

Marcinkiewicz, Edyta

**Article**

## Pension systems similarity assessment: An application of Kendall's W to statistical multivariate analysis

Contemporary Economics

**Provided in Cooperation with:**

University of Finance and Management, Warsaw

*Suggested Citation:* Marcinkiewicz, Edyta (2017) : Pension systems similarity assessment: An application of Kendall's W to statistical multivariate analysis, Contemporary Economics, ISSN 2300-8814, University of Finance and Management in Warsaw, Faculty of Management and Finance, Warsaw, Vol. 11, Iss. 3, pp. 303-314, <https://doi.org/10.5709/ce.1897-9254.244>

This Version is available at:

<https://hdl.handle.net/10419/195493>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

Primary submission: 09.07.2015 | Final acceptance: 08.08.2016

# Pension Systems Similarity Assessment: An Application of Kendall's $W$ to statistical multivariate analysis

Edyta Marcinkiewicz<sup>1</sup>

## ABSTRACT

The study aims at empirical verification of the quality of pension system clustering based on two dimensions: the extent of involvement of the state in the pension system and the level of voluntariness. To answer the question of whether these two dimensions actually determine the division into homogeneous groups that constitute pension regimes, Kendall's  $W$  concordance coefficient is employed. It is typically used in Delphi studies as an indicator of expert consensus, but it has been shown that the concept of concordance can also be applied to statistical multivariate analysis to measure intra-group similarity. This empirical research comprises 30 OECD countries grouped into three pension regimes. It employs the classical chi-square test as well as the permutation test of the coefficient of concordance previously applied to empirical problems in biology. Additionally, this study proposes a new permutation procedure that allows for verifying the quality of clustering from a different perspective.

## KEY WORDS:

Kendall's  $W$ , concordance, cluster analysis, pension regimes, pension systems

**JEL Classification:** H53, H55, J32, C12

<sup>1</sup>Lodz University of Technology - Department of Management, Poland

## Introduction

The last two decades have seen significant changes in many pension systems. Substantial pension reform has been implemented in many countries around the world. These changes arise mainly from the deteriorating demographic situation of societies, namely, population aging. Due to these reforms, current pension systems have become increasingly hybridized. Simple distinctions from the past are no longer applicable.

There are many facets and parameters that determine

the final shape of a pension system, such as defined benefit vs. defined contribution, pay-as-you-go vs. funded, voluntary vs. mandatory, publicly managed vs. privately managed, with extensive pension privileges vs. without privileges, etc. The search for the optimal combination of these factors is the main driver for pension reforms. However, this optimal system depends to a large extent on local conditions, such as demographics, politics, and history. That is why there is no single pension system that is optimal for all countries. One of the challenges of pension economics is the comparative analysis of social policy models that comprise pension regimes, mainly in order to study their outcomes in terms of providing adequacy or efficiency. In the face of a variety of solutions imple-

Correspondence concerning this article should be addressed to: **Edyta Marcinkiewicz**, Lodz University of Technology - Department of Management, ul. Piotrkowska 266, Lodz 90-924, Poland. E-mail: [edyta.marcinkiewicz@p.lodz.pl](mailto:edyta.marcinkiewicz@p.lodz.pl)

mented in different pension systems, one can only try to arrange them into similar groups that form models clearly distinguishable from each other. In this context, pension system resemblance can be studied.

The notion of a pension regime originates from welfare state regimes (Furniss & Tilton, 1977; Titmuss, 1974). The most popular typology of pension regimes was developed by Esping-Andersen (1990). It was the first one to function not only as a theoretical concept but was also based on empirical observation. Nonetheless, as stated by many authors (see Soede & Vrooman, 2008), it fails the empirical verification. Different typologies of pension regimes have also been proposed by Rhodes and Natali (2003) and later by Soede and Vrooman (2008). The present study refers directly to the classification of pension models developed by Marcinkiewicz and Chybalski (2017) and can be perceived as an extension of their research. They propose a new typology of pension system regimes based only on two opposites: voluntary vs. mandatory systems and the involvement of the state vs. the market. These dimensions are assumed to comprise the main factors that allow one to distinguish among pension systems. The aforementioned study empirically identifies three groups of contemporary pension systems in terms of these differentiating criteria and provides theoretical foundations for such a division. The present work aims at the empirical verification of the quality of this pension system clustering in order to answer the question of whether these two dimensions actually determine the division into homogeneous groups. It employs Kendall's *W* concordance coefficient (Kendall & Babington Smith, 1939), which is typically used in Delphi studies as a measure of expert consensus (Von der Gracht, 2012). However, as justified in subsequent sections, it can be useful when evaluating the homogeneity in the groups of different objects described by a multiple of variables, i.e., in the statistical multivariate analysis. For such purposes, the concordance coefficient has already been adopted in the biological sciences. For example, in community ecology, Legendre (2005) uses Kendall's *W* to identify groups of significantly associated species. Similarly, Buchholz, Hannig and Schirmel (2013) adopt it to carabid species clustering. In medical applications, Baumgartner, Somorjai, Summers and Richter (1999) employ it to assess intra-group homogeneity of clusters of voxels in functional

magnetic resonance imaging. Stausberg, Halim and Färber (2011) apply the coefficient of concordance to validate whether several indicator sets compete with one another for the assessment of effectiveness and safety of hospitals.

The contribution of this study to the current literature is twofold. First, it applies Kendall's *W* to the problems of multivariate analyses in the field of economics, namely, to the empirical verification of pension regime typology. The main advantage of the application of Kendall's *W* in cluster analysis is that the coefficient is scaled to a certain interval and is comparable across different groups, contrary to the commonly used measures of intra-group (dis)similarity. Second, this paper proposes a new application of the permutation test, which can be considered as complementary to the approach previously suggested by Legrande (2005). This study employs a dataset concerning various variables used in the classification of 30 pension systems developed by Marcinkiewicz and Chybalski (2017). The variables reflect two dimensions: the extent of the involvement of the state in the pension system and the level of voluntariness. Whereas the abovementioned study consists of theoretical and empirical identification of groups of similar pension systems forming pension regimes, this paper aims to assess the quality of this division. The statistical framework proves that the theoretical distinction into three pension regimes has counterparts in current pension system models.

## Method

In statistical multivariate analysis, particularly in cluster analysis, there are several measures used to express separation or homogeneity of a single cluster (see Hansen & Jaumard, 1997). However, they are applied in the course of finding clusters as criteria of modifying subsets or terminating the procedure. They are based on dissimilarities and rely mainly on distance metrics. They are also not interpretable. As such, they are not suitable for comparisons of the divisions of different categories of objects. On this basis, one can only conclude that the previous division across the same set of objects was better or worse than the next division. In fact, surprisingly little attention in the existing literature on statistical multivariate analysis is paid to the assessment of the quality of clustering. Of the very

few measures that enable identifying the level of intra-group resemblance, one can distinguish the intra-class correlation coefficient ICC (Koch, 1982). However, it is designed to reveal correlation mainly in data sets where the measurements are not assigned to certain units (objects). In the absence of appropriate measures of intra-group similarity in the case of multivariate analyses, the concept of concordance can be applied. As presented in Elzinga, Wang, Lin and Kumar (2011), the concept of concordance appears in at least three contexts: in voting and decision making, which is the first and basic application, in group attitude assessment, and in statistics. The research presented in this study addresses the third of the aforementioned applications using the most known index of concordance for  $m > 2$  rankings, namely, Kendall's  $W$ . Originally, it was developed to measure the consensus level in a group of experts in terms of their choices or preferences. The formula for Kendall's  $W$  is as follows:

$$W = \frac{\sum_{j=1}^n (R_j - \frac{m(n+1)}{2})^2}{\frac{1}{12} m^2 (n^3 - n)} \quad (1)$$

where  $n$  denotes the number of ranked items,  $m$  denotes the number of experts (raters) and  $R_j$  is the sum of ranks obtained by the  $j$ -th item from all the raters.

In the numerator of the ratio given in formula (1), there is a sum of squared deviations of  $R_j$  from the mean value of  $R_j$  computed for all the items. The denominator comprises the maximal sum of squared deviations, i.e., attainable only in the case of full consensus among raters, when  $m$  series of ranks are identical. Consequently, when the actual dispersion of ranks reaches its maximum and  $W$  equals 1, this indicates full agreement. When raters strongly disagree, and the sums of their ranks are randomly (equally) distributed among items, then the dispersion of sum of ranks will be equal to zero, and as a result,  $W$  will take the zero value.

However, when the ranks are tied (i.e., when at least one rater gives the same rank for at least two items), formula (1) needs to be adjusted using the correction factor:

$$T = \sum_{k=1}^g (t_k^3 - t_k) \quad (2)$$

where  $t_k$  is the number of tied ranks in each  $k$  of  $g$  groups of ties observed over  $m$  ratings. Kendall's  $W$  is then obtained from the formula:

$$W = \frac{\sum_{j=1}^n (R_j - \frac{m(n+1)}{2})^2}{\frac{1}{12} m^2 (n^3 - n) - mT} \quad (3)$$

The data set for Kendall's  $W$  calculations is typically given in the table form, such as that presented in Table 1. The table is filled with the ranks given by each  $i$ -th ( $i = 1, \dots, m$ ) rater for each  $j$ -th ( $j = 1, \dots, n$ ) item. The concordance coefficient obtained from such data set reflects the level of agreement, which can be defined as the similarity in raters' (experts') views.

The alternative approach presented here allows one to measure how a group of objects is similar (consent) in terms of the variables that the objects are characterized with. In other words, the variables or features describing particular objects in a group are ranked by each object (see Table 2).

The idea of ranking variables may seem confusing; however, the analogous concept applies to the Pearson's correlation coefficient often used in multivariate analysis. It is typically employed as a measure of correlation between two variables, but it can also be computed with respect to two objects described by a set of variables, and as such, it serves as a similarity measure. However, the advantage of Kendall's  $W$  over the Pearson correlation coefficient is that it can evaluate similarity among more than two objects. Moreover, because it is based on ranks, it is also applicable when the relationship is nonlinear. As noted by Elzinga et al. (2011), in statistics, the concept of concordance is useful to know how much large values of one random variable  $X$  correspond to large values of another random variable  $Y$ .

Baumgartner et al. (1999) noted that Kendall's  $W$  may be regarded as a cluster validity index. In this case, the concordance coefficient can be adopted to measure similarity in a group of objects in terms of all the variables used to form the cluster. However, it can also be used when only some of the variables are taken into account. Then, it allows for identifying the subsets of variables (features) that contribute the most or the least to intra-group homogeneity. Finally, one can evaluate the similarity of objects quite apart from dif-

**Table 1.** Data set presentation for the standard Kendall's  $W$  application

		Items			
		1	2	...	$n$
Raters	1				
	2				
	...				
	$m$				
Sum of ranks ( $R_j$ )					

**Table 2.** Data set presentation for the Kendall's  $W$  application in multivariate analysis

		Variables			
		1	2	...	$n$
Objects	1				
	2				
	...				
	$m$				
Sum of ranks ( $R_j$ )					

ferentiating variables, e.g., after forming the clusters of pension regimes, their homogeneity can be assessed in terms of adequacy, efficiency, and sustainability.

The standard procedure for testing Kendall's  $W$  concordance coefficient involves chi-square statistics:

$$\chi^2 = m(n-1)W \quad (4)$$

The null hypothesis states that the raters produce independent rankings of the items, i.e., the rating series are not related with each other. As shown by Legendre and Lapointe (2004), there is a close similarity between Friedman's two-way analysis of variance without replication by ranks and Kendall's  $W$ , as the only difference is the formulation of the null hypothesis. The statistics given by formula (4) are asymptotically distributed like chi-square, with  $(n-1)$  degrees of freedom.

A study by Legrande (2005) employs resampling as a very useful solution when verifying the hypothesis that the rankings are independent. He proposes the following procedure of the permutation test:

1. Transform quantitative data into ranks (tied ranks if necessary).
2. Compute Kendall's  $W$  coefficient, which will serve as a reference statistic ( $W_{ref}$ ).
3. Permute all rankings at random, independently from rater to rater, and compute Kendall's  $W$ .
4. Repeat step 3 numerous times and develop a distribution of the  $W$  statistics under permutation.
5. Calculate a p-value as the proportion of  $W$  that are larger or equal to  $W_{ref}$  and compare it to the selected significance level.

The test described above verifies the same null hypothesis as the test with chi-square statistics given by formula

(4). With respect to the data presentation in Table 2, permutations apply to the values of the variables assigned to the  $i$ -th object. They are randomly mixed numerous times independently for each object. In this manner, the test verifies the null hypothesis that the objects are not similar in terms of a given set of variables.

This study presents a complementary test based on resampling that can be applied with respect to the problems that involve validation of clustering. It can also be considered as a permutation test. The previous test involves variable permutations in a fixed set of  $m$  objects, whereas in this test, the labels of objects (belonging to the group and not belonging to the group) are permuted. This consists of sampling, without replacement, a set of  $m$  objects out of a set of  $k$  objects ( $k > m$ ). The procedure is analogous to the test presented by Legrande (2005). The steps are as follows:

1. Transform quantitative data corresponding to the  $k$  objects into ranks (tied ranks if necessary).
2. Compute Kendall's  $W$  coefficient for a preselected set of  $m$  objects as a reference statistic ( $W_{ref}$ ).
3. Extract randomly from the available set of  $k$  objects a subset of  $m$  objects and compute Kendall's  $W$ .
4. Repeat step 3 numerous times and develop a distribution of the  $W$  statistics under permutation.
5. Calculate a p-value as the proportion of  $W$  that are larger or equal to  $W_{ref}$  and compare it to the selected significance level.

This test is useful in clustering validation, especially in the case when two or more clusters are formed out of the given set of objects. It can also serve as a classifier performance assessment tool. The p-value is then a proportion of clusters that are characterized by the same or greater intra-group similarity than the original set of objects. The null hypothesis states that the given set of objects is randomly selected.

The first permutation test applies to the assessment of the similarity of objects within a preselected group, without reference to other possibilities of grouping. The second test applies strictly to the assessment of the accuracy of clustering – not the absolute (objective) accuracy but the relative one. Hypothetically, as a result of clustering, it is likely to obtain a group of dissimilar objects extracted from a larger set, but this group can still be the most homogenous compared to other possible combinations. That is why these two tests should be considered as complementary. Whereas the first test

corresponds to the question of whether the objects in a given group are similar, the second refers to the question of if there are *better* (more homogenous) groups of objects. To distinguish between these two types of tests in further empirical analysis, the first one shall be called the *absolute* permutation test and the second one the *relative* permutation test.

## Empirical Research

This study is based on the empirical results of Marcinkiewicz and Chybalski (2017). Using statistical methods, such as cluster analysis, correlation analysis, and ANOVA, they distinguish three groups of pension systems that form three pension regimes. The criteria serving for this differentiation are based on the two dimensions that can be expressed as two pairs of opposite notions: public vs. private and mandatory vs. voluntary. Table 3 presents the list of the variables (extracted from the OECD database) they used in the empirical research. The same set of variables is employed in this study to validate the intra-group pension system homogeneity. The data apply to 2011 and 2012.

Table 4 presents the data set used in cluster analysis. Three clusters of pension systems currently functioning in OECD countries were distinguished. The first group includes the following countries: Canada, United Kingdom, United States, Ireland, Czech Republic, and New Zealand. They form the cluster with a significant share of voluntary pensions in the system. The second group comprises pension systems in Australia, Denmark, Iceland, Netherlands, Sweden, Switzerland, Slovak Republic, Estonia, Israel, and Poland. This cluster covers the regime of the significant mandatory participation in private schemes. The last group is the most numerous and consists of the countries with pension systems classified to the mandatory public regime: Austria, Belgium, Finland, France, Germany, Greece, Hungary, Italy, Luxembourg, Norway, Portugal, Slovenia, Spain, and Turkey. All the figures presented in Table 4 represent the shares (rates), and all of them are in the range 0–1, so the data included in the further analysis are not adjusted in any way (e.g., standardized).

The classification into the three regimes is made for a particular moment in time, as it employs data from 2011 or 2012. Because the systems in many countries undergo reforms, the composition of the

**Table 3.** Description of variables

Variable	Description
X1	Ratio between the expenditure on old-age pension provisions from the mandatory public schemes and total expenditure on old-age pensions (public and private mandatory and voluntary private). This measures the share of mandatory and publicly managed schemes in the entire pension system;
X2	Ratio between the expenditure on old-age pension provisions from the mandatory private schemes and total expenditure on old-age pensions (public and private mandatory and voluntary private). This measures the share of mandatory and privately managed schemes in the entire pension system;
X3	Ratio between the expenditure on old-age pension provisions from the voluntary private schemes and total expenditure on old-age pensions (public and private mandatory and voluntary private). This measures the share of privately managed voluntary schemes in the entire pension system;
X4	Ratio between the estimated net pension replacement rate from the public pension system (for the person entering labor market in 2012 and earning an average wage) and the sum of the estimated net pension replacement rates from the public, mandatory private and voluntary pension system. This measures the contribution of mandatory public pension system in ensuring the income adequacy;
X5	Ratio between the estimated net pension replacement rate from the mandatory private pension system (for the person entering labour market in 2012 and earning an average wage) and the sum of the estimated net pension replacement rates from the public, mandatory private and voluntary pension system. This measures the contribution of mandatory private pension system in ensuring the income adequacy;
X6	Ratio between the estimated net pension replacement rate from the voluntary private pension system (for the person entering labor market in 2012 and earning an average wage) and the sum of the estimated net pension replacement rates from the public, mandatory private and voluntary pension system. This measures the contribution of voluntary pension system in ensuring the income adequacy;
X7	Coverage of mandatory private pension schemes by the type of plan, expressed as a percentage of the working age population (15–64 years);
X8	Coverage of voluntary private pension schemes by the type of plan, expressed as a percentage of the working age population (15–64 years), calculated as the maximum of two values: coverage of voluntary occupational schemes and coverage of voluntary personal schemes;
X9	The rate of public mandatory pension contribution (if does not exist, X9=0);
X10	Percentage contribution of all components of the first tier of the pension system to weighted average pension wealth. This measures the share of public minimum pension provision in the whole retirement income package in the mandatory system;
X11	Percentage contribution of public ER or public DC provision from the second tier of the pension system to weighted average pension wealth. This measures the share of public ER or DC provision in the whole retirement income package in the mandatory system;
X12	Percentage contribution of private DB or private DC provision from the second tier of the pension system to weighted average pension wealth. This measures the share of private pension provision in the whole retirement package in the mandatory system.

Note: Only X7, X8 and X9 variables are directly provided by the OECD, the remaining variables are transformed into ratios from the original data.

Source: Adapted from "A new proposal of pension regimes typology: Empirical analysis of the OECD countries" by E. Marcinkiewicz and F. Chybalski (2017). Retrieved from Journal of Economic Policy Reform, <http://dx.doi.org/10.1080/17487870.2016.1276454>

**Table 4.** Pension system characteristics

Group	Country	Variables											
		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
I	Canada	0.55	0.00	0.45	0.54	0.00	0.46	0.00	0.33	0.10	0.50	0.50	0.00
	United Kingdom	0.54	0.06	0.40	0.49	0.00	0.51	0.00	0.30	0.00	0.88	0.12	0.00
	United States	0.57	0.00	0.43	0.50	0.00	0.50	0.00	0.42	0.10	0.00	1.00	0.00
	Ireland	0.85	0.00	0.15	0.46	0.00	0.54	0.00	0.31	0.00	1.00	0.00	0.00
	Czech Republic	0.94	0.06	0.00	0.53	0.00	0.47	0.00	0.62	0.28	0.18	0.82	0.00
	New Zealand	1.00	0.00	0.00	0.74	0.00	0.26	0.00	0.64	0.00	1.00	0.00	0.00
II	Australia	0.67	0.05	0.28	0.26	0.74	0.00	0.68	0.20	0.00	0.39	0.00	0.61
	Denmark	0.64	0.00	0.36	0.39	0.61	0.00	0.84	0.24	0.00	0.45	0.00	0.55
	Iceland	0.46	0.00	0.54	0.09	0.91	0.00	0.85	0.42	0.00	0.15	0.00	0.85
	Netherlands	0.59	0.00	0.41	0.33	0.67	0.00	0.88	0.28	0.18	0.37	0.00	0.63
	Sweden	0.78	0.00	0.22	0.61	0.39	0.00	1.00	0.27	0.18	0.03	0.53	0.44
	Switzerland	0.60	0.40	0.00	0.58	0.42	0.00	0.70	0.00	0.10	0.00	0.65	0.35
	Slovak Republic	0.94	0.01	0.04	0.57	0.43	0.00	0.44	0.00	0.18	0.00	0.58	0.42
	Estonia	1.00	0.00	0.00	0.52	0.48	0.00	0.69	0.00	0.22	0.29	0.28	0.42
	Israel	1.00	0.00	0.00	0.30	0.70	0.00	0.82	0.00	0.07	0.34	0.00	0.66
	Poland	1.00	0.00	0.00	0.50	0.50	0.00	0.56	0.01	0.20	0.00	0.51	0.49
III	Austria	0.94	0.00	0.06	1.00	0.00	0.00	0.00	0.20	0.23	0.00	1.00	0.00
	Belgium	0.88	0.00	0.12	0.73	0.00	0.27	0.00	0.45	0.16	0.02	0.98	0.00
	Finland	0.98	0.00	0.02	1.00	0.00	0.00	0.74	0.19	0.23	0.01	1.00	0.00
	France	0.99	0.00	0.01	1.00	0.00	0.00	0.00	0.16	0.17	0.00	1.00	0.00
	Germany	0.91	0.00	0.09	0.72	0.00	0.28	0.00	0.56	0.20	0.00	1.00	0.00
	Greece	0.98	0.00	0.02	1.00	0.00	0.00	0.00	0.00	0.20	0.45	0.55	0.00
	Hungary	1.00	0.00	0.00	1.00	0.00	0.00	0.01	0.20	0.34	0.00	1.00	0.00
	Italy	0.91	0.07	0.02	1.00	0.00	0.00	0.00	0.08	0.33	0.00	1.00	0.00
	Luxembourg	0.94	0.00	0.06	1.00	0.00	0.00	0.00	0.03	0.16	0.22	0.78	0.00
	Norway	0.91	0.00	0.09	0.72	0.11	0.18	0.68	0.23	0.00	0.01	0.88	0.11
	Portugal	0.97	0.00	0.03	1.00	0.00	0.00	0.00	0.05	0.00	0.03	0.97	0.00
	Slovenia	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.24	0.06	0.94	0.00
	Spain	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.16	0.28	0.00	1.00	0.00
Turkey	1.00	0.00	0.00	1.00	0.00	0.00	0.01	0.05	0.20	0.01	0.99	0.00	

Source: Adapted from "A new proposal of pension regimes typology: Empirical analysis of the OECD countries" by E. Marcinkiewicz and F. Chybalski (2017). Retrieved from Journal of Economic Policy Reform, <http://dx.doi.org/10.1080/17487870.2016.1276454>



**Table 5.** Pension systems intra-group similarity assessment

Pension system group	m	n	W ( $W_{ref}$ )	$\chi^2$ test		abs. perm. test	rel. perm. test
				$\chi^2$ -Stat	p-value	p-value	p-value
I	6	12	0.74	48.52	0.0000	0.0000	0.0327
II	10	12	0.69	76.13	0.0000	0.0000	0.0066
III	14	12	0.79	122.06	0.0000	0.0000	0.0000

**Table 6.** Pension system intra-group similarity assessment (factors: public mandatory, private mandatory, voluntary)

Variables (factors)	Pension system group	m	n	W ( $W_{ref}$ )	$\chi^2$ test		abs. perm. test	rel. perm. test
					$\chi^2$ -Stat	p-value	p-value	p-value
X1,X4,	I	6	4	0.75	13.45	0.0038	0.0008	0.1363
X9,X11	II	10	4	0.73	21.80	0.0001	0.0000	0.0575
	III	14	4	0.69	28.97	0.0000	0.0000	0.0530
X2,X5,	I	6	4	0.33	6.00	0.1116	0.2486	0.4086
X7,X12	II	10	4	0.83	24.76	0.0000	0.0000	0.0000
	III	14	4	0.18	7.36	0.0612	0.0588	0.7511
X3,X6,X8	I	6	3	0.36	4.33	0.1146	0.1435	0.3031
	II	10	3	0.45	9.00	0.0111	0.0058	0.0640
	III	14	3	0.54	15.04	0.0005	0.0002	0.0022

groups may change over time. However, this study attempts to verify the validity of the concept that the main axes for grouping pension systems into similar clusters are the level of voluntariness and the level of involvement of the public sector. To achieve this goal, the Kendall's coefficient of concordance was calculated and the corresponding statistical tests were run: the classical test with the chi-square statistics given by formula (4), as well as two kinds of permutation tests (referring to absolute and relative similarity assessment). The number of random assignments in all permutation tests was 10,000.

According to the procedure for determining the concordance coefficient, the values characterizing particular pension systems were transformed into rankings. The results of the calculations are shown in Table 5. The most homogeneous group is the third group, as the  $W$  equals 0.79, which indicates a high level of concordance. However, the other two groups are very concordant as well. The obtained chi-square statistics also confirm that the rankings are related. This is in line with the results of the *absolute* permutation test, which verifies the same null hypothesis. The p-values computed for the *relative* permutation

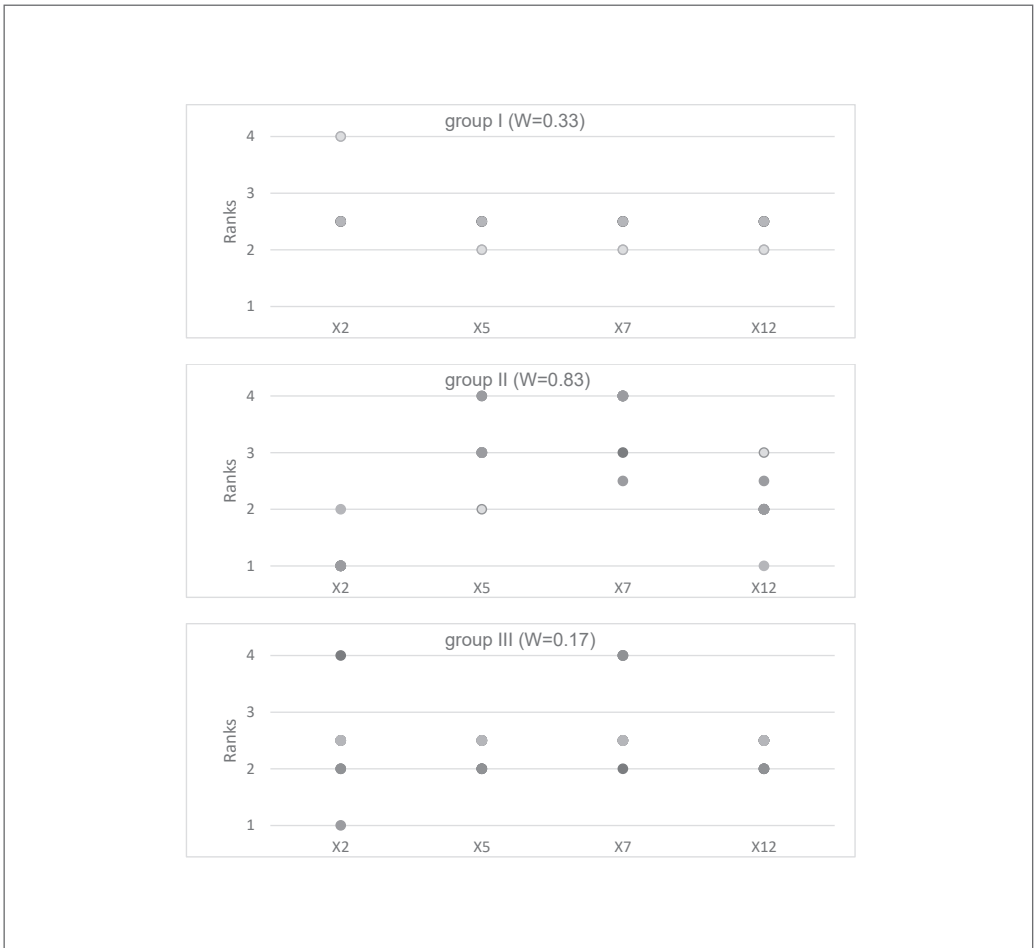


Figure 1. Rank distribution across variables

tests indicate that all three groups of pension systems are not randomly selected.

As previously stated, Kendall's coefficient of concordance can also be applied to make comparisons among different sets of variables in order to assess which factor has the greatest contribution to the overall homogeneity of a group. Therefore, the next stage of the research was to estimate the concordance within three pension regimes, but only with respect to the certain factors. The first factor is reflected by variables X1, X4, X9, and X11. It represents the publicly managed mandatory schemes. The second factor is described by variables

X2, X5, X7, and X12, which represents the significance of the mandatory and private parts of a pension system. Finally, the third factor comprises variables X3, X6, and X8, which reflect the level of voluntariness in a pension system.

Table 6 presents the calculations of Kendall's W for each pension regime with respect to the distinguished sets of variables. The results obtained for the first set of variables show that all three pension regimes are characterized by a high level of intra-group homogeneity. However, regarding the second factor, namely, mandatory and private schemes, there is a concordance

**Table 7.** Pension system intra-group similarity assessment (perspectives: beneficiaries, contributors)

Variables (factors)	Pension system group	m	n	W ( $W_{ref}$ )	$\chi^2$ test		abs. perm. test	rel. perm. test
					$\chi^2$ -Stat	p-value	p-value	p-value
X1,X2,X3	I	6	3	0.85	10.17	0.0062	0.0018	0.7273
	II	10	3	0.75	15.08	0.0005	0.0000	0.8985
	III	14	3	0.90	25.08	0.0000	0.0000	0.2078
X4,X5,X6,X7	I	6	9	0.68	32.76	0.0001	0.0000	0.0334
X8,X9,X10	II	10	9	0.66	52.89	0.0000	0.0000	0.0036
X11,X12	III	14	9	0.77	86.78	0.0000	0.0000	0.0000

only in the second pension regime, which includes the countries with significant mandatory participation in private schemes. Kendall's  $W$  is relatively small for the other two groups of countries. Still, these results can be misleading. When comparing the values of variables X2, X5, X7, and X12 in the first and third cluster (Table 4), it is clear that these groups are homogenous as well. The reason why the Kendall's  $W$  indicates a lack of concordance is the fact that the figures from the columns consist of many zeroes. In such a case, there are many equal ranks (ties), which results in a small sum of squared deviations of  $R_j$  from the mean value of  $R_j$  in the numerator of the ratio given in formula (3). This effect is visible when the distribution of respective ranks is plotted on the graph. Figure 1 presents how the ranks are spread across variables X2, X5, X7, and X12 in all three groups.

As for the third factor, namely, the level of voluntariness, expressed by variables X3, X6, and X8, the less concordant are the countries from the first group. Compared to the latter two regimes, they are characterized by the high significance of voluntary pension schemes. However, it is not a homogenous group. The intra-group similarity is relatively higher in the second and third cluster, but the mean values of X3, X6, and X8 in these groups indicate that this factor plays a minor role in differentiating the pension systems.

With respect to all three sets of variables representing different factors characterizing pension regimes, there are a few discrepancies between the results of the

relative permutation tests and the results of the other two tests. This implies that in these cases, the groups of pension systems are highly concordant, but in a different composition they could have been even more concordant. When separately considering different features of pension systems, such as public mandatory, private mandatory and voluntary, one can argue that the first group of pension systems is characterized by the smallest intra-group similarity. Moreover, there are many other combinations of countries that would constitute more homogenous groups, which is indicated by high p-values of the relative permutation tests. However, when all the factors are taken into account simultaneously, the results imply that the first group as well as the second and the third are concordant, as presented in Table 5.

When analyzing the entire set of variables, one can notice that they can be divided according to two perspectives: beneficiaries of the pension systems and contributors to the pension systems. The variables X1–X3 are associated with the expenditure on old-age pension provisions, so they are seen from the pensioners' perspective. Conversely, the variables X4–X12 are defined with respect to the working generation and their direct participation in the pension system. The results presented in Table 7 show that all three groups of pension regimes are highly homogeneous with regard to the components of the pension provision (variables X1–X3). In fact, as indicated by the results of the relative permutation tests, all the studied countries are quite similar in terms of the proportion between variables X1, X2

and X3: there are numerous random combinations of countries characterized by very high values of Kendall's concordance coefficient. This is clearly seen when comparing the figures from Table 4. In a great majority of countries from all three clusters, the pattern remains the same: the higher pension expenditure comes from the public mandatory schemes, the intermediate in terms of the share is pension provision from the voluntary private schemes, and the smallest expenditures come from the private mandatory schemes. The homogeneity with respect to the variables representing contributors to the system is smaller in all three groups of pension systems; however, it is large enough to support the conclusion about the substantial intra-group similarity.

## Discussion and Conclusion

The implications of this study can be considered in two areas: first, in reference to the pension regimes typology, and second, regarding the methodological problem of the assessment of clustering quality. This study proves that the concept of pension regime typology based on the extent of involvement of the state and the market as well as the role of voluntary schemes in a pension system is consistent with the empirically distinguished groups of pension systems operating in OECD countries. This consistency is assessed by the level of intra-group similarity. Pension regime typology can have a significant practical value for social policy when indicating the best directions for pension reforms, but only when such typology properly captures the main dimensions of pension systems. This study proves the accuracy of the typology by Marcinkiewicz and Chybalski (2017), which is quite novel and conceptually different from older typologies. Thus, it confirms that this classification can be further used in the comparative studies of social policy outcomes such as pension system adequacy or efficiency. The other implication of this study is of a methodological nature. It presents the original implementation of a new tool for assessing the quality of clustering. It has been demonstrated that the concept of concordance can be used for this purpose as a measure reflecting the similarity of the examined objects. The method employed in the empirical research is based on the Kendall's  $W$  concordance coefficient. Unlike the typical similarity measures used in the statistical multivariate analysis, it allows for the measurement of intra-group similarity when the group consists of

more than two elements. However, as in every measure, there are some limitations. First, it is based on ranks, which means the loss of information that results from the data transformation into the ordinal scale in the case of variables expressed at an interval or ratio scale. It does not distinguish whether the differences between the levels of the variables, which determine the ranking, are relevant or irrelevant in terms of the similarity of the evaluated objects. However, the analysis based on ranks is also applicable when the studied relationships are non-linear. The issue of many ties is also important, as this can influence the results of the analysis despite the presence of a correction factor in the formula describing Kendall's  $W$ . In the presented study, some of the calculated coefficients of concordance were very low because there were many columns of zeroes, and hence the ranks were equal. Another interesting problem that draws little attention in the existing literature is the issue of data adjustment for the purpose of concordance analysis. Transformations of raw data, such as standardization or scaling at the 0–1 interval, are a common practice in many procedures of statistical multivariate analysis; however, they seemed to be unnecessary in this study, as all the variables were already expressed as rates and as such were comparable to some extent. Nevertheless, data adjustment can seriously affect the results of concordance analysis.

In this study, the concordance is identified with intra-group similarity or homogeneity. However, it should be clearly stated that the term *similarity* applies to the proportions or shape and not to the same absolute level. It is a correlation-based similarity. As a result, in the case of a set of identical objects, the concept of concordance has no application; it can lead to false conclusions.

The results obtained during the course of the study reveal that the level of voluntariness and the extent of involvement of the state in the pension system are the dimensions that form the basis for distinguishing among different social policy models with respect to old-age pensions. The empirical research presented in this paper confirms that these two facets determine division into separate pension regimes. As it has been demonstrated, the level of concordance, which is identified with intra-group similarity, is high; thus, the quality of the division into pension regimes was proved. The additional conclusions can be formulated after de-

composition of the entire set of variables into subsets representing different features of pension systems. For example, when we consider only the degree of involvement of the state (i.e., the variables representing public and mandatory character), it turns out that the group of countries that represent pension regimes with the dominant role of the state is the best match in terms of intra-group similarity. The same applies to the countries representing the regime with the significant role of the market (with the substantial involvement of the private and mandatory pension schemes). The countries that constitute this group are the most consistent with respect to this feature. Only in the case of countries with a major role of voluntary pension schemes is the intra-group similarity not identified. This means, however, that in these countries, there is a considerable variation in this respect. Analysis of row data shows that compared to other regimes, they are still the group where voluntary forms of old-age provision are of great importance. However, one should keep in mind that the ultimate validity of any pension regime's typology is verified through its usefulness, as it should be formed in such a way as to capture the common patterns within the grouped systems, especially those that enable further pension system improvements.

## References

- Baumgartner, R., Somorjai, R., Summers, R., & Richter, W. (1999). Assessment of cluster homogeneity in fMRI data using Kendall's coefficient of concordance. *Magnetic Resonance Imaging*, 17(10), 1525-1532.
- Buchholz, S., Hannig, K., & Schirmel, J. (2013). Losing uniqueness - shifts in carabid species composition during dry grassland and heathland succession. *Animal Conservation*, 16(6), 661-670.
- Elzinga, C., Wang, H., Lin, Z., & Kumar, Y. (2011). Concordance and consensus. *Information Sciences*, 181(12), 2529-2549.
- Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Princeton, NJ: Princeton University Press.
- Furniss, N., & Tilton, T. (1977). *The Case for the Welfare State: From social security to social equity*. Bloomington, IN: Indiana University Press.
- Hansen, P., & Jaumard, B. (1997). Cluster analysis and mathematical programming. *Mathematical Programming*, 79(1-3), 191-215.
- Kendall, M. G., & Babington Smith, B. (1939). The problem of m rankings. *The Annals of Mathematical Statistics*, 10(3), 275-287.
- Koch, G. G. (1982). Intraclass correlation coefficient. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences* (4th ed., pp. 213-217). New York, NY: John Wiley & Sons.
- Legendre, P. (2005). Species associations: The Kendall Coefficient of Concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(2), 226-245.
- Legendre, P., & Lapointe, F. J. (2004). *Assessing the congruence among distance matrices: Single malt Scotch whiskies revisited*. *Australian and New Zealand Journal of Statistics*, 46, 615-629.
- Marcinkiewicz, E., & Chybalski, F. (2017). A new proposal of pension regimes typology: Empirical analysis of the OECD countries. *Journal of Economic Policy Reform*, <http://dx.doi.org/10.1080/17487870.2016.1276454>
- Rhodes, M., & Natali, D. (2003). Welfare regimes and pension reform agendas [PDF document]. Retrieved from [http://www.lse.ac.uk/european-Institute/research/hellenic-Observatory/pdf/pensions\\_conference/Rhodes-Pensions.pdf](http://www.lse.ac.uk/european-Institute/research/hellenic-Observatory/pdf/pensions_conference/Rhodes-Pensions.pdf)
- Soede, A., & Vrooman, C. (2008). A comparative typology of pension regimes (Research Report No. 54). European Network of Economic Policy Research Institutes. Retrieved from <https://www.files.ethz.ch/isn/93990/RR%20054.pdf>
- Stausberg, J., Halim, A., & Färber, R. (2011). Concordance and robustness of quality indicator sets for hospitals: an analysis of routine data. *BMC Health Services Research*, 11, 106-114.
- Titmuss, R. (1974). *Social policy: An introduction*. London, UK: Allen & Unwin.
- Von der Gracht, H. A. (2012). Consensus measurement in Delphi studies. Review and implications for future quality assurances. *Technological Forecasting & Social Change*, 79(8), 1525-1536.

## Acknowledgements

This work was supported by the National Science Center (Poland) under grant number DEC-2013/09/B/HS4/01516