Detection of Possible Match-fixing in Tennis Games

Kim Yong-wook[1], Jinyoung Han[2] and Choi Seoung-rak[1]
[1]*Department of Sports Science, Hanyang University, Ansan, South Korea*
[2]*Department of HCI, Hanyang University, Ansan, South Korea*

Keywords: Match-fixing, Tennis, Benford's Law.

Abstract: This study seeks ways that can immediately detect match-fixing in a tennis match. We first explore the number of rallies observed in tennis matches of the ATP and WTA leagues to determine whether they follow the Benford's law. We also artificially manipulate practice games to investigate whether the number of ralleys observed in manipulated matches also follows the Benford's law. Experimental results demonstrate that the numbers collected from fixed games and the expected frequencies predicted by Benford's Law are different. Based on the lessons learned, we develop a machine-learning-based model for detecting whether a given match is fixed or not. Our model shows a high accuracy in detecting fixed tennis matches, which has a great utility for fair tennis play.

1 INTRODUCTION

Sports must always conform to the ethics of fair play. Particularly when contending for victory, there is an underlying principle that athletes must follow fair rules in a righteous way and strive to show their best performance. However, in a world of professional sports where there are conflicts of vast interests, dramas continue to unfold, sliding back and forth between legal and illegal paths. There is a clear limit to the fundamental roles that professional athletes are basically paid to play. Furthermore, athletes fall into temptation of illegal match fixing each time their desire for fame and wealth combines with the greed of sports clubs (European Commission, 2018; Lastraa et al., 2018; Asser Institute, 2014).

In a fair competition, under the fair play system, outcomes cannot be predicted with an absolute certainty. However, an act of artificial manipulation of the game, such as a player's slowdown, the bribery of referees, or match-fixing by referees, can certainly change the outcome of a game. Match-fixing is a process of pre-determining results that might provide a certain opportunity for avid gamblers to win bets in sports gambling (Moriconi, 2018). It can be a serious act that weakens the foundation of sports. Information Commissioner's Office (ICO) and The Federation Internationale de Football Association (FIFA) have acknowledged "prevention of match-fixing" as the most critical sports issue to be addressed in the 21st century and have taken various preventive measures (IOC, 2016; FIFA, 2017; Aquilina and Chetcuti, 2014).

Match-fixing occurs extensively in a wide variety of professional sports leagues, ICO-sponsored Olympics, FIFA-sponsored World Cups, as well as the world championship games organized by various sports federations. The European Union (EU) Commission has reported more than 5,200 match-fixing cases worldwide since the year 2000 (European Commission, 2014; European Commission, 2018).

Since 2000, match-fixing cases have been most commonly found in the European league football matches. However, the cases are expected to rise if the leagues that the EU does not monitor, e.g., Asian leagues and the American leagues, are also included (European Commission, 2012; Katsarova, 2016). In particular, leagues operating in the Asian region have been under relatively less scrutiny than the European leagues, making them a softer target for match fixers (Hill, 2010). A match-fixing scandal that surfaced in Korea's professional football league in 2011 caused two footballers to commit suicide, 10 to be permanently suspended, and 59 to receive criminal penalties (THE KOREA TIMES, 2012). This incident was perpetrated by an overseas criminal organization that managed to bribe footballers with as little as USD 5,000. It delivered a devastating wound to the Korean football league that ended the lives and professional careers of many of its players forever. Even to this

Yong-wook, K., Han, J. and Seoung-rak, C.
Detection of Possible Match-fixing in Tennis Games.
DOI: 10.5220/0006924201240131
In *Proceedings of the 6th International Congress on Sport Sciences Research and Technology Support (icSPORTS 2018)*, pages 124-131
ISBN: 978-989-758-325-4

date since the conclusion of the investigations, the true identity and nature of the overseas crime group that was involved in the incident have not been revealed. There is a very high possibility that the criminal group is attempting to fix another game at this moment.

Match-fixing is very common in tennis leagues. It has been reported that match-fixing was attempted at the 2016 Wimbledon Championship, one of the most prestigious tennis tournaments. Novak Djokovic, who ranked first in the world in 2016, revealed in an interview that he was once offered a USD 200,000 monetary reward. At the same time, he also received physical threats from a group that was trying to fix a match (Huffington post, 2016). This happens owing to a structural problem with regard to the financial reality of professional tennis, in which players heavily rely on the prize money and monetary support from their sponsors to make a living.

According to Michael Russell's interview with Forbes, the 92nd ranked player in the world in 2013, he earned USD 270,000 in revenue in 2012, but his actual income was USD 85,000 after a deduction of personal and tour expenses that amounted to USD 75,000 and 35,000 respectively, while taxes resulted in the deduction of another significant portion of his effective earnings. He further disclosed that although he made a total of USD 2,100,000 in a 15-year-long career, there is not much left after his tour expenses and fee for coaches and trainers (Forbes, 2013). This means that the lower ranked players with less income in the current world of professional tennis are more prone to the temptation of match-fixing.

As match-fixing has become a serious social issue, a series of related researches have been carried out on match-fixing, such as studies that examine the match-fixing cases and explain their cause (Hill, 2013; Carpenter, 2012) and those that analyze the relevance of match-fixing with the sports betting industry (Bag and Saha, 2011; Boeri and Severgnini, 2011). In addition, there are studies that offer legal countermeasures to eradicate match-fixing (Rodenberg and Feustel, 2014) as well as attempts to detect match-fixing via a mathematical approach (Hill, 2011).

To completely stamp out match-fixing from the world of sports, the introduction of preventive measures along with ethics education and training for the athletes is essential. It is also critical to detect one in time, on occurrence. However, there are numerous variables to be considered in a sport like tennis, unexpected outcomes often arise depending on the condition of the athletes playing. Therefore, when a player or coach, who is artificially defeated in a fixed match, and claims later that "the player was in a tattered physical condition and hence unable to perform his best", on-site investigation no longer remains a feasible option.

Currently, the best option is to conduct a post-incident investigation by anti-corruption organizations such as the Tennis Integrity Unit over match-fixing suspicions in tennis tournaments. However, such inquiries require conclusive evidence to determine culpability, such as evidence of monetary transactions surrounding suspected athletes; moreover, only the authorities with jurisdiction to conduct criminal investigation, such as the prosecution or the police, have access to them. As a result, organizations such as the Tennis Integrity Unit, which lack the authority, begin with probing into suspicious events. Once sufficient evidence is gathered to support the allegations, they request international investigative authorities to open official investigations into the alleged match-fixing incidents (TIU, 2018).

This approach investigates players belonging to different nationalities under the jurisdiction of the country where the match-fixing is suspected to have occurred, and it inevitably accompanies highly complicated and cumbersome administrative procedures. Therefore, if clear evidence such as testimonies of fraudulent deals or any circumstantial evidence regarding illegal monetary transactions is not secured, ongoing investigations often get suspended. As discussed earlier, a post-match investigation into a match-fixing incident requires tremendous effort to gather sufficient evidence to support a finding of actual unlawful transactions, making it highly difficult to identify match-fixing in reality (HM Government, 2017). Consequently, the current method of revealing fixed matches and imposing severe penalties on the conspirators is not enough to eradicate the problem of match-fixing.

Then how do we prevent match-fixing? The best approach for now would be to hinder the match-fixing attempts in advance. The following two types of preventive measures are currently being implemented at the sports scenes. First, we can encourage the athletes to be aware of the seriousness of match-fixing by continuously providing them with ethics education and training (Department for Culture, Media, and Sport: UK, 2010). Second, if pre-game signs of possible match-fixing appear, such as a sudden change in the dividend rates offered by sports betting sites, it would be prudent to prevent attempts of match-fixing by informing the players that the impending match is most likely, a fixed match. These methods use trainings and warnings to discourage the players from attempting to fix a match and they are already being used by IOC, professional leagues, and various sports federations (IOC, 2018; FIFA, 2018),

However, if athletes continue with their attempts to fix matches despite the measures in force, there are almost no additional preventive measures available that can be applied at the scene. Even if there is a method to immediately detect match-fixing at a game, how would that prevent such incidents from occurring? In that scenario, the athletes will be alerted that a match-fixing attempt will most likely be exposed and they will inevitably reduce such attempts. Therefore, this study introduces a method to detect match-fixing immediately in the field of tennis.

2 RESEARCH METHOD

The research method applied to this study for the detection of match-fixing is briefly explained as follows. First, we observe the number of rallies overserved in matches of the ATP and WTA leagues to determine whether they follow the Benford's law. Second, we artificially manipulate practice games to collect the data on the number of rallies in each match and verify whether the distribution follows the Benford's law. Lastly, we develop a machine-learning-based model to detect whether a given match is fixed or not. For leaning and testing the model, A-Set (training set), which is the number of rallies recorded from ATP and WTA, and B1-Set (validation set) and B2-Set (test set) that reflect the data collected from two artificially manipulated matches are used.

2.1 Benford's Law

Benford's Law is an observation that numerical values that can be observed in our daily lives appear in accordance with certain rules. When we look at datasets such as the population numbers, death rates, passwords, and lengths of rivers, the probability of the first digit being number 1 is approximately 31%, while the numbers 5 and 9 appear as the most significant digits 8%, and 5% of the time, respectively. This shows that the lower numbers are observed more frequently than the higher numbers (Figure 1).

Benford's Law was first discovered in 1881 by an American mathematician, Simon Newcomb. Newcomb noticed that the earlier pages in logarithm tables were much more worn than the other pages, and realized that the smaller digits were more likely to appear than the larger digits as naturally occurring real-life numbers. It was an empirically interesting discovery, but it lacked mathematical and logical

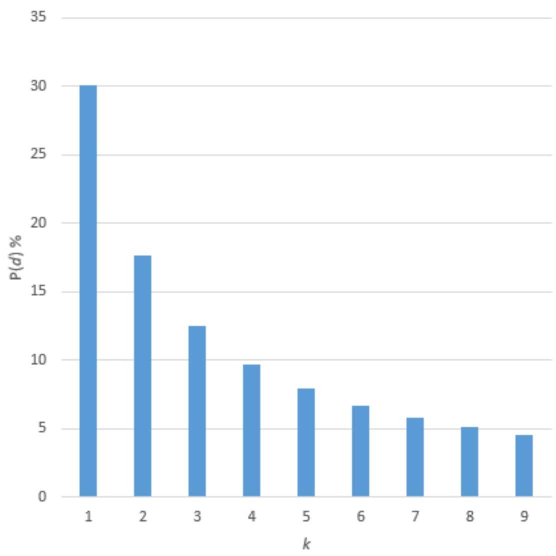

Figure 1: The distribution of first digits according to Benford's law.

explanation, and was not accepted as a law. It was again noted in 1938 by an American physicist Frank Benford, who observed a phenomenon that supported Newcomb's claim in naturally occurring collections of data. He analyzed 20 unrelated domains, such as river lengths, population numbers, and also a number of magazines and as a result, the probability of the first digit being number 1 appeared as 31%, while 19% began with number 2. He measured the probabilities for the occurrence of the digits ranging from 1 to 9 and announced the results. This observational phenomenon was later named after him as the Benford law (Nigrini, 2012).

The Benford's Law reveals the probability distribution of naturally occurring numbers, and as artificial numbers do not follow this law, we can identify human intervention by verifying the numbers observed. For example, numbers that reflect people's thoughts and purposes, such as the phone numbers, postal codes, and the price of goods, do not follow the Benford's law. Therefore, we believe the law can be applied in this study. We conjecture that the numbers observed in matches played with a specific intention, such as match-fixing, might differ from the data recorded at the games that were played to win.

2.2 Machine Learning

Machine learning algorithms are classified into supervised learning, unsupervised learning, and reinforcement learning. Supervised learning utilizes input and labeled data and accurate results of input data are achieved using the training data. A program is generated based on the analysis of these patterns, enabling it to predict the results of the incoming input data. Figure 2 shows the process of analyzing the training data and creating the evaluation model to verify the test data. Here, the extract features are the data format expressed and processed for the algorithm to judge. Machine learning algorithm generates a model that predicts an output based on an input. In the end, the output created based on the input can be predicted as a result of the learning carried out by the machine learning algorithm. In this study, an analysis of the deep learning method of supervised learning is used.

2.3 Research Data

This study used a number of rallies observed in tennis matches for analysis. A rally can be seen in a tennis match where a ball is played with a net, such as a tennis ball, and it refers to a process where the ball passes over to the opposing side of the court and then returns. In other words, the number of rallies indicates the number of times the ball has passed over to the other side of the court and one round trip is referred to as one rally. However, in this study, we used the number of times that the opponent passes the ball for a more accurate analysis and named it as h-rally. The h-rally data is collected by implementing the following two methods.

2.3.1 Data Collection from ATP and WTA Leagues

The matches studied for the data collection were the tennis matches held in the ATP and WTA leagues from January 2016 to March 2018. We selected and collected data from 220 ATP and WTA matches that were broadcasted television or were available online. The data was recorded by researchers who watched every selected game, and the h-rally that occurs for each score was directly observed and collected. The dataset collected here is referred to as “Dataset A”.

2.3.2 Data Collection from Fixing Game

We artificially manipulated practice games by asking one of the players with similar skills to lose the match, and we measured the outcomes. The experiments were conducted twice on H and S tennis courts in Ansan, South Korea on April 21 and May 13, 2018. Four players, C, K, J, and P participated in twelve games in total, six times on each day. We made a prior arrangement with player C to artificially lose the game while players K, J, and P remained unaware of player C’s intention to fix the game throughout the entire experiment. In the end, h-rallies recorded from all six matches that player C participated were collected as a dataset for fixed games. The datasets collected on the first and second days of experiments were labeled as “Dataset B1” and “Dataset B2”, respectively.

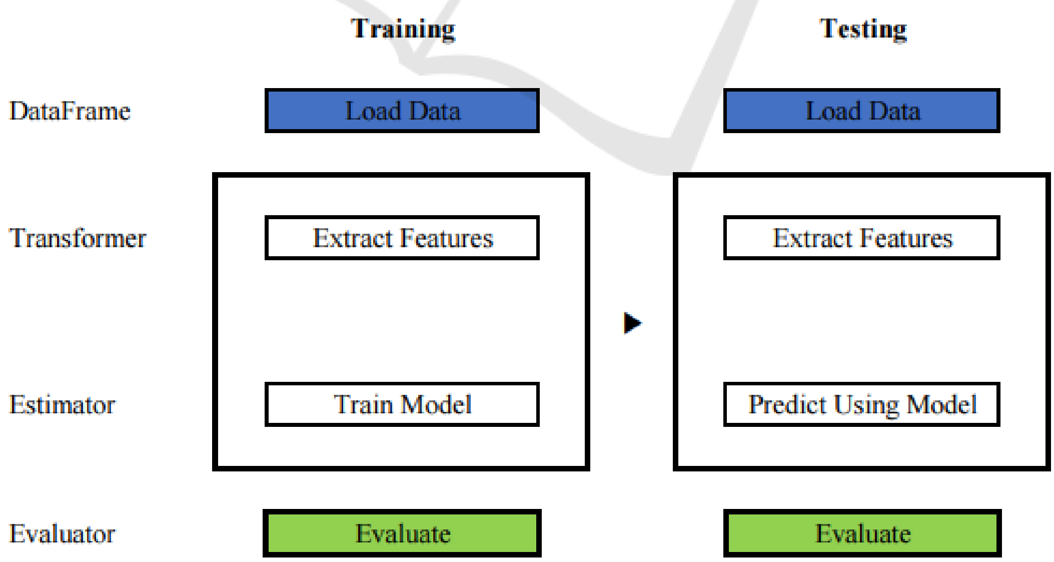

Figure 2: Supervised Learning workflow.

2.4 Analysis Method

The collected data was analyzed as follows. First, we verified whether the data by applying x^2 test statistics followed by the Benford's law. Second, we used "Dataset A" collected from ATP and WTA as a training set and "Dataset B1" that was collected from artificially manipulated matches as a test set. Third, we used additional dataset, "Dataset B2", for validating our model.

3 RESULTS AND DISCUSSION

Previous studies that adopted the Benford's Law analysis simply concluded that it can be used in detecting the occurrence of match-fixing. Therefore, this study investigated whether the distribution of the number of ralleys follows the Benford's Law. We also explored whether the expected frequencies of ralleys measured through the Benford's Law can be used in detecting match-fixing attempts.

3.1 Benford's Law Analysis

Figure 3 graphically reflects the comparison of Data Set B, collected through the artificial match-fixing, and Data Set A, gathered from ATP and WTA, against Benford's Law. Because there are not enough confirmed cases of match-fixing, obtaining data from confirmed fixed matches was not simple. The statistical analysis of these values produces slightly different results. Table 1 shows that normal games follow Benford's Law. The data for that can be found in "Dataset A". The data from fixed matches found in "Dataset B" shows that it doesn't follow Benford's Law. Data from sets A and B were not compared with each other but with the 'Probability for Benford's Law'. The results of the data in Table 1 was verified by the use of chi-square value. These results differ from those of a previous study that applied Benford's Law to fixed badminton matches (Choi and Park, 2017). In other words, it signifies that unlike in badminton games, the relationship between the number of rallies and the Benford's Law is not statistically significant in tennis matches. This suggests that it is difficult to judge whether a game is fixed based only on the number of rallies made in each tennis match.

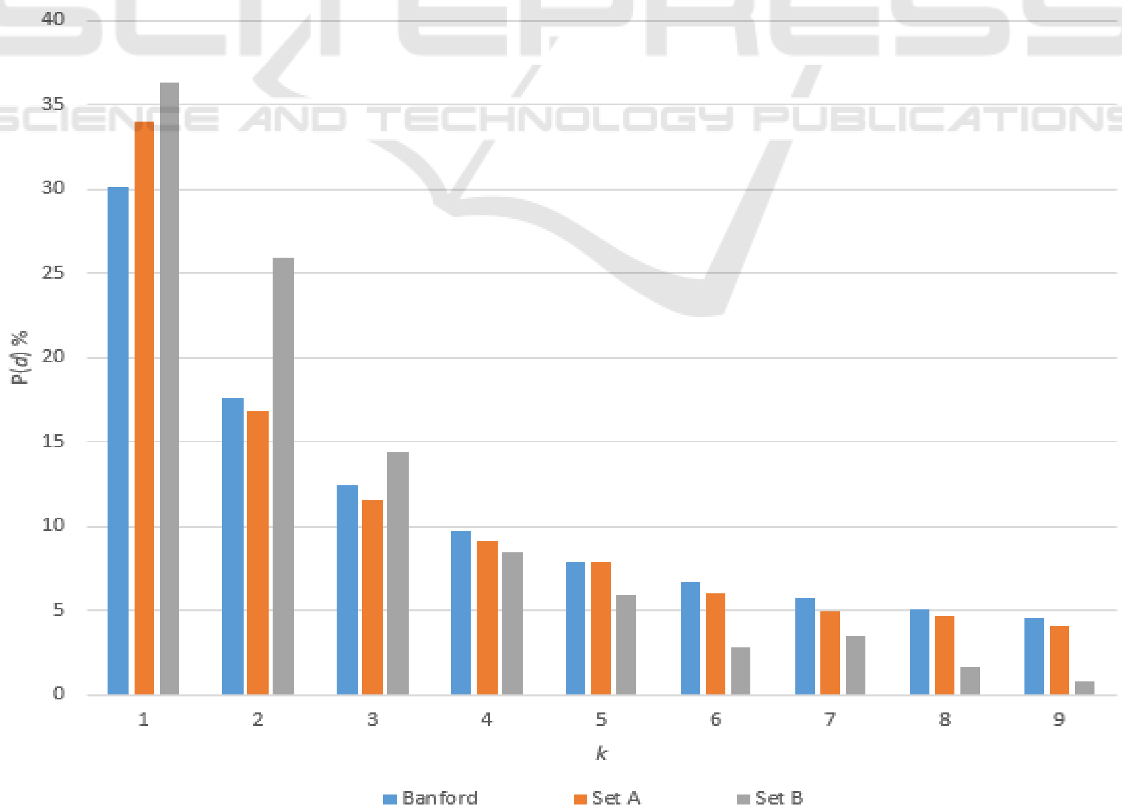

Figure 3: Distribution of Data Set.

3.2 Detecting Possible Match-fixing

In the above analysis, we confirmed that the number of rallies recorded in a tennis game with a high probability of match-fixing is different from the expected frequency predicted by the Benford's Law (Figure 3). Based on lessons learned, we develop a machine-learning-based model to detect whether a given game is fixed or not. The model used for this study contained an h-rally frequency of first digit feature. The performance results of the proposed model are presented in the following section.

First, the learning process was performed by applying 113 sets that account for the 50% of "Dataset A" gathered from ATP and WTA, and "Dataset B1" collected from fixed matches. We apply the artificial neural network as a classifier. The nine expected frequencies of Benford's Law were used in the model. The number of layers of the artificial neural network was set to 10. The training process is presented in Figure 4. The model A was obtained by applying the above learning method.

Second, the evaluation data was composed of 113 games that amount to the remaining 50% of "Dataset A" and collected "Dataset B2". The corresponding test results (game number and the possibilit of its match-fixing) are shown in Figure 5. Note that the last three games (no. 111, 112, and 113) were the match-fixed ones. As shown in Figure 5, the last three sets of values show a near 1.0 value (.948, .972, and .941). The other 110 games are identified as normal ones.

This shows that our proposed model can successfully deliver a positive outcome for detecting all three fixed matches.

4 CONCLUSIONS

In this study, Benford's Law and machine learning were used to examine the possible detection of match-fixing. Benford's Law is an empirical rule and a phenomenon that occurs every day in nature. Experimental results of this research show that the numbers collected from fixed games and the expected frequencies predicted by Benford's Law are different. However, it is not easy to simply say that a game is fixed "simply because it does not follow Benford's law". Therefore, in this study, we made a detection model for match-fixing through machine learning and confirmed the results by analyzing the test data. Experimental results show that fixed games can be detected by the proposed model implemented through machine learning.

"Dataset B" used in this study is collected from artificially fixed matches and it cannot be guaranteed that a similar set of data will appear in actual games. However, improved results will be obtained by constructing the judgment model for machine learning more precisely and using other features in addition to the expected frequencies of the Benford's Law used in this study.

Table 1: Comparison of Benford's Law and Data Set A, B.

first digits(d)	Probability for Benford's law	Data Set A	Data Set B
1	30.10	33.97313	36.29547
2	17.61	16.84369	25.93724
3	12.49	11.55616	14.35712
4	9.69	9.180898	8.49764
5	7.92	7.869342	5.915024
6	6.69	6.082734	2.804777
7	5.8	4.957067	3.526798
8	5.12	4.667906	1.638434
9	4.58	4.110236	0.827492
n		220	6
mean		58.21	59.26
sd		47.77	67.75
x^2		.439	9.210
p		.999	.324

```
# weights:  131
initial  value 28.500480
iter  10 value 3.055997
iter  20 value 2.998755
iter  30 value 2.993691
iter  40 value 2.981852
iter  50 value 2.944008
iter  60 value 2.540907
iter  70 value 0.251033
iter  80 value 0.218691
iter  90 value 0.204201
iter 100 value 0.202962
iter 110 value 0.202942
final  value 0.202939
converged
```

Figure 4: Training result.

Sports such as horse racing and motorboat racing play a major role in sports betting and strict monitoring and control are being systematically implemented to prevent match-fixing. However, organizations that make profit illegally through match-fixing are still making attempts to manipulate games while evading the surveillance network. Since the match-fixing attempts are carried forward covertly, the cases exposed by the monitoring agencies are very rare, which implies that the match-fixing is occurring even at this moment.

1	0.079276113	94	0.001778566
2	0.019730565	95	0.032312940
3	0.073483354	96	0.032312940
4	0.015390691	97	0.001671565
5	0.070259913	98	0.012914062
6	0.067507593	99	0.011867684
7	0.062596705	100	0.001697511
8	0.064375336	101	0.001719169
9	0.011756237	102	0.011938163
10	0.036359412	103	0.032112234
11	0.011563981	104	0.011431564
12	0.064192517	105	0.001640597
13	0.010314325	106	0.001528148
14	0.008399448	107	0.030354372
15	0.059460804	108	0.010658210
16	0.053702679	109	0.001368157
17	0.053702679	110	0.001354006
18	0.007771323	111	0.948129456
19	0.053547749	112	0.972948381
20	0.007237585	113	0.941203181

Figure 5: Testing result.

Sports games classify their results as data. Until recently, it was very difficult to detect match-fixing attempts and therefore, punitive measures were difficult to implement. The lack of punishment has inevitably made the eradication of match-fixing more challenging. However, various scientific analysis techniques that can determine match-fixing attempts through simple data analysis are being studied extensively. If these methods can be successfully implemented in the sports policies, match-fixing attempts could get completely eradicated as a result. The initial prevention of match-fixing attempts will be able to help sports advance a step further.

ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their helpful comments on an earlier version of this paper.

REFERENCES

Asser Institute., 2014. *Study on risk assessment and management and prevention of conflicts of interest in the prevention and fight against betting-related match fixing in the EU 28*. Asser Institute. Luxembourg.

Aquilina, Dawn, and Chetcuti, Angelo. 2014). Match-fixing: the case of Malta. *International Journal of Sport Policy and Politics*. 6(1). 107-128.

Bag, Parimal Kanti. and Saha, Bibhas., 2011. Match-fixing under competitive odds. *Games and Economic Behavior*. 73(2011). 318–344.

Boeri, Tito. and Severgnini, Battista., 2011. Match rigging and the career concerns of referees. *Labour Economics*, 18(2011). 349–359.

Carpenter, Kevin, 2013. Match-Fixing: The Biggest Threat to Sport in the 21st Century?. *Sweet and Maxwell's International Sports Law Review*. 2012(2). 13-24.

Choi, C. H., and Park, J. H., 2017. The Possibility of Detecting Match-fixing: Benford's Law in Sports Data. *The Korean Jounml of Measurement and Evaluation in Physical Education and Sport Science*. 19(1). 69-89.

Department For Culture Media And Sport: UK., 2010. *Report of the Sports Betting Integrity Panel*. Department For Culture Media And Sport. London.

European Commission, 2012. *Match fixing in sport*. http://ec.europa.eu/assets/eac/sport/library/studies/study-sports-fraud-final-version_en.pdf

European Commission, 2014. *Expert Group on Match Fixing - State of Play on the fight against match fixing*. European Commission. Luxembourg.

European Commission, 2018. *Match fixing*. https://ec.europa.eu/sport/policy/organisation-of-sport/match-fixing_en

FIFA, 2017. *FIFA ACTIVITY REPORT 2017*. FIFA. Zurich.

Forbes, 2013. *How The 92nd-Ranked Tennis Player In The World Earns A Comfortable Living*. https://www.forbes.com/sites/miguelmorales/2013/08/26/aces-into-assets-how-michael-russell-has-made-a-profitable-career-in-the-demanding-world-of-pro-tennis/#f585f5247548

Hill, Declan, 2010. *The Fix: Soccer and Organized Crime*. McClelland and Stewart. Toronto.

Hill, Declan, 2013. *The Insider's Guide to Match-Fixing in Football*. Anne McDermid and Associates Limited. Toronto.

Hill, T. P., 2011. Benford's Law Blunders. *The American scientist.* 86(4). 358-363.

HM Government., 2017. *UNITED KINGDOM ANTI-CORRUPTION STRATEGY 2017-2022*. HM Government. London.

Huffington post, 2016. *Tennis Match-Fixing Claims Involve Wimbledon And French Open Competitors*. https://www.huffingtonpost.co.uk/2016/01/18/tennis-match-fixing-wimbledon-french-open-competi tors_n_9 0 06 842.html

IOC., 2016. *Rules for the Application during the Games of the XXXI Olympiad in 2016 in Rio de Janeiro of the Articles 7, 9 and 10 of the IOC Code of Ethics and of the Olympic Movement Code on the Prevention of the Manipulation of Competitions*. https://stillmed.olympic.org/media/Document%20Library/OlympicOrg/Documents/Code-of-Ethics/2016/Rules-on-the-Prevention-of-Manipulation-Rio-2016.pdf#_ga=2.185496480.867491679.1524447313-1193877039.1524447313

Katsarova, Ivana, 2016. Match-fixing: Issues and policy responses. *European Parliamentary Research Service*. 2016(April). 1-10.

Lastraa, Reynald, Bellb, Peter, and Bonda, Christine, 2018. Sports betting and the integrity of Australian sport: Athletes' and non-athletes' perceptions of betting-motivated corruption in sport. *International Journal of Law, Crime and Justice*. 52(2018). 185–198.

Moriconi, Marcelo, 2018. The official football match-fixing prevention discourse as a cognitive limitation (the cases of Iberian countries). *Soccer and Society.* 19(2). 271-287.

Nigrini, Mark, 2012. *Benford's Law*. WILEY. New Jersey.

Rodenberg, Ryan M., and Feustel, Elihu D., 2014. Forensic Sports Analytics Detecting And Predicting Match-fixing In Tennis. *The Journal of Prediction Markets*. 8(1). 77-95.

The Korea Times, 2012. *Over 10 footballers under probe for match-rigging scam*. http://www.koreatimes.co.kr/www/nation/2018/04/113_87763.html

TIU, 2018. *TENNIS INTEGRITY UNIT Annual Review 2017*. TENNIS INTEGRITY UNIT. London.