An urgent issue of modern football competitions is the detection of fixed matches. Known methods for predicting the outcome of a match by analyzing bets on a match or analyzing the actions of football players on the field use a large amount of data that is not always available. To overcome this obstacle, there can be applied a method for detecting suspicious fixed match results based on conformal predictors and power martingales, which uses publicly available public data. But in practice, this method does not always detect such matches with high precision. An improved method for determining a suspicious match is proposed, based on the theory of conformal predictors using a modified Stepanets indicator function, which is compared with a threshold. The modified Stepanets indicator function is applied to the power martingale and shows the relative change in the martingale value of the current match compared to the previous match. The threshold value was determined experimentally according to the criterion of the maximum of the F1 metric. Data from the 2013–2014 season of the French II League were used as a training sample, and data from the 2014–2015 season of Serie B in Italy were used as a test sample. Team clustering was performed on all samples. For each of the formed classes of matches on both samples, the measure of non-conformity, the degree of non-conformity, the power martingale, and the modified Stepanets indicator function were calculated. The resulting indicators of precision metrics and F1 are higher (average values of metrics P=0.84, F1=0.87) than the same indicators of martingale and p-value rules (average values of metrics P=0.75, F1=0.78), applied to the same data. The proposed method reveals 4 out of 5 matches of the 2014–2015 Serie B season in Italy, which are considered fixed according to information from official Italian law enforcement sources

Keywords: fixed result, power martingale, measure of non-conformity, offline algorithm, p-value, F1 metric

UDC 004.852
DOI: 10.15587/1729-4061.2023.282645

DETECTION OF FIXED FOOTBALL MATCHES BASED ON THE THEORY OF CONFORMAL PREDICTORS USING THE MODIFIED STEPANETS INDICATOR FUNCTION

Oleg Chertov
Corresponding author
Doctor of Technical Sciences, Professor*
E-mail: chertov@i.ua
Ivan Zhuk
Postgraduate Student*
*Applied Mathematics Department
National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»
Peremohy ave., 37, Kyiv, Ukraine, 03056

Received date 04.04.2023
Accepted date 08.06.2023
Published date 30.06.2023

How to Cite: *Chertov, O., Zhuk, I. (2023). Detection of fixed football matches based on the theory of conformal predictors using the modified Stepanets indicator function. Eastern-European Journal of Enterprise Technologies, 3 (4 (123)), 22–32. doi: https://doi.org/10.15587/1729-4061.2023.282645*

1. Introduction

Today, the field of sports competitions is becoming an area for looking for anomalies, in particular, due to the simplification of the possibility to make bets on sports and the global spread of the activities of bookmakers [1]. One of the tasks in this area is the search for matches suspicious of a fixed result (match-fixing, hereinafter MSFR), which can be interpreted as an anomaly. Such matches are a problem of sports competitions no less than the problem of doping among athletes [2]. The importance of this task is due, on the one hand, to the risk of uncontrolled illegal earnings from betting, and on the other hand, to the risk of public distrust of both individual teams, sports competitions, and sports as a whole [1]. A competition that is not considered authentic cannot expect continued public support [1]. In the case when the fixed result of the match is the result of an agreement between a certain organization and the main players of the team, the fact of a possible agreement can be traced through financial transactions or an unnatural distribution of the volumes of bets on the corresponding results of the match [1, 3]. But these data are not always publicly available. Therefore, research aimed at finding potentially suspicious results of sports matches based on the processing of exclusively publicly available public data should be considered relevant.

2. Literature review and problem statement

A separate category of research in the field of sports, in particular football, competitions are mathematical methods of identifying sports matches with a fixed result. The search for such anomalies in sports competitions, at least in the field of football, is reduced to predicting the outcome of the match, and analysis of bets [4–6].

Statistical methods [5] (in particular, Bayesian networks) and machine learning methods [6] are used to predict the outcome of a football match. These methods can be used to identify «anomalies» in match results. But their disadvantage is the need to use a significant number of match attributes, which are not always available, and the lack of an opportunity to obtain analytical regularities for predicting the result. These shortcomings make it necessary to devise methods that do not require a large amount of non-publicly available data.

Methods based on betting analysis are also used to detect matches with a fixed result [4]. If during a match the difference between the actual betting volume and the predicted volume is statistically significant, the match is considered fixed. However, a recent study [7] showed that most cases of match-fixing do not involve betting.

To analyze the properties of downward convex sequences and power functions, in particular, their convergence to zero, a special α-function-indicator was proposed in [8]. In [9], the inverse α-function-indicator was used to compare ordered patterns of results of football teams obtained during the season. It is shown that in many cases this approach allows better detection of deviations from the standard behavior of the team during the season than a direct comparison of it with the current behavior of the team.

In the monograph [10], a mathematical apparatus of conformal predictors and power martingales was built for the solution of anomaly detection. The advantages of this mathematical apparatus involve the combination of the process of learning and forecasting in one stage and, as a result, in using the results of decision-making on previous objects to predict the result for the current object. In particular, in [11] it is shown how exactly this mathematical apparatus can be practically used to improve the results of classification by methods of support vectors and nearest neighbors and to detect changes in data flows.

In [12], it is proposed to use the apparatus of conformal predictors and power martingales to find MSFR of football tournaments. Since the precision metric of the martingale rule proposed in the study [12] for some classes of matches was less than 70 %, there is a need to improve the developed rule. Therefore, it is expedient to conduct a study aimed at improving the precision of the detection of MSFR based on the theory of conformal predictors due to the use of the above-mentioned Stepanets indicator function.

3. The aim and objectives of the study

The purpose of the research is to identify MSFR in football tournaments based on the theory of conformal forecasting using the modified Stepanets indicator function and data only on the results of the matches. This will make it possible to increase the quality indicators of MSFR detection.

To accomplish the aim, the following tasks have been set:

– to analyze the application of the Stepanets indicator function to the power martingale and perform its modification;

– to devise an improved method for detecting matches suspicious of a fixed result based on the theory of conformal predictors using the modified Stepanets characteristic;

– to investigate and evaluate the effectiveness of the devised method for detecting football matches suspicious of a fixed result on real data by applying classification metrics that are basic in machine learning.

4. The study materials and methods

4. 1. The object and hypothesis of the study

The object of our study is the process of identifying suspicious football matches regarding the fixed result.

The main idea of the method is to detect MSFR by recording the corresponding changes in the value of the power martingale using the modified Stepanets characteristic. The martingale value for a match is formed based on the results of the conformal predictor for all matches of the season. The conformal predictor determines the probability of how close the current match corresponds to the match group to which it was assigned according to the classification of participating teams. The classification of participating teams and the calculation of the specified probability is based on data on the success of the participating teams and the results of all matches of the season. Match data is obtained from an open resource [13] related to soccer betting research.

4. 2. Method for detecting matches suspected of a fixed result based on the theory of conformal predictors

The method for detecting MSFR proposed in [12] consists of four actions that must be performed for each new z_k match:

1) the measure of non-conformity (difference) a_k is calculated, which is the first step in the calculation of the conformal predictor:

$$a_k = 1.5^{1-\operatorname{sgn}\left(avg_{k(i,j)}*(\alpha_k-\beta_k)\right)}\left|\left(\alpha_k-\beta_k\right)-avg_{k(i,j)}\right|, \quad (1)$$

where the function sgn(x) indicates the sign of the number x and is equal to 1 if $x>0$, 0 if $x=0$, and –1 if $x<0$, the characteristic $avg_{k(i,j)}$ is the arithmetic mean difference of goals scored by home and away teams in all class matches. The $avg_{k(i,j)}$ characteristic is considered the expected result of the match in the class of matches and is calculated according to the formula:

$$avg_{k(i,j)} = \underset{class(z_k)=class(z_l)}{\operatorname{mean}}\left\{\alpha_l-\beta_l\right\}, \quad (2)$$

where k is the number of the current match in chronological order; l is the number of another match that belongs to the same class as the current match; i and j are, respectively, the rank of the home and away teams of the current match. The calculation of characteristic (2) is carried out for the entire set of matches z_k in the class (i, j), that is, the entire set of observations is considered known and available for research in its entirety. Because of this, the proposed MSFR detection algorithm is termed an offline algorithm.

The result of the match in the format of a victory of the home team, a draw, or a victory of the away team will be hereafter called the character of the result of the match.

The measure of non-conformity a_k takes into account the absolute results of the teams during the match, the difference between the actual and expected results of the match;

2) the degree of non-conformity p_k of the match z_k of the set of observations $\{z_1, z_2,..., z_{k-1}, z_k\}$is calculated:

$$p_k = p\left(z_1, z_2, \ldots, z_k\right) = \frac{\#\left\{i: a_i \ge a_k, 1 \le i \le k\right\}}{k}, \quad (3)$$

where the operation $\#\{i: a_i \ge a_k, 1 \ge i \ge k\}$ returns the number of matches whose degree of non-conformity (1) is not less than the degree of non-conformity of the match z_k;

3) a modified power martingale $M_k^{(\varepsilon)}$ is calculated for an arbitrary value $\varepsilon \in [0; 1]$:

$$M_k^{(\varepsilon)} = \varepsilon \prod_{i=1}^{k} p_i^{\varepsilon-1}; \quad (4)$$

4) the MSFR set **S** is formed on the basis of the value of the modified power martingale $M_k^{(\varepsilon)}$ (martingale rule) and the degree of non-conformity p_k (p-value rule) of matches z_k, respectively, according to the formulas:

$$\mathbf{S} = \left\{z_k \mid M_k^{(\varepsilon)} > M_{k-1}^{(\varepsilon)}\right\}, \quad (5)$$

$$\mathbf{S}=\left\{z_k \mid \left(p_k < p_{k-1}\right) \vee \left(\left(z_{k-1}\in S\right)\wedge\left(p_{k-1}<p_k+\Delta\right)\right)\right\}. \qquad (6)$$

where the sign $\vee$ is a disjunction (logical operation OR), the sign $\wedge$ is a conjunction (logical operation AND); $\Delta\in(0;\,1)$ is the permissible difference between the degree of conformity p_{k-1} of the previous and p_k of the current object.

The evaluation of research results is performed according to the principle of positive-negative binary classification problems, where a positive match is a match for which the «Potentially Suspicious» characteristic is equal to 1. The metrics of precision (P), recall (R), and measure F_1 are used as performance indicators:

$$P=\frac{TP}{TP+FP}, \qquad (7)$$

$$R=\frac{TP}{TP+FN}, \qquad (8)$$

$$F_1=\frac{2}{\frac{1}{P}+\frac{1}{R}}=\frac{2TP}{2TP+FP+FN}. \qquad (9)$$

The value TP (true positives) is equal to the number of matches that are potentially suspicious, and the algorithm detected them as such. The value of FP (false positives) is equal to the number of matches that are not potentially suspicious but were detected as such by the algorithm. The value FN (false negatives) is equal to the number of matches that are potentially suspicious but were mistakenly missed by the algorithm.

Similar to [12], an offline algorithm is applied, that is, an algorithm in which the match result is calculated based on all matches in the class. However, the calculation of the degree of non-conformity (3) was carried out based on the degree of non-conformity of all matches in the class, and not only those matches that have already ended before the current one, that is:

$$p_k=p\left(z_1,\,z_2,\ldots,z_k,\ldots,\,z_N\right)=\frac{\#\left\{i: a_i\geq a_k, 1\leq i\leq N\right\}}{k}, \qquad (10)$$

where N is the number of matches in class (i, j).

4. 3. Stepanets indicator function

To study the dependence of martingale for different classes of matches, it was proposed to use the function-indicator of Stepanets [8]:

$$\alpha\left(x; f(x)\right)=\frac{f(x)}{x\,|f^{'}(x)|}. \qquad (11)$$

In [8], this indicator function was used to analyze the properties of downward convex sequences and power functions.

This characteristic was used, in particular, for the analysis of the rate of convergence to zero of downwardly convex functions $f(x)$ by comparing these functions with power functions x^n, $n\in\mathbf{R}$ and further classifying the function $f(x)$ according to the value of this characteristic. For downward convex functions of form $f(x)=Cx^n$, the characteristic $a(x, f(x))=1/n$.

In work [9], the Stepanets characteristic $\tilde{\alpha}\left(x; f(x)\right)$ was used to compare ordered patterns of results of football teams obtained during the season:

$$\tilde{\alpha}\left(x; f(x)\right)=\frac{x\,|f'(x)|}{f(x)}. \qquad (12)$$

It is shown that in some cases this approach makes it possible to better detect deviations from the standard behavior of the team during the season than a direct comparison of the standard pattern of team behavior with the current behavior of the team [9].

5. Results of the study of an improved method of detecting fixed matches using a modified characteristic of Stepanets

5. 1. Results of the study of the modified Stepanets function-indicator

During the study of martingale charts for different classes of matches on the data of the 2013–2014 season of the French II League [13], a property related to the change in the value of the martingale was established (Fig. 1). At points that corresponded to suspicious fixed matches, the value of the martingale increased significantly faster than at points that corresponded to regular matches but were defined as suspicious by the martingale rule.

In Fig. 1, correctly detected matches (TP) are highlighted in green, and falsely detected matches (FP) are highlighted in red. In the vast majority of points that correspond to correctly detected matches, the derivative is larger than in points that are close to correctly detected and correspond to false detections. For example, for the class of matches (1, 1) (Fig. 1, *a*) at points $k\in\{4, 8, 9\}$, which correspond to potentially suspicious matches, the magnitude of the change in the value of the martingale is greater than at point $k=11$, which corresponds to a falsely detected match. For class (2, 2) (Fig. 1, *b*) at points $k\in\{5, 8, 10, 13, 15, 16, 19, 20\}$, which correspond to potentially suspicious matches, the magnitude of the change in the martingale value is greater than at points $k\in\{6, 7, 9, 12\}$, which corresponds to a falsely detected match.

It is proposed to use the characteristic $\tilde{\alpha}\left(x; f(x)\right)$ also for the task of searching for MSFR.

From a practical point of view for data analysis, this characteristic is interesting since its value depends on the ratio of the value of the derivative $f'(x)$ to the value of function $f(x)$ at point x. Additionally, this ratio is scaled by multiplying by the value of point x. The ratio of the value of the derivative $f'(x)$ to the value of function $f(x)$ can be interpreted as the relative rate of change of the value o function $f(x)$ at point x. If x is considered an observation number, and function $f(x)$ is considered a characteristic of a certain process, then the characteristic $\tilde{\alpha}\left(x; f(x)\right)$ is the value of the relative rate of change of function $f(x)$. This value is additionally weighted by the observation number: the more recent (larger) the observation number x is, the more weight the value of the relative rate of change of function $f(x)$ has.

Since the current study uses an offline algorithm, i.e., an algorithm in which all observations are assumed to be known and equal, weighting more relevant observations is not appropriate in this case. Therefore, as part of the current study, this characteristic was modified and represented in the following form:

$$s\left(x; f(x)\right)=C\,\frac{f'(x)}{f(x)}, \qquad (13)$$

where the number C is a certain weighting factor that is constant for all observations of x.

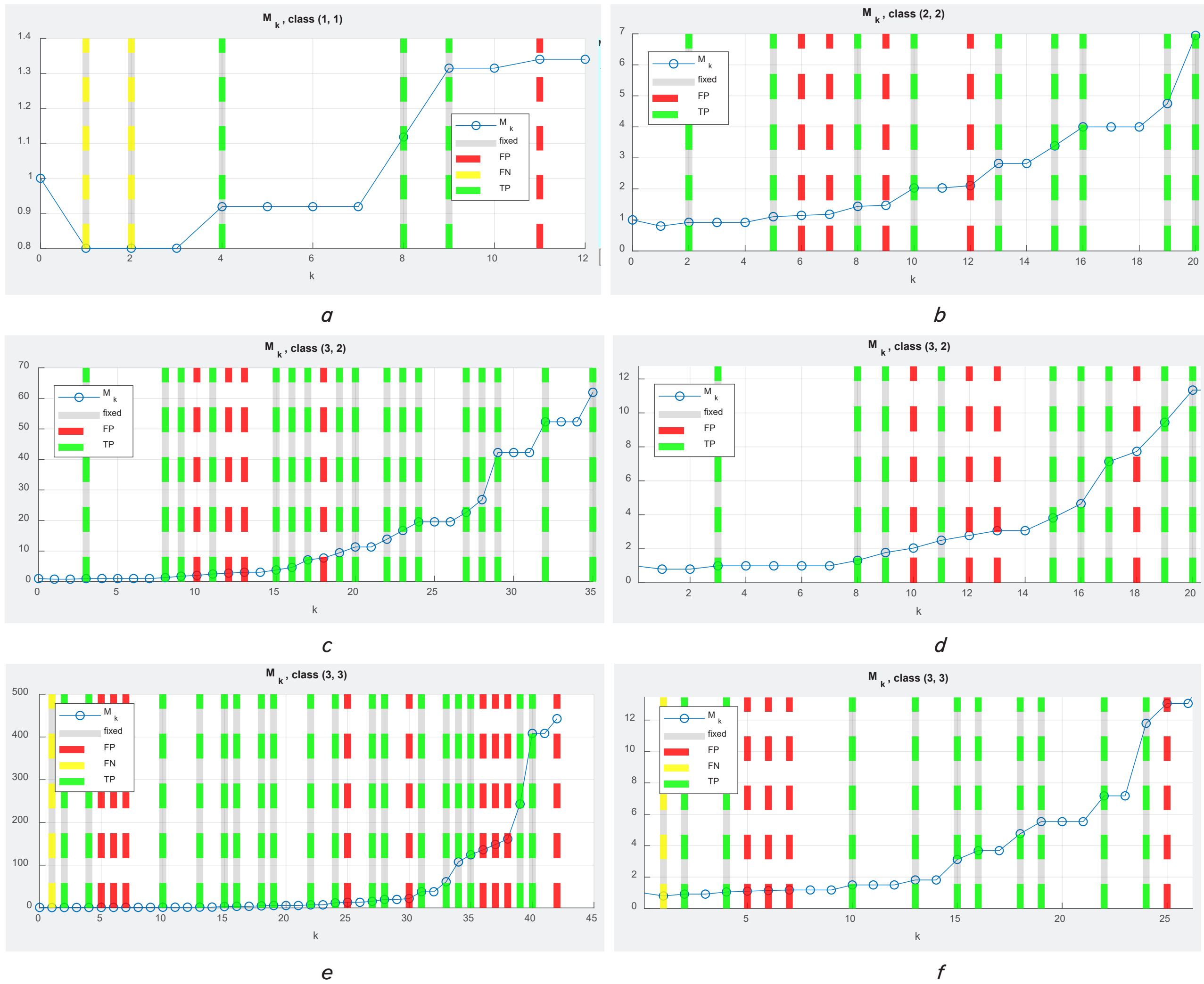

Fig. 1. Chart of characteristics M_k for several classes of matches of the 2013–2014 season of the French II League: *a* – for the class of matches (1, 1); *b* – for the class of matches (2, 2); *c, d* – for the class of matches (3, 2); *e, f* – for the class of matches (3, 3)

This coefficient is introduced for the same weighting of the values of characteristic $s(x;\ f(x))$ regardless of the observation number. A modulus was removed from the derivative of function $f'(x)$ to take into account not only the magnitude of the relative change of the function $f(x)$ but also the direction of its change.

5. 2. Improved method for detecting matches suspected of a fixed result using the modified Stepanets characteristic

The modified Stepanets function (13) was applied to martingale M_k. Since, in the framework of this study, the martingale is a function of a discrete value k, which is the observation number, then $f(x)=M_k$.

The martingale M'_k derivative is calculated approximately using the difference scheme:

$$M'(j) \approx \frac{M(j)-M(j-1)}{j-(j-1)} \approx M(j)-M(j-1). \qquad (14)$$

Characteristic $s(k;\ M(k))$ is also further normalized to reduce its value in each match class to a common range, according to the following formula:

$$s_N(k;M(k)) = \frac{s(k;M(k))}{\max\limits_k \{s((k;M(k))\}}. \qquad (15)$$

The following threshold rule with the threshold S_ε was proposed for the classification of MSFRs with numbers k into the set of MSFR **S** in each class of matches:

$$k \in \mathbf{S} \leftrightarrow s_N(k;M(k)) > S_\varepsilon. \qquad (16)$$

According to this rule, a match suspicious of a fixed result is considered a match with the number k, for which the value of the normalized modified Stepanets indicator function (15) exceeds the threshold value S_ε.

The search for the threshold S_ε of the decisive rule (16) was carried out on the basis of the 2013–2014 French League II season as a training season. Using the data of this season, the ranges for the S_ε threshold, at which the highest values of the F_1 metric are achieved, were obtained. The search was performed separately on each class of matches. The search results are given in Table 1. Cells in the columns of metrics P, R and F_1 have a range of four-color coloration. Cells with values from the range [0.75; 0.9) – yellow, and green – with values from the range [0.9; 1]. When calculating martingale (4), the value of the parameter $\varepsilon=0.8$ was used.

Table 1

Ranges of the best S_ε threshold values for the threshold rule (16) by each class of matches in the 2013–2014 French League II season

Class	[min (S_ε); max (S_ε)]	Number of cases			Metrics		
		TP	FP	FN	*P*	*R*	*F*1
(1, 1)	[0.1179; 0.6074]	5	0	0	1	1	1
(1, 2)	[0; 0.2110]	8	0	1	1	0.89	0.94
(1, 3)	[0; 0.3232]	14	0	1	1	0.93	0.97
(1, 4)	[0.2365; 0.4202]	6	0	0	1	1	1
(2, 1)	[0.2753; 0.5063]	7	0	0	1	1	1
(2, 2)	[0.0722; 0.211]	9	0	0	1	1	1
(2, 3)	[0.0672; 0.2303]	12	0	0	1	1	1
(2, 4)	[0; 0.211]	9	0	0	1	1	1
(3, 1)	[0.1108; 0.269]	9	0	0	1	1	1
(3, 2)	[0.1534; 0.2448]	17	0	0	1	1	1
(3, 3)	[0.1216; 0.2472]	18	0	1	1	0.95	0.97
(3, 4)	[0.1002; 0.1912]	15	0	0	1	1	1
(4, 1)	[0; 0.2425]	7	0	0	1	1	1
(4, 2)	[0.0781; 0.1538]	13	0	0	1	1	1
(4, 3)	[0.0713; 0.3278]	12	0	0	1	1	1
(4, 4)	[0; 0.2309]	6	0	0	1	1	1
Groups	Threshold value range S_ε	Average characteristics					
Group I	[0.1534; 0.1538]	11	0	0.21	1	0.98	0.99
Group II	[0.2753; 0.4202]	6.50	0	0	1	1	1

After determining the ranges of the best values for the threshold S_ε, a common threshold value for most match classes was searched using the following algorithm:

1. The maximum set of classes of matches (Group I) is defined, for which the ranges of threshold values have an intersection with *C*. The classes of matches (Group I) that fall into the maximum set are highlighted in Table 1 in blue. Classes of matches that do not fall into the maximum set are highlighted in Table 1 in beige color.

2. From the found intersection of the ranges *C*, the common threshold value S_ε is chosen as the closest to the ranges of those classes of matches that did not fall into Group I. In this example, this value is chosen near the right border of intersection C since this border is open, and the ranges of other classes' matches (Group II) are to the right of Group I. The chosen value is S_ε=0.1538.

The performance indicators of the threshold rule at this threshold value are given in Table 2. Cells in the columns of metrics *P*, *R*, and *F*1 have the same color as in Table 1. Cells with values from the range [0.6; 0.75) – orange, and, as in Table 1, yellow – with values from the range [0.75; 0.9), and green – with values from the range [0.9; 1]. For convenience, the first column Class has the same color scheme that this column has in Table 1. Match classes that belong to the maximum set of match classes (Group I from Table 1) are marked in blue, other match classes are marked in beige.

For classes (1, 4) and (2, 1), the values of the precision metric *P* are in the range of (0.6, 0.7), and the values of metric F_1 are in the range of (0.75, 0.8). Metric values for classes of matches belonging to Group I are equal to the values of the metrics for the same classes in Table 1. This is because the threshold S_ε=0.1538 according to Table 1 belongs to the interval [min (S_ε); max (S_ε)] for these match classes.

Table 2

Estimates of the effectiveness of the threshold rule (16) at S_ε=0.1538 for each class of matches of the 2013–2014 season of the French II League

Class	Number of cases			Metrics		
	TP	FP	FN	*P*	*R*	*F*1
(1, 1)	5	0	0	1	1	1
(1, 2)	8	0	1	1	0.89	0.94
(1, 3)	14	0	1	1	0.93	0.97
(1, 4)	6	3	0	0.67	1	0.80
(2, 1)	7	4	0	0.64	1	0.78
(2, 2)	9	0	0	1	1	1
(2, 3)	12	0	0	1	1	1
(2, 4)	9	0	0	1	1	1
(3, 1)	9	0	0	1	1	1
(3, 2)	17	0	0	1	1	1
(3, 3)	18	0	1	1	0.95	0.97
(3, 4)	15	0	0	1	1	1
(4, 1)	7	0	0	1	1	1
(4, 2)	13	0	0	1	1	1
(4, 3)	12	0	0	1	1	1
(4, 4)	6	0	0	1	1	1
Average characteristics				0,96	0,99	0.97

The average values of the metrics, precision *P*, recall *R*, and *F*1, for the threshold rule (16) with S_ε=0.1538 on the data of the 2013–2014 season of the French II League are in the range of [0.9; 1].

5. 3. Results of the application of the devised improved method for detecting matches suspicious of a fixed result for the season of 2014–2015 in Italian Serie B

Before searching for MSFR, teams are grouped into clusters according to their success in the season, and based on this grouping, each team is assigned a rank $i \in \{1, ..., K\}$, where *K* is the number of clusters. Rank 1 is awarded to the best teams of the season, and rank *K* is awarded to a certain number of the last teams of the season, depending on the rules of the tournament. Examples of a grouping of tournament teams according to the final result are shown in works [9, 12]. The grouping principle is as follows: the first cluster is formed from *M*+1 first teams of the season, and the last cluster is formed from *L*+1 last teams of the season. The numbers *M* and *L* are selected from the increase or decrease rules. For Italian Serie *B*, *M*=3 and *L*=4. Other teams are divided into 2 groups: the first group consists of teams close to the teams of the first cluster, and the second group consists of teams close to the teams of the last cluster. To obtain these two groups, the 2-means algorithm was used in the current study. The results of team grouping are shown in Table 3 on the example of the teams of the 2014–2015 season of Serie *B* of Italy, *K*=4. Certain groups of teams in Table 3 are highlighted in separate colors: the teams of cluster No. 1 are highlighted in light red, the teams of cluster No. 2 – in green, the teams of cluster No. 3 – in yellow, and the teams of cluster No. 4 – in white.

Table 3

Grouping of teams of the 2014–2015 season of the Serie *B* in Italy

Team number in the overall standings	Team name	Number of points scored	Number of wins	Number of draws	Cluster No., Rank
1	’Carpi’	80	22	14	1
2	’Bologna’	73	18	19	1
3	’Frosinone’	71	20	11	1
4	’Pescara’	70	18	16	1
5	’Vicenza’	69	18	15	2
6	’Spezia’	67	18	13	2
7	’Perugia’	66	16	18	2
8	’Avellino’	65	17	14	2
9	’Livorno’	59	15	14	2
10	’Bari’	54	14	12	3
11	’Trapani’	53	13	14	3
12	’Ternana’	51	13	12	3
13	’Lanciano’	50	10	20	3
14	’Latina’	50	11	17	3
15	’Catania’	49	12	13	3
16	’Entella’	49	10	19	3
17	’Modena’	49	10	19	3
18	’Pro Vercelli’	49	12	13	4
19	’Brescia’	48	12	12	4
20	’Crotone’	48	12	12	4
21	’Cittadella’	44	9	17	4
22	’Varese’	39	9	12	4

Using the resulting team grouping, matches are classified by the rank of the participating teams. As a result, a set of match classes $C=\{(i,j)|\ i,j\in\{1,\ldots,K\}\}$ is formed.

For each class of matches (i, j), the expected result of the match is calculated as the average arithmetic difference of goals scored by the home and away teams of all matches of class (2).

Since there is no information on all fixed matches of the season, in each class of matches (i, j) MSFRs were determined according to the expert principle. If the match result did not match the expected $avg_{k(i,j)}$ or the goal difference in the match was one or more goals more than expected, then such a match is considered potentially suspicious. For example, if the expected result is 1.5, then matches in which the home and away goal difference is less than 1 goal or more than 2 goals are considered potentially suspicious. On the other hand, if a draw is expected, then matches in which the goal difference is not zero are considered potentially suspect. Table 4 shows the result of the marking of matches on the example of class (3, 3) matches of the 2014–2015 Italian Serie *B* season. Potentially suspicious matches are marked in gray. An additional yellow color indicates matches that are recognized as match-fixing according to information from external sources [14]. Table 4 is one of the results of the classification of matches. Data on the date of the match, the participating teams, and the result of the match were taken from an open resource [13] related to the study of bets on football matches.

Table 4

Marking of class (3,3) matches of the 2014–2015 season of Serie *B* in Italy

Match class (3,3) of the 2014–2015 season of the Italian II Division					
No.	Host team	Guest team	Result	Date of the event	Potentially suspicious
1	’Catania’	’Lanciano’	’3:3’	30-Aug-2014	0
2	’Entella’	’Bari’	’0:2’	30-Aug-2014	1
3	’Lanciano’	’Modena’	’2:0’	07-Sep-2014	1
4	’Catania’	’Modena’	’0:0’	20-Sep-2014	0
5	’Trapani’	’Entella’	’2:2’	23-Sep-2014	0
6	’Latina’	’Ternana’	’1:1’	27-Sep-2014	0
7	’Lanciano’	’Bari’	’1:1’	29-Sep-2014	0
8	’Bari’	’Modena’	’1:1’	04-Oct-2014	0
9	’Trapani’	’Latina’	’1:0’	04-Oct-2014	0
10	’Catania’	’Bari’	’2:3’	12-Oct-2014	1
11	’Ternana’	’Trapani’	’1:2’	24-Oct-2014	1
12	’Catania’	’Entella’	’5:1’	28-Oct-2014	1
13	’Entella’	’Lanciano’	’0:0’	01-Nov-2014	0
14	’Entella’	’Ternana’	’2:1’	04-Nov-2014	0
15	’Bari’	’Ternana’	’0:1’	10-Nov-2014	1
16	’Trapani’	’Catania’	’2:2’	16-Nov-2014	0
17	’Latina’	’Lanciano’	’1:0’	16-Nov-2014	0
18	’Bari’	’Trapani’	’2:1’	22-Nov-2014	0
19	’Catania’	’Latina’	’1:0’	23-Nov-2014	0
20	’Ternana’	’Catania’	’1:0’	29-Nov-2014	0
21	’Lanciano’	’Trapani’	’2:2’	06-Dec-2014	0
22	’Ternana’	’Lanciano’	’0:1’	13-Dec-2014	1
23	’Entella’	’Modena’	’1:1’	16-Dec-2014	0
24	’Bari’	’Latina’	’1:0’	20-Dec-2014	0
25	’Modena’	’Trapani’	’2:1’	20-Dec-2014	0
26	’Latina’	’Entella’	’1:1’	24-Dec-2014	0
27	’Ternana’	’Modena’	’1:0’	24-Dec-2014	0
28	’Modena’	’Latina’	’0:0’	28-Dec-2014	0
29	’Bari’	’Entella’	’0:0’	17-Jan-2015	0
30	’Lanciano’	’Catania’	’3:0’	17-Jan-2015	1
31	’Modena’	’Lanciano’	’1:1’	25-Jan-2015	0
32	’Entella’	’Trapani’	’1:1’	14-Feb-2015	0
33	’Ternana’	’Latina’	’0:2’	21-Feb-2015	1
34	’Bari’	’Lanciano’	’2:0’	21-Feb-2015	1
35	’Latina’	’Trapani’	’1:0’	27-Feb-2015	0
36	’Modena’	’Bari’	’0:1’	28-Feb-2015	1
37	’Bari’	’Catania’	’1:1’	03-Mar-2015	0
38	’Ternana’	’Entella’	’0:1’	03-Mar-2015	1
39	’Modena’	’Catania’	’0:0’	10-Mar-2015	0
40	’Trapani’	’Ternana’	’4:2’	14-Mar-2015	1
41	’Entella’	’Catania’	’2:0’	21-Mar-2015	1
42	’Lanciano’	’Entella’	’1:0’	29-Mar-2015	0
43	’Ternana’	’Bari’	’2:0’	02-Apr-2015	1
44	’Catania’	’Trapani’	’4:1’	11-Apr-2015	1
45	’Lanciano’	’Latina’	’1:1’	11-Apr-2015	0
46	’Modena’	’Entella’	’2:0’	11-Apr-2015	1
47	’Trapani’	’Bari’	’1:1’	18-Apr-2015	0
48	’Latina’	’Catania’	’1:2’	19-Apr-2015	1
49	Catania’	’Ternana’	’2:0’	24-Apr-2015	1
50	’Trapani’	’Lanciano’	’1:0’	28-Apr-2015	0
51	’Lanciano’	’Ternana’	’1:1’	02-May-2015	0
52	’Latina’	’Bari’	’2:0’	09-May-2015	1
53	’Trapani’	’Modena’	’3:0’	09-May-2015	1
54	’Entella’	’Latina’	’2:0’	16-May-2015	1
55	’Modena’	’Ternana’	’1:2’	16-May-2015	1
56	’Latina’	’Modena’	’1:1’	22-May-2015	0
57	’Entella’	’Modena’	’2:2’	30-May-2015	0
58	’Modena’	’Entella’	’1:1’	06-Jun-2015	0

According to information from Italian law enforcement agencies [14], in the 2014–2015 season of the Serie *B* in Italy, the owner of the Catania team admitted to organizing the fixing of the result in 5 matches in 2015. We are talking about 4 matches of this team against the teams of Varese, Trapani, Latina, Ternana, which took place in April, and 1 match of this team against the team of Livorno, which took place in May. According to the applied team grouping, 3 of the 4 mentioned matches (Catania – Trapani, Latina – Catania, Catania – Ternana) are in the match class (3, 3). Therefore, lines No. 44, No. 48, and No. 49 from Table 4 that correspond to these matches are marked with a yellow background color. In addition, the mentioned matches were identified as potentially suspicious according to the expert marking principle. Therefore, these lines from Table 4 are also marked with a gray background color.

Table 5 shows the result of the marking of matches on the example of class (4, 3) matches of the 2014–2015 Italian Serie *B* season. Table 5 is one of the results of the classification of matches. Data on the date of the match, the participating teams, and the result of the match are also taken from an open resource [13] related to the research of bets on football matches.

According to the applied grouping of teams, 1 of the 4 matches of April 2015 (Varese – Catania) for which there is a confirmation of the fixed result was included in the match class (4, 3). Therefore, line No. 35 from Table 5, which corresponds to this match, is marked with a yellow background color. In addition, the specified match was identified as potentially suspicious according to the expert marking principle. Therefore, this line from Table 5 is also marked with a gray background color.

The proposed threshold rule (16) was applied with the chosen threshold value of S_ε=0.1538 to the data of the 2014–2015 Serie *B* Italy season as test data. Because this, the season match data was processed in the manner discussed above to obtain the match classes for a given season. This particular season of the Serie *B* tournament in Italy was chosen since external sources [14] have information about matches that are officially recognized as match-fixing in this season. The results of applying the threshold rule (16) to the specified data are shown in Table 6. Cells in the columns of metrics *P*, *R*, and *F*1 have the same color design as in Table 3. Cells with values from the range [0.3; 0.6) – red, in orange – with values from the range [0.6; 0.75), yellow – with values from the range [0.75; 0.9), and green – with values from the range [0.9; 1]. For convenience, the first column Class has the same color scheme that this column has in Table 2. Match classes that belong to the maximum set of match classes are marked in blue, other match classes are marked in beige.

The results obtained by the threshold rule (16) based on the average values of the *P* and F_1 metrics are better than the results of applying the algorithms proposed in the previous study [12] to the same data.

The results of detecting matches suspicious of a fixed result in the class of matches (3, 3) of the season 2014–2015 Serie *B* Italy are shown in Fig. 2. Fig. 2, *a* shows the chart of the measure of non-conformity (1) calculated for each match from Table 4. Fig. 2, *b* shows the chart of the degree of non-conformity (*p*-value) (10) and the result of applying the *p*-value rule (6) at Δ=0.2 to the matches listed in Table 4. Fig. 2, *c* shows the chart of the power martingale (4) and the result of applying the martingale rule (5) to the matches listed in Table 4. Fig. 2, *d* shows the chart of the normalized modified function of the Stepanets indicator (15) and the result of applying the threshold rule (16) to the matches listed in Table 4.

Table 5

Marking of the class (4, 3) matches of the season 2014–2015 Serie *B* Italy

Match class (4.3) of the 2014–2015 season of the Italian II division					
No.	Host team	Guest team	Result	Date of the event	Potentially suspicious
1	'Crotone'	'Ternana'	'0:2'	30-Aug-2014	1
2	'Pro Vercelli'	'Catania'	'3:2'	07-Sep-2014	0
3	'Varese'	'Lanciano'	'1:1'	13-Sep-2014	0
4	'Brescia'	'Ternana'	'0:0'	20-Sep-2014	0
5	'Brescia'	'Lanciano'	'1:1'	23-Sep-2014	0
6	'Crotone'	'Catania'	'1:1'	23-Sep-2014	0
7	'Varese'	'Trapani'	'5:2'	27-Sep-2014	1
8	'Cittadella'	'Lanciano'	'2:3'	04-Oct-2014	1
9	'Cittadella'	'Entella'	'0:1'	20-Oct-2014	1
10	'Varese'	'Bari'	'2:1'	25-Oct-2014	0
11	'Pro Vercelli'	'Bari'	'3:0'	01-Nov-2014	1
12	'Varese'	'Modena'	'2:1'	02-Nov-2014	0
13	'Cittadella'	'Latina'	'1:1'	08-Nov-2014	0
14	'Crotone'	'Bari'	'3:0'	16-Nov-2014	1
15	'Pro Vercelli'	'Entella'	'2:0'	22-Nov-2014	1
16	'Crotone'	'Modena'	'1:4'	29-Nov-2014	1
17	'Varese'	'Entella'	'2:2'	06-Dec-2014	0
18	'Cittadella'	'Bari'	'0:1'	13-Dec-2014	1
19	'Pro Vercelli'	'Ternana'	'2:1'	20-Dec-2014	0
20	'Brescia'	'Bari'	'2:1'	24-Dec-2014	0
21	'Cittadella'	'Catania'	'3:2'	24-Dec-2014	0
22	'Pro Vercelli'	'Trapani'	'1:0'	28-Dec-2014	0
23	'Varese'	'Ternana'	'2:0'	28-Dec-2014	1
24	'Cittadella'	'Modena'	'1:1'	17-Jan-2015	0
25	'Crotone'	'Latina'	'2:1'	24-Jan-2015	0
26	'Brescia'	'Entella'	'3:0'	31-Jan-2015	1
27	'Cittadella'	'Trapani'	'1:0'	31-Jan-2015	0
28	'Pro Vercelli'	'Modena'	'1:1'	31-Jan-2015	0
29	'Brescia'	'Modena'	'0:1'	03-Mar-2015	1
30	'Crotone'	'Trapani'	'1:0'	07-Mar-2015	0
31	'Brescia'	'Latina'	'1:2'	14-Mar-2015	1
32	'Pro Vercelli'	'Lanciano'	'2:1'	21-Mar-2015	0
33	'Cittadella'	'Ternana'	'0:0'	29-Mar-2015	0
34	'Brescia'	'Trapani'	'1:1'	29-Mar-2015	0
35	'Varese'	'Catania'	'0:3'	02-Apr-2015	1
36	'Crotone'	'Lanciano'	'1:1'	18-Apr-2015	0
37	'Pro Vercelli'	'Latina'	'1:1'	25-Apr-2015	0
38	'Varese'	'Latina'	'1:2'	02-May-2015	1
39	'Brescia'	'Catania'	'4:2'	09-May-2015	1
40	'Crotone'	'Entella'	'0:0'	22-May-2015	0

Table 6

Estimates of the effectiveness of the threshold rule (16) at S_ε=0.1538 and the p-value rule at Δ=0 and Δ=0.2 for each class of matches of the 2014–2015 Italian Serie B

Rule p-value, Δ=0						
Class	Number of cases			Metrics		
	TP	FP	FN	P	R	$F1$
(1, 1)	4	0	1	1	0.80	0.89
(1, 2)	7	2	2	0.70	0.78	0.78
(1, 3)	8	3	0	0.73	1	0.84
(1, 4)	7	2	3	0.78	0.70	0.74
(2, 1)	6	4	2	0.60	0.75	0.67
(2, 2)	8	2	1	0.80	0.89	0.84
(2, 3)	7	7	5	0.50	0.58	0.54
(2, 4)	6	2	5	0.75	0.55	0.63
(3, 1)	12	1	1	0.92	0.92	0.92
(3, 2)	14	3	5	0.82	0.74	0.78
(3, 3)	17	4	6	0.81	0.74	0.77
(3, 4)	11	5	1	0.69	0.92	0.79
(4, 1)	8	2	2	0.80	0.80	0.80
(4, 2)	8	2	4	0.80	0.67	0.73
(4, 3)	12	1	4	0.92	0.75	0.83
(4, 4)	2	4	4	0.33	0.33	0.33
Average characteristics				0,75	0,74	0.75
Rule p-value, Δ=0.2						
Class	Number of cases			Metrics		
	TP	FP	FN	P	R	$F1$
(1, 1)	4	0	1	1	0.80	0.89
(1, 2)	8	3	1	0.73	0.89	0.80
(1, 3)	8	7	0	0.53	1	0.70
(1, 4)	8	2	2	0.80	0.80	0.80
(2, 1)	8	4	0	0.67	1	0.80
(2, 2)	9	2	0	0.82	1	0.90
(2, 3)	10	10	2	0.50	0.83	0.63
(2, 4)	6	3	5	0.67	0.55	0.60
(3, 1)	12	4	1	0.75	0.92	0.83
(3, 2)	16	4	3	0.80	0.84	0.82
(3, 3)	22	5	1	0.80	0.96	0.88
(3, 4)	12	8	0	0.60	1	0.75
(4, 1)	10	3	0	0.77	1	0.87
(4, 2)	11	4	1	0.73	0.92	0.82
(4, 3)	15	1	1	0.94	0.94	0.94
(4, 4)	4	5	2	0.44	0.67	0.53
Average characteristics				0,72	0,88	0.78
Threshold rule, S_ε=0.1538						
Class	Number of cases			Metrics		
	TP	FP	FN	P	R	$F1$
(1, 1)	4	3	1	0.57	0.80	0.67
(1, 2)	9	0	0	1	1	1
(1, 3)	8	0	0	1	1	1
(1, 4)	10	0	0	1	1	1
(2, 1)	8	7	0	0.53	1	0.70
(2, 2)	9	4	0	0.69	1	0.82
(2, 3)	8	0	4	1	0.67	0.80
(2, 4)	11	5	0	0.69	1	0.82
(3, 1)	13	0	0	1	1	1
(3, 2)	13	0	6	1	0.68	0.81
(3, 3)	23	0	0	1	1	1
(3, 4)	12	13	0	0.48	1	0.65
(4, 1)	10	0	0	1	1	1
(4, 2)	12	0	0	1	1	1
(4, 3)	16	0	0	1	1	1
(4, 4)	6	8	0	0.43	1	0.60
Average characteristics				0.84	0.95	0.87

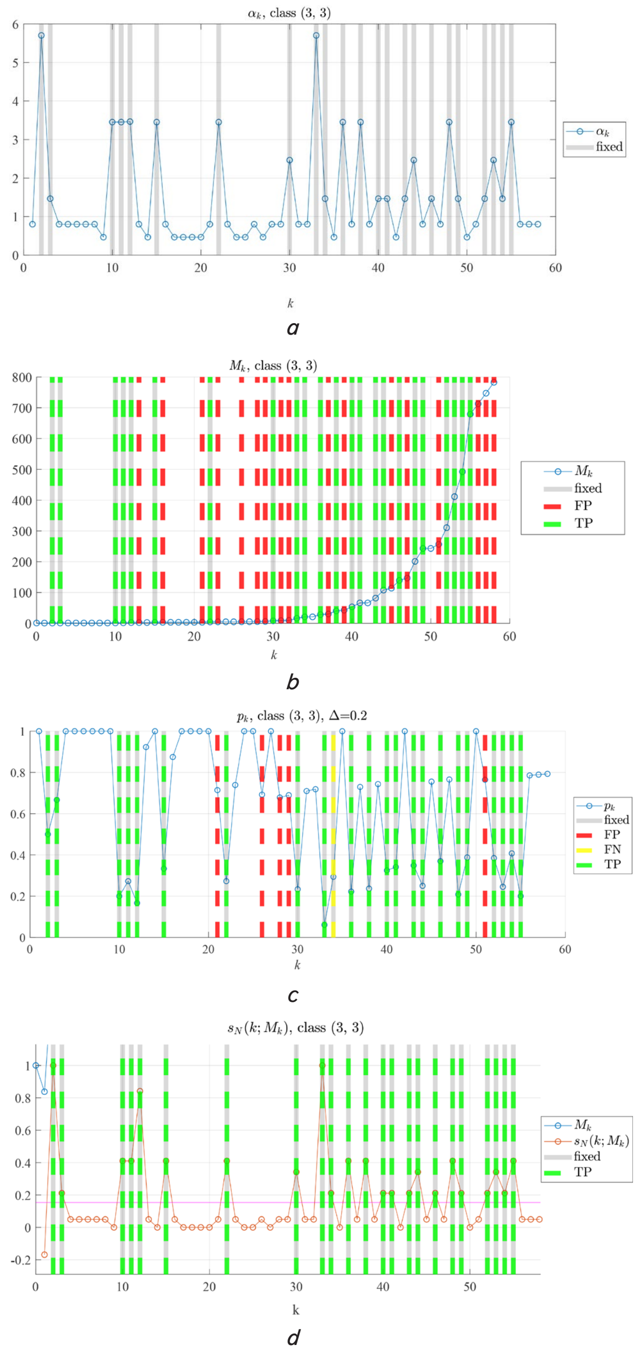

Fig. 2. The application of different rules for detecting MSFR in the class of matches (3, 3) of the 2014–2015 season of the Serie B Italy: a – measure of non-conformity of matches; b – martingale rule; c – p-value rule; d – threshold rule (16)

The results of detecting matches suspicious of a fixed result in the class of matches (4, 3) of the season 2014–2015 Serie B Italy are shown in Fig. 3. Fig. 3, a shows the chart of the measure of non-conformity (1) calculated for each match from Table 5. Fig. 3, b shows the chart of the degree of non-conformity (p-value) (10) and the result of applying the p-value rule (6) at Δ=0.2 to the matches listed in Table 5. Fig. 3, c shows the chart of the power martingale (4) and the result of applying the martingale rule (5) to the matches listed in Table 5. Fig. 3, d shows the chart of the normalized modified function of the Stepanets indicator (15) and the result of applying the threshold rule (16) to the matches listed in Table 5.

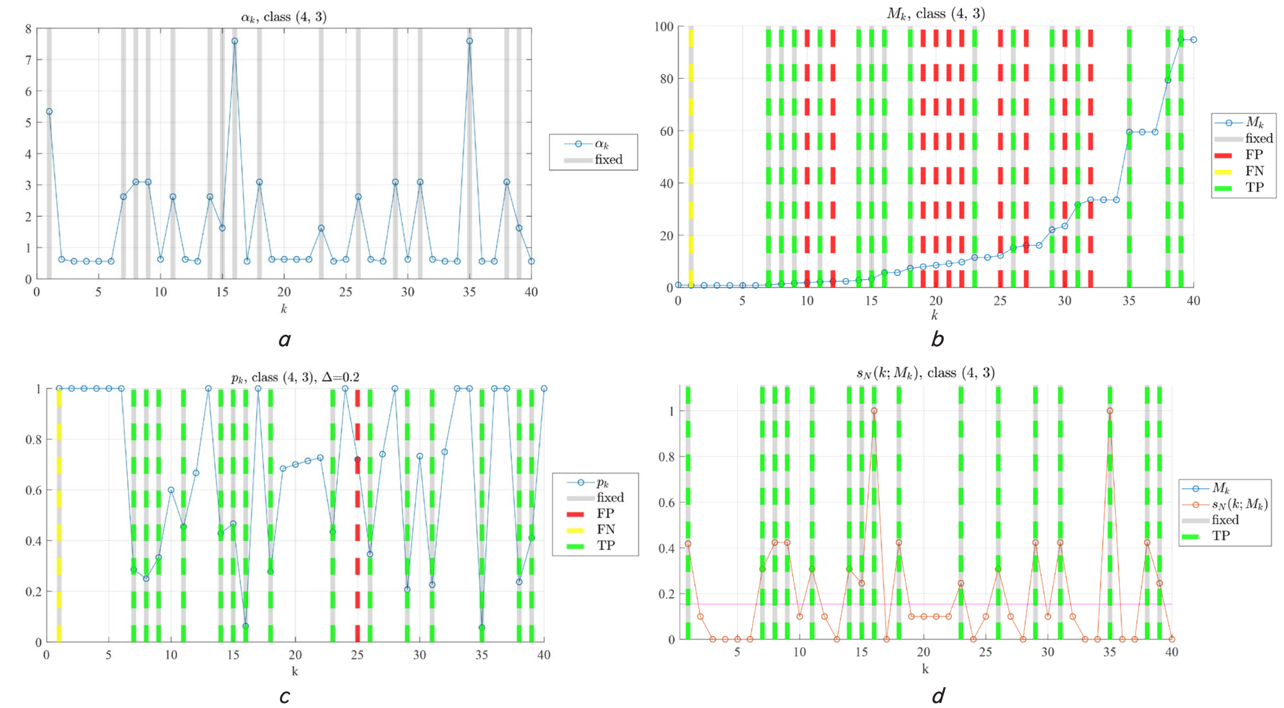

Fig. 3. The application of different rules for detecting MSFR in the class of matches (4, 3) of the 2014–2015 season of the Serie *B* Italy: *a* – measure of non-conformity of matches; *b* – martingale rule; *c* – *p*-value rule; *d* – threshold rule (16)

On each chart in Fig. 2, 3, match numbers are placed on the horizontal axis, which are in the ordinal (first) columns. Thus, the charts of the characteristics are arranged horizontally in the same way as the matches in the tables of the match classes of Tables 4, 5. Fig. 2, 3, *d*, in addition to the results of applying rule (16), the threshold S_ε at which the results were obtained is shown in the form of a pink straight line. In Fig. 2, 3, on each chart gray lines highlight the points that correspond to MSFR from the tables (Table 4 for Fig. 2 and Table 5 for Fig. 3, respectively). The shaded green lines on the charts (*b*–*d*) highlight those matches that were found to be suspicious according to the relevant rule and are such according to the tables. Accordingly, the red dashed lines on the charts (*b*–*d*) highlight those matches that are identified as suspicious according to the relevant rules but are not such according to the tables.

Regarding the detection of matches that were indicated in the news [14]:

– Varese VS Catania (class (4, 3) #35, 2 April 2015, 0:3) was detected by the proposed threshold rule (16). Graphic confirmations are shown in Fig. 2;

– Catania VS Trapani (class (3, 3) #44, April 11, 2015, 4:1), Latina VS Catania (class (3, 3) #48, April 19, 2015, 1:2), Catania VS Ternana (class (3, 3) No. 49, April 24, 2015, 2:0) are detected by the proposed threshold rule (16). Graphic confirmations are shown in Fig. 3.

6. Discussion of results of the application of the improved method for detecting matches suspected of a fixed result

The results of the application of the improved method of detecting MSFR based on the proposed threshold rule (16) and the previously devised method based on the *p*-value rule (6) and the martingale rule (5) are shown in Fig. 2. After applying the normalized modified Stepanets indicator function (15) $s_N(k,M(k))$ to the martingale (4) $M(k)$ (Fig. 2, *b*), we obtain a chart of the function (Fig. 2, *d*). This chart shows that the points marked with gray lines have a value of characteristic (15) $s_N(k,M(k))$ greater than the points corresponding to normal matches and not marked with gray lines. Therefore, it is possible to introduce a certain threshold value S_ε at which, applying the threshold rule (16), all or most MSFRs in the class (3, 3) will be detected. Similar results of applying the threshold rule are shown in Fig. 3, *c*, for the class of matches (4, 3).

Let us first consider the results of the application of the methods on the class of matches (3, 3) of the 2014–2015 season of the Serie *B* in Italy (Table 4, Fig. 2). According to the precision (*P*) and F_1 metric, the applied rules were ranked in the following order: the third place was taken by the martingale rule (5), the second place – the *p*-value rule (6), the first place – the proposed threshold rule (16). At the same time, the values of the metrics (*P*) and F_1 for this class of matches are close to the average values of these metrics in the case of the threshold rule (16) (Table 6). This can be explained by several factors: the range of metric values for the threshold rule, and the quality of the performance of the non-conformity measure on this class of matches.

The range of metric values for the threshold rule is still large: from 0.43 to 1 for the precision metric and from 0.6 to 1 for the F_1 metric, which leads to a decrease in the average values of these metrics. At the same time, in most classes of matches, the values of the metrics are equal to 1 (for 10 out of 16 according to the precision metric and 8 out of 16 according to the F_1 metric).

On the class of matches (3,3), potentially suspicious matches are qualitatively distinguished by the measure of non-conformity (Fig. 2, *a*): the values of the measure of non-conformity for potentially suspicious matches are greater than the values of the measure of non-conformity for normal matches. This leads to the fact that the application of the p-value rule (6) at $\Delta=0.2$ makes it possible to detect most of the potentially suspicious matches.

The fact that the martingale rule (5) has the lowest precision is due to the fact that the value of the martingale (5) increases with any deviation of the degree of non-conformity (5) from 1: $M_{k-1}<M_k$ when $p_{k-1}\geq 1$ and $p_{k-1}\geq 1$. On the one hand, the martingale rule (5) is an effective rule for the case that all normal matches will receive a p-value equal to 1. This, in turn, is possible by choosing an appropriate nonconformity measure (1), which may be one of the directions of further research.

On the other hand, the application of the normalized modified Stepanets indicator function (15) to the martingale (4) made it possible to clearly distinguish all or most of the matches that belong to the potentially suspicious class. According to Fig. 2, 3, *d*, it can be seen that all matches belonging to the potentially suspicious class are clearly distinguished from the others by the value of characteristic (15). Thus, the combination of the martingale characteristic (4) and the normalized modified Stepanets indicator function (15) made it possible to propose an improved rule (16). Using this threshold rule with an appropriate threshold value allows one to achieve the highest precision rates for detecting potentially suspicious matches. Therefore, in the framework of the current study, it was shown that the proposed threshold rule allows qualitative detection of potentially suspicious matches.

In turn, the proposed rule has a limitation related to the determination of the threshold value. In this study, a procedure for finding a threshold using training data is proposed. A sample of data for the 2013–2014 season of the II League of France was used as a training set. And as a test sample, data for the 2014–2015 season of the Serie *B* in Italy was used. The training and test data are similar, as both leagues are second divisions in the French and Italian football league systems, respectively. The experimental result of the study confirmed the hypothesis about the similarity of the data: the application of rule (16) with the threshold $S_\varepsilon=0.1538$ determined on the training data provided the highest average indicators of the quality of detection of potentially suspicious matches. Thus, if potentially suspicious matches need to be detected using rule (16) in a certain season, one can choose a similar season and use it to determine the threshold value for the rule (16).

On the other hand, the need to use data from another season can be considered as a disadvantage of this approach for determining the threshold since the procedure for establishing the similarity between two seasons is not formally defined. Therefore, one of the directions of further research may be the definition of an analytical method for finding a threshold that is optimal according to a specified criterion within the framework of the task of detecting MSFR.

In comparison with the results given in [12], the proposed method for detecting MSFR based on the threshold rule (16) has better indicators of the metrics of precision P, recall R, and $F1$ (Table 2) than the method of detecting MSFR based on the martingale rule (5) both on average and in most match classes. In the classes of matches (1, 4) and (2, 1) of the 2013–2014 season of the II French League, the proposed method provides the same indicators of the metrics of precision P, recall R and $F1$ (Table 2) as the method of detecting MSFR based on the martingale rule (5).

Also, the proposed method for detecting MFPR based on the threshold rule (16) has better indicators of the precision P, recall R and $F1$ metrics (Table 2) than the method for detecting MFFR based on the p-value rule (6) [12], on all classes of matches of season 2013–2014 II League of France, except (1, 4) and (2, 1).

In contrast to [4–6], the devised method uses input data of a much smaller size, which are always publicly available (date and score of the match, names of the teams participating in the match, for which team the match is considered home). Therefore, a direct comparison of its effectiveness with methods based on the analysis of bets on a match or the movement of players during a match is incorrect. However, the devised method can be supplemented with the models proposed in [4–6] to improve the performance of its work due to the use of additional data of a different nature. Conversely, the devised method can be considered as an effective filter for matches suspicious of a fixed result. To the matches filtered with its help, it will be possible, if the necessary data are available, both to apply the methods known from [4–6] and to use non-mathematical profile methods of the relevant law enforcement agencies.

7. Conclusions

1. The modified Stepanets indicator function characterizes the relative rate of change of the value of the function regardless of its argument and takes into account not only the magnitude of the relative change of the function but also the direction of its change. Additionally, the normalization of the modified function-indicator of Stepanets makes it possible to reduce its value in each class of matches to a common range. The modified Stepanets indicator function is applied to the power martingale and shows the relative change in the martingale value of the current match compared to the previous match, taking into account the above features.

2. The method for detecting atypical events in the theory of conformal predictors has been improved, which is distinguished by the use of a modified Stepanets indicator function. The martingale value for a match is formed based on the results of the conformal predictor for all matches of the season. In the proposed method, matches suspicious for a fixed result are detected by comparing the values of the normalized modified Stepanets indicator function with the threshold. The threshold value is determined experimentally according to the criterion of the maximum of the F_1 metric on the training sample.

3. On the training sample, for which data from the 2013–2014 season of the French II league were used, the improved method provided better indicators of precision metrics and $F1$ (average values of metrics $P=0.96$, $F1=0.97$) than martingale rules and p-value (average values of metrics $P=0.87$, $F1=0.85$). On the test sample, for which data from the 2014–2015 season of the Italy Serie *B* were used, the indicators of precision metrics and $F1$ for the improved method are higher (average metric values $P=0.84$, $F1=0.87$) than the same indicators for martingale rules and p-value (average values of metrics $P=0.75$, $F1=0.78$) applied to the same data. On the other hand, on certain classes of matches, the proposed method achieves lower precision metrics than martingale rules and p-value. This can occur on match classes for which the best threshold value ranges do not intersect with the maximum set of match classes that share a common

intersection of threshold ranges that achieve the highest values of the F_1 metric. The proposed threshold rule turns out to be 4 out of 5 matches of the 2014–2015 Serie *B* season in Italy, which are considered fixed according to information from official Italian law enforcement agencies.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study and the results reported in this paper.

Funding

The study was carried out within the framework of research work No. 2310p «Information and analytical system for mathematical modeling and management of social risks with application in technology and medicine» (State registration number 0120U102216).

Data availability

The data will be provided upon reasonable request.

Acknowledgments

The authors would like to express their deep gratitude to the staff of the Applied Mathematics Department of the Faculty of Applied Mathematics of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» for their support in conducting research within the framework of research work No. 2310p «Information and analytical system for mathematical modeling and management of social risks with application in technology and medicine» (State registration number 0120U102216).

References

1. Abarbanel, B., Johnson, M. R. (2018). Esports consumer perspectives on match-fixing: implications for gambling awareness and game integrity. International Gambling Studies, 19 (2), 296–311. doi: https://doi.org/10.1080/14459795.2018.1558451
2. Lilley, E. (2015). A Review of the recommendations of the 'Report of the Sports Betting Integrity Panel' in assessing the progress towards tackling Match-fixing in Sport. Laws of the Game, 1 (1).
3. Huggins, M. (2018). Match-Fixing: A Historical Perspective. The International Journal of the History of Sport, 35 (2-3), 123–140. doi: https://doi.org/10.1080/09523367.2018.1476341
4. Forrest, D., McHale, I. G. (2019). Using statistics to detect match fixing in sport. IMA Journal of Management Mathematics, 30 (4), 431–449. doi: https://doi.org/10.1093/imaman/dpz008
5. Razali, N., Mustapha, A., Yatim, F. A., Ab Aziz, R. (2017). Predicting Football Matches Results using Bayesian Networks for English Premier League (EPL). IOP Conference Series: Materials Science and Engineering, 226, 012099. doi: https://doi.org/10.1088/1757-899x/226/1/012099
6. Anfilets, S., Bezobrazov, S., Golovko, V., Sachenko, A., Komar, M., Dolny, R., Kasyanik, V. et al. (2020). Deep multilayer neural network for predicting the winner of football matches. International Journal of Computing, 19 (1), 70–77. doi: https://doi.org/10.47839/ijc.19.1.1695
7. Spapens, T., Olfers, M. (2015). Match-fixing: The Current Discussion in Europe and the Case of The Netherlands. European Journal of Crime, Criminal Law and Criminal Justice, 23 (4), 333–358. doi: https://doi.org/10.1163/15718174-23032077
8. Stepanets, A. I. (2005). Methods of Approximation Theory. Boston. doi: https://doi.org/10.1515/9783110195286
9. Chertov, O., Zhuk, I., Serdyuk, A. (2021). Search of the Deviation from the Natural Process Using Stepanets Approach for Classification of Functions. 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). doi: https://doi.org/10.1109/idaacs53288.2021.9660997
10. Vovk, V. (2014). The Basic Conformal Prediction Framework. Conformal Prediction for Reliable Machine Learning, 3–19. doi: https://doi.org/10.1016/b978-0-12-398537-8.00001-8
11. Ho, S.-S., Wechsler, H. (2010). A Martingale Framework for Detecting Changes in Data Streams by Testing Exchangeability. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32 (12), 2113–2127. doi: https://doi.org/10.1109/tpami.2010.48
12. Zhuk, I., Chertov, O. (2023). Framework based on conformal predictors and power martingales for detection of fixed football matches . Eastern-European Journal of Enterprise Technologies, 2 (4 (122)), 6–15. doi: https://doi.org/10.15587/1729-4061.2023.276977
13. Historical analysis of closing odds. Available at: https://github.com/Lisandro79/BeatTheBookie
14. Catania's owner admits to match fixing in five Serie B games. The Guardian. Available at: https://www.theguardian.com/football/2015/jun/30/catania-match-fixing-serie-b