Journal of Applied Mathematics and Computational Mechanics 2025, 24(2), 71-85
www.amcm.pcz.pl p-ISSN 2299-9965
DOI: 10.17512/jamcm.2025.2.06 e-ISSN 2353-0588

ROUNDING ERRORS IN PAIRWISE COMPARISONS MATRICES BY MONTE CARLO ANALYSIS

Tomasz Starczewski

Department of Mathematics, Czestochowa University of Technology
Częstochowa, Poland
tomasz.starczewski@pcz.pl

Received: 7 February 2025; Accepted: 22 June 2025

Abstract. A pairwise comparisons technique is used to obtain preferences of decision makers. It is especially useful in multi-criteria decision analysis. One of the most famous methods based on pairwise comparisons is the Analytic Hierarchy Process (AHP). In this method, the comparisons are limited to discrete values of a rating scale. This limitation of comparisons may cause rounding errors. In the AHP, inaccuracy in decision-making assessments may be detected by inconsistency analysis. However, the rounding is not a decision maker's mistake. In this article, the size of rounding errors and the values of inconsistency indices related to this kind of error are investigated. Presented results are obtained using the Monte Carlo experiment.

MSC 2010: *90B50, 90C29, 90C31, 91B16*
Keywords: *AHP, multicriteria decision making, pairwise comparisons, inconsistency indices*

1. Introduction – pairwise comparisons and AHP

One of the crucial tasks of multi-criteria decision-making (MCDM) methods is to evaluate preferences of a decision maker (DM) on a number scale. There are various approaches to this issue [1–4]. Among them, one of the more interesting ways to obtain the DM's opinions is asking about the relative importance of objects in pairs [5]. In this approach, in each step, the DM compares two alternatives in terms of one criterion. The most popular method, which employs the pairwise comparisons technique is the Analytic Hierarchy Process (AHP) [6]. In this method, the DM evaluates ratios of preferences for pairs of alternatives by properly designed responses, which belong to a rating scale. These comparisons are next replaced by numbers. A collection of numerical values and their associated statements is referred to as a rating scale [7, 8].

Although a set of numbers (and up to a point a set of rating options) could contain almost infinitely many items, it must be limited because of the ability of human cognitive resolution [9] to discrete values of a scale. This limitation influences on the accuracy of the method and may cause some drawbacks such as ranking reversals and the others (cf. [3, 10]). Moreover, in the AHP procedure, a DM usually compares each pair of the objects only once since it is assumed that the comparisons of the same objects in reverse order are reciprocal. However, it has been proven that enforcing reciprocity leads to deterioration of the AHP methodology too [10]. The coupling of language statements with numbers can sometimes also be misleading (cf. [8]).

The comparisons of all pairs of alternatives are collected in a matrix, which is called the Pairwise Comparisons Matrix (PCM). Then the values of relative importance of each alternative are calculated based on such PCMs. These values are usually collected in the Priority Vector (PV). There are some methods of calculating values of a PV, which give the same results in the case when PCMs satisfy the consistency condition. In this case the values of preferences gathered in the PV are fully coherent with the values obtained from the DM's opinions. However, fully consistent PCMs do not arise often, also because of the limitation of using scale. Therefore, the obtained results are almost always inaccurate.

The pairwise comparisons technique aids the DM in obtaining true values of preferences. However, because of the limit of a rating scale applied in the AHP (related to the limitation of the DM's reasoning), the usability of this technique is limited. The PCM's elements are the numbers from only the rating scale. As a consequence, the values in the PV are also limited, and this is one of the reasons for their inaccuracy. Also, enforcing reciprocity of the PCM usually leads to definitely worse results of the procedure (cf. [10]). This inaccuracy, which is not the result of a DM's mistake but of procedure deficiency, we call the rounding error (ROE). The ROE also has an influence on PCM consistency and may lead to the wrong order of alternatives [3, 10, 11]. The ROEs are not mistakes, although they may sometimes cause basically correct PCMs to be subjected to an "improving" procedure (which is proven later).

The ROEs may have a significant impact on the results of the AHP, but they are usually ignored in literature. Some remarks have been noted in the articles [11, 12], in which the magnitude of absolute errors in the PV caused by the rounding procedure and their relationship with frequencies of changes in alternatives order have been presented. The impact of reciprocity-enforcing procedures has been analysed in [10] in turn. In our paper, we would like to show more results concerning the analysis of ROEs. In the presented article, the absolute and relative value of errors and a value of some CIs connected with ROEs are analysed. Some histograms of the values of CIs are presented as well. Moreover, values of CIs for rounded PCMs are juxtaposed with thresholds of inconsistency used in AHP methodology. This comparison shows whether the commonly used threshold would allow us to accept correct PCMs, which only contain internal inaccuracies of the procedure.

2. Preliminaries

Let's assume the DM needs to gain values of importance of n alternatives, which he/she compares in terms of some criteria. Let's denote by $\mathbf{w}$ the PV, which contains true values of this relative importance:

$$\mathbf{w} = [w_1, w_2, \ldots, w_n] \tag{1}$$

We assume the $\mathbf{w}$ is normalized, i.e., it satisfies $\sum w_i = 1$.

In the pairwise comparisons technique, the DM does not give the values w_i directly, but he/she compares alternatives in pairs. So, an ideal PCM consists of the ratios of elements w_i:

$$\mathbf{M} = \begin{bmatrix} 1 & w_1/w_2 & \ldots & w_1/w_n \\ w_2/w_1 & 1 & \ldots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \ldots & 1 \end{bmatrix} \tag{2}$$

We call a matrix given by formula (2) the Matrix of Priority Ratios (MPR). However, such a matrix is usually not realizable.

In practice, the PCM consists of the items of used scale, which are approximations of the elements of the MPR at best (cf. [10]). The rating scale, which is usually used in the AHP, consists of 9 natural numbers and their inverses [13]:

$$S = \left\{ \frac{1}{9}, \frac{1}{8}, \ldots, 1, 2, \ldots, 9 \right\} \tag{3}$$

Therefore, it is assumed the elements of the PCM given by the DM satisfy:

$$a_{ij} = \frac{w_i}{w_j} + \delta_{ij} \tag{4}$$

An element δ_{ij} in (4) indicates the difference between true priority ratios and the nearest values from the rating scale. In reality, the elements of the PCM are estimation of priority ratios, so the factor δ_{ij} may also involve violations in DM assessments. However, in our simulation, we only investigate disruptions caused by the procedure itself (cf. [14, 15]).

Another weakness of AHP methodology is enforcing of the reciprocity, which means that the entries below the main diagonal of the PCM are calculated as the inverses of entries above the main diagonal. Although, in practice, the DM's opinions do not always satisfy this condition, and enforcing it leads to weaker results (cf. [10, 16]). Therefore, the DM is only asked for single comparisons, which are inserted into the upper triangle of the PCM, and the entries below the main diagonal are calculated by employing the formula:

$$a_{ji} = \frac{1}{a_{ij}} \tag{5}$$

The diagonal entries a_{ii}, which satisfy this equation, are equal to 1, of course. The PCM that satisfies equation (5) is called a reciprocal matrix. In our investigation, we take into account such reciprocal PCMs.

Let's denote by $\mathbf{A}$ the PCM given by the DM. Based on $\mathbf{A}$, the estimation of the PV can be calculated. Let's denote this approximation of vector $\mathbf{w}$ gained in the AHP by $\mathbf{v}$. Let's note the true PV given by formula (2) satisfies the equation:

$$\mathbf{M}\mathbf{w} = n\mathbf{w} \tag{6}$$

Therefore, in the AHP, the estimation of the PV is given as the principal right eigenvector of matrix $\mathbf{A}$, i.e., the vector that is associated with the maximal eigenvalue $\lambda_{\max}$ of matrix $\mathbf{A}$. Although this original method, called the right eigenvector method (REV), was introduced by the inventor of the AHP, it is usually displaced by Geometric Mean Procedure (GM), which is considered better by some authors [17, 18].

Among the various advantages of GM, it is more efficient in usage than the REV. The PV elements in the GM are obtained as normalized geometric means of the rows of the PCM:

$$v_i = \frac{\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}}{\sum_{i=1}^{n}\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}} \tag{7}$$

Both methods yield the same PV when the PCM is given by (2), but in practice, the obtained estimations by each method are usually a little different [19]. Each prioritization method also has both supporters and opponents. Although applying different methods may lead to slightly different results, the differences are not very meaningful, so in our investigation the GM is applied.

The estimation of a true PV, obtained in the AHP, is a good approximation of this vector when the PCM is close to the MPR (cf. [6, 17]). Although we do not know the true PV and we do not know the MPR in reality, there is the necessary and sufficient condition of the existence of such a unique PV. It is a consistency condition. The PCM satisfies the consistency condition when all triples of its elements a_{ij}, a_{ik}, a_{kj} satisfy equation (cf. [13]):

$$a_{ik} \cdot a_{kj} = a_{ij} \tag{8}$$

This requirement seems to be reasonable, but it is usually not satisfied [10, 20, 21]. Therefore, to verify the usability of PCMs, an inconsistency index (CI) is applied. When the value of the CI exceeds the established threshold, it is considered that the PCM should be corrected (cf. [22, 23]).

In literature, there are various definitions of the CIs [24]. Each CI ”measures”, in a specific way, the distance between the PCM given by the DM and a consistent version of it. Although some authors propose a set of axioms for measuring inconsistency [25, 26], in fact, the only requirement that the CI should meet is that it should be equal to 0 for a consistent matrix and positive in other cases [27, 28]. Most of CIs have no clear interpretation as a distance between the PCM and the MPR and between a true PV and its estimation (some remarks are proven in [3, 29]). However, simulations basically show a proper relationship between these values (cf. [12, 15, 28]).

In the AHP, a simple rule of PCM usefulness verification is used. The rule proposed by the inventor of the AHP is to accept the PCM when the CI value is less than the established threshold, which is calculated as 0.1 of the average value of the CI for random PCMs. One can also consider different thresholds for various sizes of PCM (cf. [3]). In spite of the critique received [3, 21], the rule of establishing a threshold has been used up to now. Therefore, in this paper, values of these thresholds are juxtaposed with the ROEs.

In our simulation, we take into account some CIs, whose definitions are presented in Table 1 (cf. [15,24,30]). There are definitions of 4 indices, i.e., Saaty's Index (SI), Koczkodaj's Index (KI), Average Triad Index (ATI), and Salo-Hamalainen Index (SHI). The SI is related to REV, but the remaining indices (KI, ATI, and SHI) are related to consistency definition (8). The triple of elements a_{ik}, a_{kj}, a_{ij} is called a triad in definitions of KI and ATI, and these definitions only differ in a function used to aggregate single triad inconsistency [15], so their values are the same for the size of a matrix $n = 3$ (only one triad is in such a matrix). Alternative formulations of the KI definitions are also documented (cf. [30]). Obviously, all presented definitions satisfy the necessary condition for the CI.

Table 1. Inconsistency indices definitions

	CI definition	Notes
SI	$\frac{\lambda_{\max} - n}{n-1}$	$\lambda_{\max}$ – maximal eigenvalue of the PCM
KI	$\max\limits_{i<k<j} \left\{ \frac{\lvert a_{ik}a_{kj} - a_{ij}\rvert}{\max\{a_{ik}a_{kj}, a_{ij}\}} \right\}$	$i < k < j$ means triples a_{ik}, a_{kj}, a_{ij} are from the upper triangle of the PCM
ATI	$\frac{1}{\binom{n}{3}} \sum\limits_{i<k<j} \frac{\lvert a_{ik}a_{kj} - a_{ij}\rvert}{\max\{a_{ik}a_{kj}, a_{ij}\}}$	like above, but the average value is calculated – there are $\binom{n}{3}$ triples
SHI	$\frac{2}{n(n-2)} \sum\limits_{i<j} \frac{\max\{r_{ij}\} - \min\{r_{ij}\}}{\left(1-\max\{r_{ij}\}\right)\left(1-\min\{r_{ij}\}\right)}$	$\{r_{ij}\} = \{a_{ik}a_{kj} \mid k = 1, \ldots, n\}$

All definitions presented in Table 1 are given for a reciprocal PCM, so they take into account only entries above the main diagonal. However, each definition, apart from SI, could be customized to no reciprocal matrix, but in our investigation we consider only reciprocal ones.

3. Simulation framework

Results presented in this paper are obtained by using the Monte Carlo experiment. The method is used since, in the AHP, values of the DM's preferences are unknown. In fact, we are not sure about the ”true” values of the evaluated preferences. Therefore, it is unavoidable to conduct a proper simulation to investigate the dependencies between true priorities and their estimations or other factors related to the AHP methodology. In our experiment, a true PV is assumed, and an impact of various factors on ultimate results is investigated (cf. [15,28]).

We take into account ”true” PVs consisting of $n = 3,4,\ldots,10$ elements, which are randomly generated. This PV is drawn from a uniform distribution over the range $[1/n;n]$ and then normalized. Next, we calculate the matrix of priority ratios (MPR), given by formula (2). The PCM, which is given by the DM in reality, is obtained by rounding elements of the MPR to the nearest value of the rating scale (3). We calculate values of the ROEs by comparing rounded PCM with the MPR. For the PCM, we also calculate values of indices presented in Table 1. Next, based on the PCM, the estimation of the PV is calculated with the use of GM. Finally, we yield values of errors in the PV by comparing values of PV estimation with the initially given ”true” PV (cf. [11,12]).

We analyze two types of errors in the PCM and in the PV: absolute error (AE) and relative error (RE). The absolute error is defined as the sum of the difference between the true and estimated value:

$$\mathrm{AE}(v) = \sum_{i=1}^{n} |v_i - w_i|; \qquad \mathrm{AE}(\mathbf{A}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \left|a_{ij} - m_{ij}\right| \tag{9}$$

As stated previously, the w_i means elements of a true PV given by (1), and m_{ij} means true priority ratios calculated as in the MPR (2). In turn, the RE is a sum of ratios of absolute error and true value of the PV or the PCM. It is defined as follows:

$$\mathrm{RE}(v) = \sum_{i=1}^{n} \frac{|v_i - w_i|}{w_i}; \qquad \mathrm{RE}(\mathbf{A}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\left|a_{ij} - m_{ij}\right|}{m_{ij}} \tag{10}$$

Using introduced notations, a framework of our simulation can be described by the following steps:

1. Randomly generating a true PV
2. Calculating the MPR
3. Rounding entries of the MPR to the rating scale – yielding the PCM
4. Calculating values of the CIs defined in Table 1
5. Obtaining a PV based on the PCM (using the GM)
6. Calculating the AE and RE for the PCM and for the PV
7. Recording values of the CIs, AE and RE in a database

The presented simulation steps have been performed $N = 10\,000$ times for each number of alternatives $n = 3,\ldots 10$. In each loop of simulation, a new ”true” PV with

n elements has been generated, and the following steps have been carried out. In such a way, for each n we obtained a database that contained 10000 records with values of four CIs and two types of errors in the PCM and the PV.

4. Results

Based on the database obtained in simulation, we calculate some statistics of the errors and the distributions of the CIs. We calculate the mean, maximum, and the minimum values of the errors that are presented in Tables 2 and 3. In Table 4, the average and maximum values of CIs related to ROEs are presented, and in Table 5 they are juxtaposed with consistency thresholds. Then, Figures 1-4 show the distribution of CI's values, using histograms, in which the values of consistency thresholds are also marked out.

The results gathered in Table 2 show us rather large values of errors caused by the rounding procedure. We can see that the average values of RE are always bigger than 10 %, and for $n > 6$, we observe values more than 20 %. Moreover, the maximum values of ROEs are almost always bigger than 100 %, so for each number of alternatives, the observed ROEs can visibly change values of priority ratios. We can also notice the values of AE and RE in whole PCMs increase along with the number of alternatives n. It is natural when we realize that the values of errors are obtained as the sum of all deviations, which occur in the matrix. However, this growth is not bigger than the growth of the size of matrices. When we take a look at the column that contains the maximum values of errors divided by the size of the matrix, we can observe only some fluctuations of these values. A similar observation can be made when we divide the values of the mean by the number n, so one can conclude the values of errors in the single matrix elements are the same or even decrease when the size of the matrix increases.

Table 2. Statistics of distribution of the AE and RE in PCMs rounded to rating scale

	AE				RE			
n	min	mean	max	max/n	min	mean	max	max/n
3	0.002	0.334	5.52	1.84	0.001	0.122	0.85	0,283
4	0.007	0.532	7.15	1.79	0.005	0.157	1.04	0,260
5	0.024	0.674	8.80	1.76	0.010	0.180	1.26	0,252
6	0.027	0.791	9.35	1.56	0.020	0.197	1.29	0,215
7	0.032	0.894	11.59	1.66	0.022	0.212	1.54	0,220
8	0.064	0.979	14.02	1.75	0.041	0.224	1.80	0,225
9	0.072	1.056	14.24	1.58	0.048	0.234	1.82	0,202
10	0.074	1.132	15.14	1.51	0.047	0.245	1.92	0,192

When we look at Table 3, we can make similar observations. The ROEs in the PCM entail the errors in the PV, so the maximum REs in PVs are always more than 100 %, and in some cases, they exceed 200 %. In turn, the values of maximum REs divided by n only for $n = 5$ exceed 50 %, and they are rather smaller for a bigger n, whereas the average values of REs do not exceed 17 %. The magnitude of the mean AE in Table 3 balances between 0.01 and 0.02, and the maximum values of the AE do not exceed 0.1. In turn, the maximum AEs per one alternative (divided by n) explicitly decrease along with the number n. These values may seem quite small, but we should remember the PV values are normalized, so the sum of all entries in the PV is always equal to 1. Therefore, the AE (for the whole matrix) close to 0.1 is quite big.

These observations align with the error analysis presented in [10], which demonstrated that the inaccuracy of the final results increases with the number of alternatives. However, in the same paper, the authors have shown the probability of the inaccuracy of each comparison related to the enforcing of reciprocity decreases along with the number of available comparisons. At this point it is worth noticing we use the interval $[1/n;n]$ for generating values of a "true" PV.

On the other hand, when we analyse values of errors divided by the number of alternatives, we can generally observe the decrease of the values. It is coherent with the restriction proposed for maximum values of errors in [3]. Although the average values in Table 3 do not exceed the upper estimate of permissible errors proposed in this work, the maximum values (even if they are divided by n) in most cases are larger than these limits.

Table 3. Statistics of distribution of the AE and RE in PVs yielded from PCMs rounded to rating scale

	AE				RE			
n	min	mean	max	max/n	min	mean	max	max/n
3	0.0001	0.0132	0.0725	0.0242	0.0002	0.070	1.033	0.344
4	0.0002	0.0154	0.0875	0.0219	0.0014	0.104	1.814	0.454
5	0.0004	0.0156	0.0971	0.0194	0.0029	0.126	2.595	0.519
6	0.0006	0.0147	0.0832	0.0139	0.0045	0.139	1.732	0.289
7	0.0005	0.0133	0.0806	0.0115	0.0042	0.149	1.760	0.251
8	0.0007	0.0119	0.0811	0.0101	0.0085	0.154	1.965	0.246
9	0.0004	0.0105	0.0631	0.0070	0.0046	0.158	2.268	0.252
10	0.0008	0.0094	0.0539	0.0054	0.0114	0.162	1.810	0.181

Table 4 introduces us to values of 4 CIs, whose definitions have been presented in Table 1. These values are obtained from rounded PCMs. Since the values of CIs are not normalized, they cannot be compared together – we can observe different magnitude levels for each CI (cf. [31]). However, we can compare the variation patterns of each CI, as well as the differences between their average and maximum values.

When we look at average values of the CI in Table 4, we can usually observe an increase in values as the number n grows. However, this increase is not the same for each index. We can observe bigger growth for smaller alternative numbers and smaller growth for a larger n (especially for 3 indices). On the other hand, for ATI, the average values decrease for the number of alternatives $n \geq 5$. Moreover, the maximum values of ATI obtained in simulation decrease along with a growing n. The maximum values of KI only increase as n grows for $n < 5$, and for $n \geq 5$ they are the same in turn. For the SI and the SHI, the visible growth of maximum values is observed for a small n too, and for a bigger n, we can observe more fluctuation than increase.

Table 4. An average and a maximum value of 4 CIs for PCMs rounded to rating scale

	SI		KI		ATI		SHI	
n	mean	max	mean	max	mean	max	mean	max
3	0.005	0.091	0.185	0.719	0.185	0.719	0.038	0.179
4	0.014	0.127	0.363	0.816	0.210	0.606	0.076	0.245
5	0.021	0.151	0.480	0.889	0.217	0.585	0.105	0.285
6	0.025	0.163	0.556	0.889	0.218	0.566	0.126	0.299
7	0.028	0.163	0.607	0.889	0.216	0.527	0.142	0.309
8	0.029	0.153	0.643	0.889	0.215	0.507	0.154	0.307
9	0.031	0.160	0.671	0.889	0.213	0.494	0.164	0.316
10	0.031	0.162	0.693	0.889	0.211	0.468	0.172	0.315

The next table shows us the values of CIs juxtaposed with thresholds of inconsistency used in the AHP. As it was mentioned, such thresholds are evaluated based on pure random PCMs, i.e., matrices, of which entries above the diagonal are randomly drawn from the rating scale (3). The thresholds are obtained as 0.1 of the gained average value of the CI (cf. [21, 32]). As we can see in Table 5, the thresholds are different for various indices and usually depend on matrix size n. However, as in the previous table, we can also observe similarities in the values of the KI and the ATI now. For the first one, the values of thresholds are the same or almost the same for $n \geq 5$, and for the second one, we can observe the same or almost the same values for all presented sizes of matrix n.

Table 5. A comparison of average and maximum value of 4 CIs with their consistency thresholds

n		**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
SI	mean	0.005	0.014	0.021	0.025	0.028	0.029	0.031	0.031
	max	0.091	0.127	0.151	0.163	0.163	0.153	0.160	0.162
	thresh.	**0.053**	**0.088**	**0.111**	**0.125**	**0.134**	**0.141**	**0.145**	**0.149**
KI	mean	0.185	0.363	0.480	0.556	0.607	0.643	0.671	0.693
	max	0.719	0.816	0.889	0.889	0.889	0.889	0.889	0.889
	thresh.	**0.078**	**0.095**	**0.098**	**0.099**	**0.100**	**0.100**	**0.100**	**0.100**
ATI	mean	0.185	0.210	0.217	0.218	0.216	0.215	0.213	0.211
	max	0.719	0.606	0.585	0.566	0.527	0.507	0.494	0.468
	thresh.	**0.078**	**0.078**	**0.077**	**0.077**	**0.077**	**0.077**	**0.077**	**0.077**
SHI	mean	0.038	0.076	0.105	0.126	0.142	0.154	0.164	0.172
	max	0.179	0.245	0.285	0.299	0.309	0.307	0.316	0.315
	thresh.	**0.038**	**0.057**	**0.069**	**0.076**	**0.081**	**0.084**	**0.087**	**0.088**

In fact, the comparisons of CIs' values with proper thresholds are the most interesting. When we look at the average values of SI, we can observe that they are always much smaller than the values of its thresholds. Although the maximum values of this index always exceed the proper threshold, the differences depend on n. The maximum value of SI in Table 5 is more than 70 % bigger than the threshold for $n = 3$ and less than 10 % for $n = 10$. Nevertheless, based on the obtained results, we conclude that correct matrices (which only contain errors caused by the rounding procedure) may sometimes be rejected in the AHP because they exceed the consistency threshold (cf. [15, 28]).

Actually, the results seem to be even more pessimistic when we compare values of other CIs with their thresholds. For KI, ATI, and SHI, not only maximum values but also average values exceed proper thresholds. For KI, the average value is more than two times bigger than the proper threshold for $n = 3$ and almost seven times bigger for $n = 10$. The values for ATI are more balanced, so we observe no more than three times bigger values of the mean of the ATI than values of the threshold. In turn, the average value of SHI for $n = 3$ is equal to the proper threshold, but for $n > 3$ we observe the mean values bigger than thresholds up to two times in Table 5. Obviously, the maximum values of these indices are much bigger than the average values, so they exceed the proper thresholds as well. Therefore, in the case of these indices, the proposed thresholds entail rejection of the majority of correct PCMs, which only contain ROEs.

The visible difference between the range of CIs' values below and above the threshold can be seen in Figures 1-4. These figures show us a distribution of CIs' values for two numbers of alternatives: $n = 4$ (on the left) and $n = 7$ (on the right). On the axis OX, one can observe the values of particular indices, and on the axis OY, the frequencies of this value are presented. As it has been mentioned, 10 000 matrices of each size have been generated, so the values on the OY axis means the ratio of the number of matrices, whose inconsistency value belongs to a particular subinterval (divided by 10 000). We can observe more regular shapes for bigger sizes of PCMs. A smooth, regular shape is especially observed for $n = 7$ for indices SI, ATI, and SHI. However, for these indices, we do not observe the same shape. A histogram of SI distribution has a unimodal, right-skewed shape, while the histograms for ATI and SHI are rather bell-shaped with the peak a little shifted to the left. The histograms for $n = 4$ have similar properties but are more irregular than for $n = 7$. Instead, the histograms for KI do not have a regular shape, but for $n = 4$ we can observe a few peaks, and for $n = 7$, there is one peak and a rather flat shape all around.

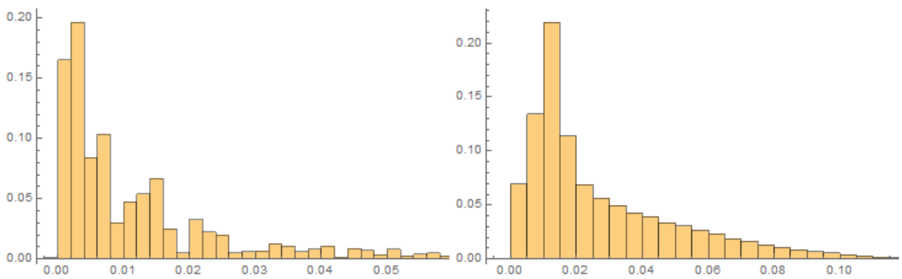

Fig. 1. Histograms of the values of SI for PCMs rounded to scale for $n = 4$ (left) and $n = 7$ (right). The values of the index are marked on the axis OX, and their frequencies are on the axis OY

On each histogram, in Figures 2-4, the threshold points for proper CIs have been marked out. There are no such points in Figure 1, since thresholds for the SI are

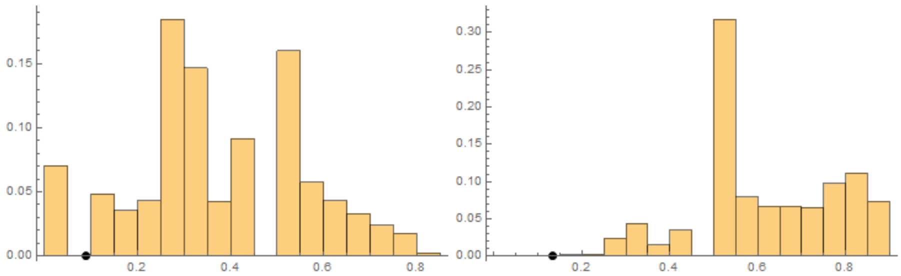

Fig. 2. Histograms of the values of KI for PCMs rounded to scale for $n = 4$ (left) and $n = 7$ (right). The values of the index are marked on the axis OX, and their frequencies are on the axis OY

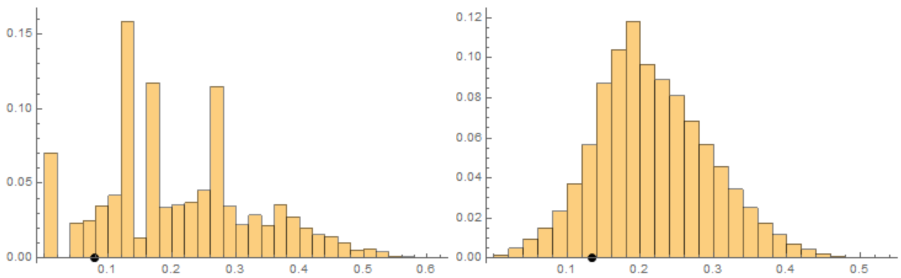

Fig. 3. Histograms of the values of ATI for PCMs rounded to scale for $n = 4$ (left) and $n = 7$ (right). The values of the index are marked on the axis OX, and their frequencies are on the axis OY

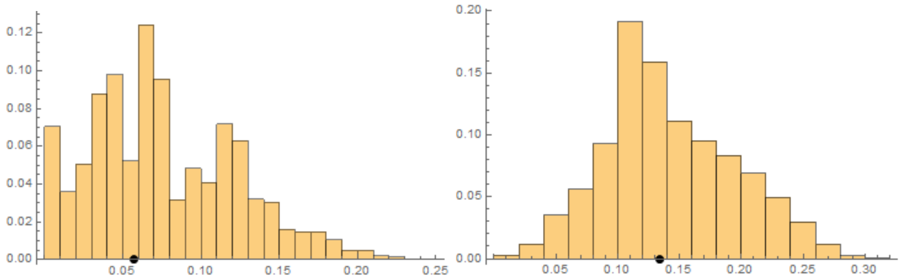

Fig. 4. Histograms of the values of SHI for PCMs rounded to scale for $n = 4$ (left) and $n = 7$ (right). The values of the index are marked on the axis OX, and their frequencies are on the axis OY

bigger than the end points of the histograms' range, which have been automatically matched (some outliers are omitted at these histograms). This indicates that indeed, the most of the values of the SI are below thresholds. However, we should remember the maximum values of this index, which are invisible at the histograms, are always bigger than thresholds.

A completely different situation can be observed in the remaining figures. For KI, only a small part of the histogram is below the threshold for $n = 4$, and all results are above the threshold for $n = 7$. A similar situation presents us histograms in figure 3, where only a small part of results is located below thresholds. More parts of the histograms are placed below thresholds in figure 4, but on these histograms, we also observe a definitely bigger part of the results placed above threshold. Moreover, the main peaks of histograms in Figures 2-4 are placed above thresholds, so the values of indices KI, ATI, and SHI are more frequently above than below presented thresholds.

As it has been explained, the histograms show us values of CIs for "error-free" PCMs, which have only been rounded to scale (3). The only errors that they contain are ROEs, which are unavoidable in the AHP. However, on the presented histograms, we notice the values of most indices usually far exceed these thresholds. This is a bothersome observation, because when we use such thresholds in the AHP, we have to reject or at least improve correct (!) PCMs (cf. [22, 33]).

5. Conclusions

In this article, we have investigated the influence of the errors caused by the limitation of available comparisons in the AHP by its criticized rating scales. Moreover, the enforcement of reciprocity has also been applied. The impact of this procedure on the consistency of PCMs as well as on the quality of final priorities has been analysed. We have calculated some characteristics of quality indicators distribution and prepared some diagrams of obtained distributions.

Presented tables concerning errors in PCMs and in PVs have demonstrated that ROEs have significant meaning in the AHP methodology. The average values of considered errors usually exceed dozens of percent of correct values, and maximum errors exceed a hundred percent of these values. The values are also substantial when they are calculated for each alternative (divided by n). Although such errors are unavoidable in the AHP procedure, the obtained results show the significance of them for AHP outcomes. Although they would be reduced by using other rating scales, what has been shown in [11, 12], the problem is also in human cognitive resolution. The obtained values sometimes also exceed the values that have been indicated in [3] as a guarantee of good properties of the methodology.

In practice, the errors in PCMs and in PVs can not be observed in the AHP. Therefore, the CIs are used as indicators of PCMs' quality, and the consistency thresholds determine if a PCM can be used in further process of obtaining PVs. The proposed methods for establishing a threshold are not based on proven theory (e.g., [3, 28, 32]).

AHP methodology uses 0.1 for inconsistency as a heuristic since no formal theory exists to support it. This threshold has been originally formulated for SI. Results presented in this paper show inappropriate relationships between such consistency thresholds and values of various CIs for correct PCMs.

Since the PCMs entries are limited to the scale, even if the DM perfectly assesses priority ratios, the PCM would be inconsistent. Although it seems obvious that a perfectly filled-out PCM, which only contains ”errors” related to the AHP procedure itself, should be accepted, it turns out this does not always occur. In this paper, average and maximum values of 4 indices are presented, among which only average values of one index (i.e., the SI) are always below consistency thresholds. For remaining indices, both average and maximum values are usually much bigger than proper thresholds. This observation shows us the uselessness of the originally proposed establishing technique of the consistency threshold. The same conclusions can be made on the basis of histograms, which show us the distribution of CIs' values for randomly generated perfect PCMs. Although the shape of the histogram differs for various CIs, most of the obtained values are placed above consistency thresholds, which have been marked out on these diagrams.

To sum up, our results show that the manner of obtaining consistency thresholds in the AHP does not work in case of the presented indices ATI, KI, and SHI, and for the SI it should be cautiously used. The better way of determining such thresholds would be using the maximum value of correct PCMs (rounded to rating scale). Such maximum values have been presented in Tables 4 and 5. Other propositions of the threshold would be values calculated on the basis of average CIs (for rounded PCMs), which have also been presented, or other statistics of distribution of CIs obtained for randomly generated, correct (rounded) PCMs. We can also consider values of thresholds based on other types of somehow correct matrices – for example, matrices satisfying the transitivity condition. Such a simulation would be presented in the future. Certain other propositions of establishing the usefulness of PCMs and their limitations are also presented in extensive literature (e.g., [3, 10, 21, 28, 32]).

References

[1] Saaty, T.L. (2004). Decision making – the Analytic Hierarchy and Network Processes (AHP/ANP). *Journal of Systems Science and Systems Engineering*, *13*(1), 1-35.

[2] Ho, W. (2008). Integrated analytic hierarchy process and its applications – A literature review. *European Journal of Operational Research*, *186*, 211-228.

[3] Tomashevskii, I.L. (2015). Eigenvector ranking method as a measuring tool: Formulas for errors. *European Journal of Operational Research*, *240*, 774-780.

[4] Greco, S., Słowiński, R., & Wallenius, J. (2024). Fifty years of multiple criteria decision analysis: From classical methods to robust ordinal regression. *European Journal of Operational Research*, *323*, 351-377, DOI: 10.1016/j.ejor.2024.07.038.

[5] Thurstone, L. (1927). A law of comparative judgments. *Psychological Review*, *34*, 273-286.

[6] Saaty, T.L. (1980). *The Analytic Hierarchy Process*. NewYork: McGraw Hill.

[7] Dong, Y., Xu, Y., Li, H., & Dai, M. (2008). A comparative study of the numerical scales. *European Journal of Operational Research*, *186*, 229-242.

[8] Kułakowski, K. (2020). *Understanding the Analytic Hierarchy Process*. Boca Raton: Chapman and Hall – CRC Press.

[9] Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97.

[10] Triantaphyllou, E., & Yanase, J. (2024). The use of pairwise comparisons for decision making may lead to grossly inaccurate results. *Computers & Industrial Engineering*, *198*, 110653.

[11] Grzybowski, A.Z., & Starczewski, T. (2018). Simulation analysis of prioritization errors in the AHP and their relationship with an adopted judgement scale. *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2018*, 547-551.

[12] Grzybowski, A.Z., & Starczewski, T. (2019). Analysis of errors in priority vector estimation and their relationship with the correctness of the final ranking of decision alternatives. *IAENG International Journal of Applied Mathematics*, *49*(3), 381-386.

[13] Brunelli, M. (2015). *Introduction to the Analytic Hierarchy Process*. SpringerBriefs in Operations Research. P. 83. 978-3-319-12502-2 (electronic), DOI: 10.1007/978-3-319-12502-2.

[14] Zahedi, F. (1986). A simulation study of estimation methods in the analytic hierarchy process. *Socio-Economic Planning Sciences*, *20*(6), 347-354.

[15] Grzybowski, A.Z. (2016). New results on inconsistency indices and their relationship with the quality of priority vector estimation. *Expert Systems with Applications*, *43*, 197-212.

[16] Hovanov, N.V., Kolari, J.W., & Sokolov, M.V. (2008). Deriving weights from general pairwise comparison matrices. *Mathematical Social Sciences*, *55*, 205-220.

[17] Lin, C.-C. (2007). A revised framework for deriving preference values from pairwise comparison matrices. *European Journal of Operational Research*, *176*, 1145-1150.

[18] Dijkstra, T.K. (2013). On the extraction of weights from pairwise comparison matrices. *Central European Journal of Operations Research*, *21*, 103-123.

[19] Ishizaka, A., & Lusti, M. (2006). How to derive priorities in AHP: a comparative study. *Central European Journal of Operations Research*, *14*, 387-400.

[20] Temesi, J. (2011). Pairwise comparison matrices and the error-free property of the decision maker. *Central European Journal of Operations Research*, *19*, 239-249.

[21] Cavallo, B. (2017). Computing random consistency indices and assessing priority vectors reliability. *Information Sciences*, *420*, 532-542.

[22] Ishizaka, A., & Siraj, S. (2020). Interactive consistency correction in the analytic hierarchy process to preserve ranks. *Decisions in Economics and Finance*, *43*, 443-464.

[23] Finan, J.S., & Hurley, W.J. (1997). The Analytic Hierarchy Process: does adjusting a pairwise comparison matrix to improve the consistency ratio help? *Computers & Operations Research*, *24*(8), 749-755.

[24] Brunelli, M. (2018). A survey of inconsistency indices for pairwise comparisons. *International Journal of General Systems*, *47*(8), 751-771.

[25] Koczkodaj, W.W., & Szwarc, R. (2013). On axiomatization of inconsistency indicators for pairwise comparisons. *Fundamenta Informaticae*, *132*(4), 485-500.

[26] Koczkodaj, W.W., & Urban, R. (2018). Axiomatization of inconsistency indicators for pairwise comparisons. *International Journal of Approximate Reasoning*, *94*, 18-29.

[27] Grzybowski, A.Z., & Starczewski, T. (2019). Comparative study of inconsistency indices based on a notion of a triad inconsistency. *Proceedings of Informatics 2019, IEEE 15th International Scientific Conference on Informatics*, Poprad, Slovakia.

[28] Grzybowski, A.Z., & Starczewski, T. (2020). New look at the inconsistency analysis in the pairwise-comparisons-based prioritization problems. *Expert Systems with Applications*, *159*, 113549.

[29] Kułakowski, K. (2015). On the properties of the priority deriving procedure in the pairwise comparisons method. *Fundamenta Informaticae*, *139*(4), 403-419.

[30] Koczkodaj, W.W., & Szybowski, J. (2016). The limit of inconsistency reduction in pairwise comparisons. *International Journal of Applied Mathematics and Computer Science*, *26*(3), 721-729.

[31] Koczkodaj, W.W., Magnot, J.-P., Mazurek, J., Peters, J.F., Rakhshani, H., Soltys, M., Strzałka, D., Szybowski, J., & Tozzi, A. (2017). On normalization of inconsistency indicators in pairwise comparisons. *International Journal of Approximate Reasoning*, *86*, 73-79.

[32] Amenta, P., Lucadamo, A., & Marcarelli, G. (2020). On the transitivity and consistency approximated thresholds of some consistency indices for pairwise comparison matrices. *Information Sciences*, *507*, 274-287.

[33] Linares, P. (2009). Are inconsistent decisions better? An experiment with pairwise comparisons. *European Journal of Operational Research*, *193*, 492-498.