, 15% of salary is credited to the member's account at the end of every year. An interest credit is calculated using a real rate of 4% p.a. This means that a lump sum equal to 10 times the final salary is promised at retirement. If the investment returns differ from that assumed, the employer's contribution rate will need to be adjusted. However, it is the member who bears the longevity risk in this scheme. If annuity prices change (due to changes in longevity), a lump sum of 10 times final salary may not be enough to purchase the desired benefit amount. Hence the member's contribution will need to change. As the member's contribution rate in this scheme is fixed at 7.5% of salary, the extra contributions made by the member due to changes in longevity can be considered to be paid as additional voluntary contributions.

In the self-annuitising scheme, annuity conversion rates are fixed at 16.67. It is assumed that the employer will buy out the liability with the insurance company, hence any changes in annuity prices will cause the employer to adjust their contribution rate accordingly. The member bears the investment risk, so changes in investment returns will need to be rectified by variation in the member's contribution rate.

Finally, a new type of hybrid pension scheme, proposed by Gierusz (2019) is modelled. In this scheme, each member has an individual account into which contributions made by the employer and member are paid. At retirement, an annuity providing the desired benefit level for life is purchased. A target benefit level (60% of final salary) is chosen, and the base contribution rate required to achieve that benefit is calculated. Using the assumptions described above, this will be a contribution rate of 7.5% for the member and 7.5% for the employer. Every year, a change to the contribution rate, required to achieve the target benefit in the case of changes in investment returns or longevity, is calculated and divided between member and employer as follows: a change of up to 5 percentage points (p.p.) is paid by the employer, while anything required above 5 p.p. is paid by the member.

Next, four different scenarios are considered in which investment returns or survival probabilities change, causing a change in the required contribution rate.

5. Calculation Results

Scenario 1

The following model of future investment returns is considered: the rate of return every year is a random variable following normal distribution with set mean and variance. In scenario 1, this mean is equal to 4% and the standard deviation is equal to 2%. A random sample of 40 rates of return was drawn (one for each year of service in the scheme), and year by year a contribution rate required to achieve the target benefit in each of the considered schemes was calculated. Figures 1 and 2 illustrate the contribution rates payable in each of the considered schemes.

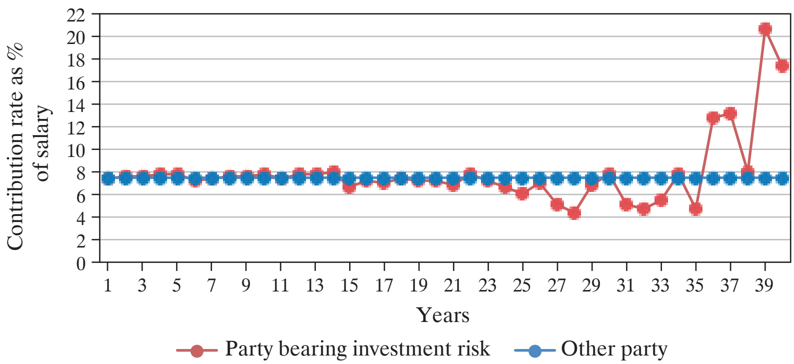


Fig. 1. Contribution Rates under Scenario 1 in a DC, DB, Cash Balance or Self-annuitising Scheme

Source: author's own work.

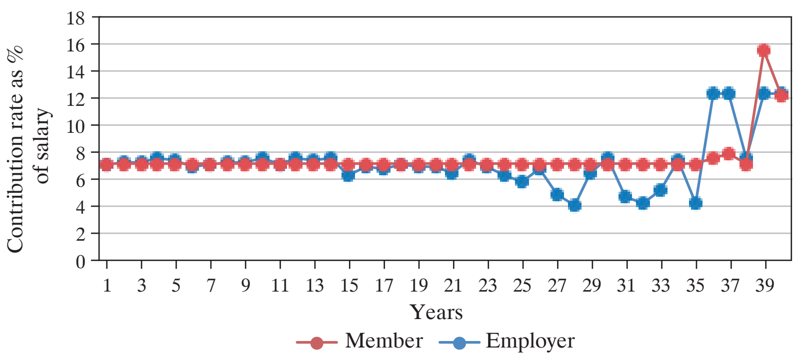


Fig. 2. Contribution Rates under Scenario 1 in the Proposed Scheme

Source: author's own work.

As shown by Figure 1, a change in investment return rates according to scenario 1 causes a fluctuation in the contribution rate of a party bearing the investment risk (a member in a DC or self-annuitising scheme, an employer in a DB or cash balance scheme). On average, they pay a contribution rate which differs by 1.5 p.p. from the base contribution rate. If the member chooses not to adjust their contribution rate in a DC or self-annuitising scheme, the benefit at retirement decreases slightly to 59% of final salary.

In the proposed hybrid scheme, as shown in Figure 2, the extra contribution is shared between employer and member. The employer's contribution rate is different from the base rate by 1.1 p.p. on average, the member's contribution rate by 0.4 p.p. The member needs to adjust their contribution in 10% of all years.

Scenario 2

In this scenario, a decrease in the real rate of the scheme's investment returns is considered. The rate of investment returns follows normal distribution with a mean of 2% and standard deviation of 2%. Figures 3 and 4 illustrate the contribution rates payable in each of the considered schemes.

As shown by Figure 3, a change in investment return rates according to scenario 2 causes an increase in the contribution rate of a party bearing the investment risk. In the case of a DC or self-annuitising scheme it is the member, while in the case of a DB or cash balance scheme it is the employer. They pay a contribution rate which on average differs by 11.3 p.p. from the base contribution rate. If the member chooses not to adjust their contribution rate in a DC or self-annuitising scheme, the benefit at retirement decreases to 34% of final salary.

In the proposed hybrid scheme, as shown in Figure 4, the extra contribution is shared between employer and member. The employer's contribution rate is different from the base rate by 3.9 p.p. on average, the member's contribution rate by 7.4 p.p. The member needs to adjust their contribution in 63% of all years.

Scenario 3

In this scenario, an increase in the real rate of the scheme's investment returns is considered. The rate of investment returns follows normal distribution with a mean of 5% and standard deviation of 2%. Figures 5 and 6 illustrate the contribution rates payable in each of the considered schemes.

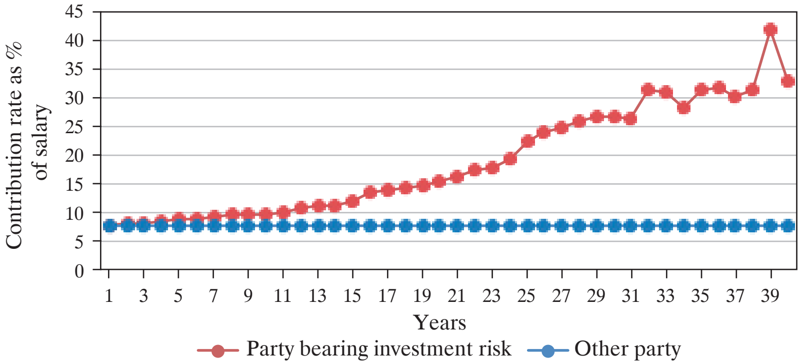


Fig. 3. Contribution Rates under Scenario 2 in a DC, DB, Cash Balance or Self-annuitising Scheme

Source: author's own work.

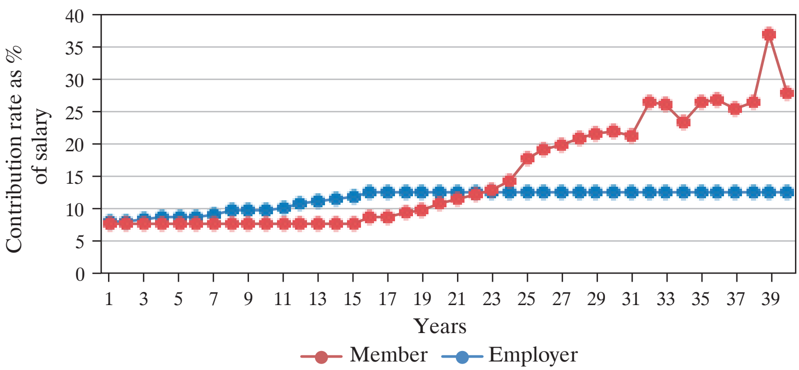


Fig. 4. Contribution Rates under Scenario 2 in the Proposed Scheme

Source: author's own work.

As shown by Figure 5, a change in investment return rates according to scenario 3 causes a decrease in the contribution rate of a party bearing the investment risk. In the case of a DC or self-annuitising scheme it is the member, while in the case of a DB or cash balance scheme it is the employer. On average, they pay a contribution rate different by 2.3 p.p. from the base contribution rate. If the member chooses not to adjust their contribution rate in a DC or self-annuitising scheme, the benefit at retirement increases to 68% of final salary.

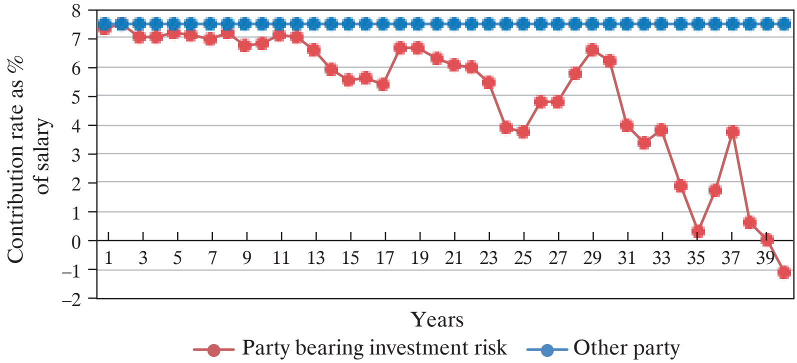


Fig. 5. Contribution Rates under Scenario 3 in a DC, DB, Cash Balance or Self-annuitising Scheme

Source: author's own work.

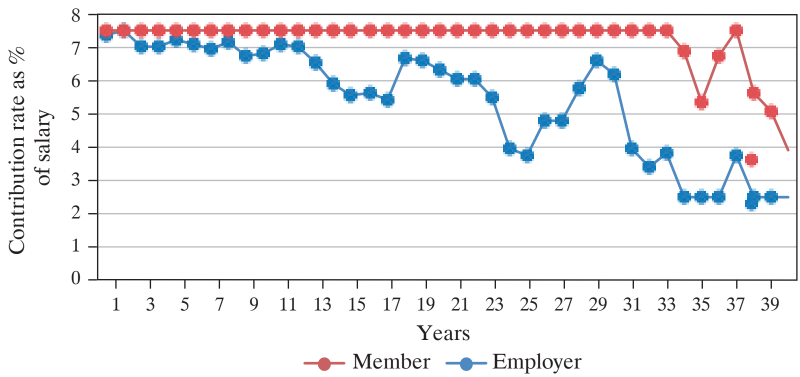


Fig. 6. Contribution Rates under Scenario 3 in the Proposed Scheme

Source: author's own work.

In the proposed hybrid scheme, as shown in Figure 6, the extra contribution is shared between employer and member. The employer's contribution rate is different from the base rate by 2.1 p.p. on average, the member's contribution rate by 0.3 p.p. The member needs to adjust their contribution in 15% of all years.

Scenario 4

In addition to changes in rates of return (according to Scenario 1), an increase in future longevity is considered. Survival rates change in a way which corresponds to rating the life table down by 5 years, i.e. the survival

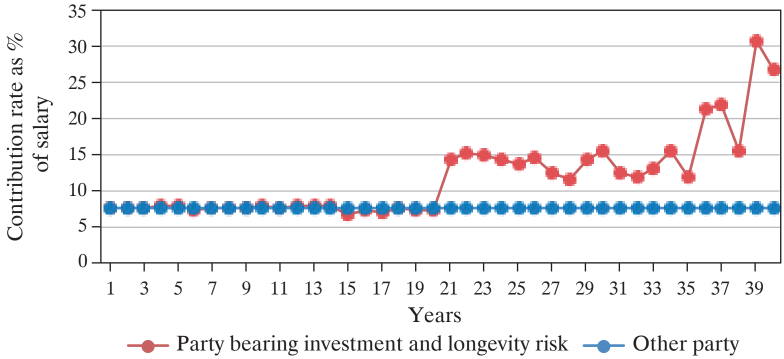


Fig. 7. Contribution Rates under Scenario 4 in a DC or DB Scheme
Source: author's own work.

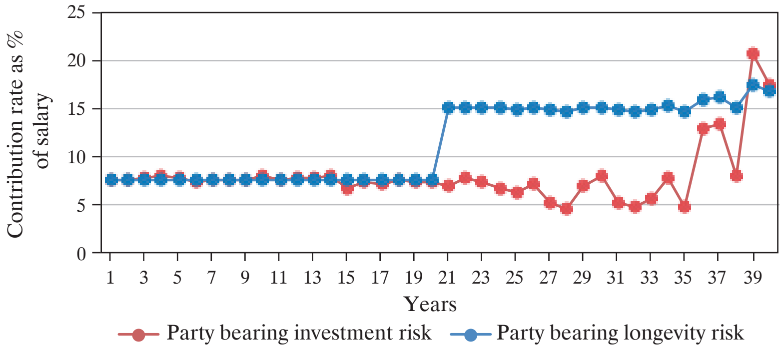


Fig. 8. Contribution Rates under Scenario 4 in a Cash Balance or Self-annuitising Scheme
Source: author's own work.

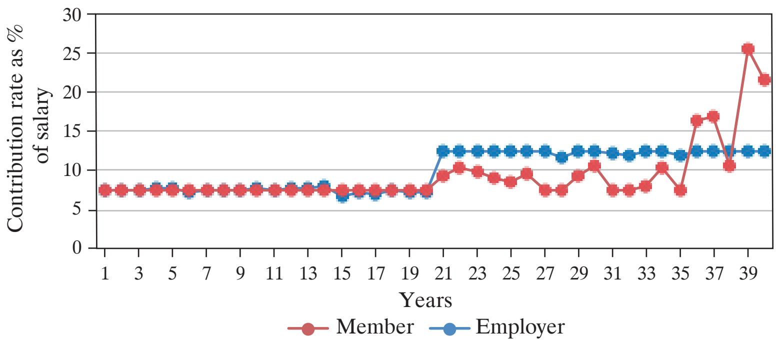


Fig. 9. Contribution Rates under Scenario 4 in the Proposed Scheme
Source: author's own work.

probabilities which applied to a person aged 60 in the original table now apply to a person aged 65. This means lower mortality rates and an increase in annuity conversion rates (due to increased longevity) from 16.67 to 19.86. This increase happens once, at the start of year 21. Figures 7, 8 and 9 illustrate the contribution rates payable in each of the considered schemes.

As shown by Figure 7, a change in conditions according to scenario 4 causes a fluctuation in the contribution rate of a party bearing the investment and longevity risk. In the case of a DC scheme it is the member, while in the case of a DB scheme it is the employer. On average, they pay a contribution rate different by 4.4 p.p. from the base contribution rate. If the member chooses not to adjust their contribution rate in a DC scheme, the benefit at retirement decreases to 50% of final salary.

Figure 8 shows that in the case of cash balance and self-annuitising schemes, the contribution of a party bearing the investment risk (the employer in a cash balance scheme, the member in a self-annuitising scheme) fluctuates. The party bearing the longevity risk (the member in a cash balance scheme, the employer in a self-annuitising scheme) has the contribution rate increased at the end of year 21. The rate then fluctuates slightly due to changes in the investment return rate (thus even though this party is said to bear only the longevity risk, the need to provide extra funds due to a change in longevity while investment conditions change means that they also bear some investment risk). On average, the party bearing the investment risk pays a contribution rate different by 1.5 p.p. from the base contribution rate. The party bearing the longevity risk pays a contribution rate different by 3.9 p.p. on average. If the member chooses not to adjust their contribution rate in a cash balance scheme, the benefit at retirement decreases to 50% of final salary. If the member chooses not to adjust their contribution rate in a self-annuitising scheme, the benefit at retirement decreases to 59% of final salary.

In the proposed hybrid scheme, as shown in Figure 9, the extra contribution is shared between employer and member. The employer's contribution rate is different from the base rate by 2.6 p.p. on average, the member's contribution rate by 1.8 p.p. The member needs to adjust their contribution in 40% of all years.

6. Practical Considerations

There are several issues that need to be considered if such a theoretical concept for a scheme was to be implemented in practice. Firstly, members of a pension scheme tend to appreciate a fixed, known level of benefits, but may be unwilling to increase their contribution rates. In many countries participation in occupational pension schemes is voluntary, hence members cannot be forced to adjust their contributions. In such cases the scheme setup can be adjusted so that while the employer needs to adjust their contribution rate as described above, the member is not required to do so. As a result, the benefit amount achieved in the scheme will be variable, although to a lesser extent than in a DC scheme due to adjustments in the employer's contribution.

Communication with members in such a scheme would be crucial. Members might not understand or be willing to accept the fact that the employer's contribution rate can decrease. In practice it is more likely that instead of decreasing their contribution rate, employers will allow the surplus to accumulate as a buffer for when circumstances are adverse.

In the proposed scheme, the contribution rate is recalculated annually. This should be done by an independent party (e.g. an actuary) so that there is no moral hazard of the employer attempting to influence the contribution rate. However, if the actuarial fees are covered by the employer, this can in turn lead to moral hazard for the actuary. A regulator may need to be involved to provide guidance and oversee the process of setting the contribution rate. Annual recalculation of the contribution rate can be costly, especially if there are many members in the scheme. It can be changed to a bi- or triannual recalculation, but the variability of the contribution rate could increase as a result.

Under the considered model, it is possible for contribution rates to be negative. In practice this would mean funds being returned to the employer and/or member. Such payments may be forbidden or subject to an unfavourable tax treatment. It is therefore possible that instead of funds being returned, a member will opt for the funds to remain in the scheme and increase the benefit amount. The issue of the return of surplus to the employer is very complicated and, depending on laws of the country concerned, may not be possible or may be heavily restricted.

It is also worth noting that the regulations in a given country may prohibit hybrid schemes or limit their form. In Poland, for example, only DC schemes are possible.

7. Conclusions

In a DB scheme, both investment risk and longevity risk are borne by the employer. Modest investment returns and low interest rates have caused the costs and risk of such schemes to increase, leading some employers to close their DB schemes (Clark & Monk 2006, p. 44). They often choose to offer employees a DC scheme instead. However, in such a scheme the risk is borne by the member, who often does not have adequate means to manage the risk. Hybrid schemes provide a mixture of both traditional forms, allowing risk to be shared between employer and member. The exact approach to risk sharing depends on the form of hybrid scheme. In the two forms considered in this article, it is achieved by allocating a given type of risk to a certain party. In a cash balance scheme, the employer bears the investment risk and the member the longevity risk, while in a self-annuitising scheme the opposite is the case.

In the proposed scheme, the risk is shared between member and employer irrespective of its type. An upper and lower limit for extra employer contributions is set. If an increase in the required contribution means the extra contribution exceeds the upper limit, the member is called upon to pay the remainder of the extra contribution, and analogously in the case of a decrease in contributions. In this way, the risk is shared between both parties. The employer will always need to adjust their contribution when such adjustment is required, but only up to a certain limit, so the employer is sure that the extra contribution (and therefore the variability in contributions) will be no greater than a known value. The member's contribution will only need to be adjusted in a certain number of instances, if actual experience is very different that assumed. These features allow for better financial planning for the two parties involved. However, there are several practical issues that must be considered if such a theoretical concept for a pension scheme were to be implemented in practice.

Bibliography

Blake, D. (2006) *Pension Economics.* Chichester: John Wiley & Sons.

Blommestein, H., Janssen, P., Kortleve, N. and Yermo, J. (2009) "Moving Beyond the 'DB vs. DC' Debate: the Appeal of Hybrid Pension Plans". *Rotman International Journal of Pension Management* 2 (2).

Bureau of Labor Statistics (2018) *National Compensation Survey: Employee Benefits in the United States.* March 2018, https://www.bls.gov (accessed: 15.06.2020).

Clark, G. L. and Monk, A. H. B. (2006) "The 'Crisis' in Defined Benefit Corporate Pension Liabilities. Part I: Scope of the Problem". *Pensions* 12, https://doi.org/10.1057/palgrave.pm.5950041.

Cooper, D. R. (2005) *Comparing Pension Outcomes from Hybrid Schemes.* Department for Work and Pensions Research Report 269. London.

Davis, R. (2013) "The Tracker Plan: a Controlled Risk Defined Contribution Retirement Program". *The John Marshall Law Review* 46 (3).
Davis, R. and Madland, D. (2013) *American Retirement Savings Could Be Much Better.* Center for American Progress.
European Commission (2017) *Country Fiche on Pensions for the Netherlands – the 2017 Round of Projections for the Ageing Working Group*, https://ec.europa.eu/info/sites/info/files/economy-finance/final_country_fiche_nl.pdf (accessed: 15.06.2020).
GCAE Groupe Consultatif Actuariel Europeen (2001) *Actuarial Methods and Assumptions Used in the Valuation of Retirement Benefits in the EU and other European Countries*, D. Collinson (ed.), http://actuary.eu/documents/actuarial_methods.pdf (accessed: 4.10.2018).
Gierusz, A. (2019) *Modele podziału ryzyka w hybrydowych pracowniczych programach emerytalnych.* Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.
GUS (2018) *Trwanie życia w 2017 r.* [Life expectance tables of Poland 2017]. Warszawa.
J.P. Morgan (2018) *Long-term Capital Market Assumptions. Time Tested Projections to Build Stronger Portfolios*, www.jpmorgan.com (accessed: 15.06.2020).
KNF (2019) *Pracownicze Programy Emerytalne w 2018 roku.* Warszawa: Urząd Komisji Nadzoru Finansowego.
Mackenzie, S. (2010) *Hybrids and Other Alternatives to the Traditional Pension.* AARP Public Policy Institute.
Palmer, B. (2008) *2008 Replacement Ratio Study*, AON Consulting, http//www.aon.com/about −aon/intellectual − capital/attachments/human − capital −consulting/RRStudy070308.pdf (accessed: 15.06.2020).
Petelczyc, J. (2016) *Pracownicze programy emerytalne w krajach Unii Europejskiej.* Warszawa: Dom Wydawniczy Elipsa.
Pugh, C. and Yermo, J. (2008) "Funding Regulations and Risk Sharing". *OECD Working Papers on Insurance and Private Pensions* 17, OECD Publishing, https://doi.org/10.1787/241841441002.
Sweeting, P. (2007) "DB or Not DB – the Choice of Pension Plan Provision by Employers". *Pensions Institute Discussion Paper* PI-0710. The Pensions Institute, Cass Business School, City University London.
Szczepański, M. and Brzęczek, T. (2016) *Zarządzanie ryzykiem w pracowniczych programach emerytalnych. Uwarunkowania instytucjonalne, ekonomiczno-fiskalne i demograficzne.* Poznań: Polskie Towarzystwo Ekonomiczne.
Takayama, N. (2013) "Matching Defined Contribution Pension Schemes in Japan" in R. Hinz, R. Holzmann, D. Tuesta, N. Takayama (eds) *Matching Contributions for Pensions*. World Bank.
Turner, J. A. (2014) *Hybrid Pensions: Risk Sharing Arrangements for Pension Plan Sponsors and Participants.* The Society of Actuaries Research Project.
Wesbroom, K., Reay, T. (2005) *Hybrid Pension Plans: UK and International Experience.* London: Department for Work and Pensions Research Report 271.
World Bank (1994) *Averting the Old-age Crisis. Policies to Protect the Old and Promote Growth.* Oxford: Oxford University Press.

ISSN 1642-168X
e-ISSN 2545-3866