



Are betting returns a useful measure of accuracy in (sports) forecasting?

Fabian Wunderlich^{*}, Daniel Memmert

Institute of Training and Computer Science in Sport, German Sport University Cologne, Cologne, Germany



ARTICLE INFO

Keywords:

Sports forecasting
Error measures
Betting market
Betting odds
Betting returns
Market efficiency
Profitability of forecasts
Statistical measures
Economic measures
Simulation

ABSTRACT

In an economic context, forecasting models are judged in terms not only of accuracy, but also of profitability. The present paper analyses the counterintuitive relationship between accuracy and profitability in probabilistic (sports) forecasts in relation to betting markets. By making use of theoretical considerations, a simulation model, and real-world datasets from three different sports, we demonstrate the possibility of systematically or randomly generating positive betting returns in the absence of a superior model accuracy. The results have methodological implications for sports forecasting and other domains related to betting markets. Betting returns should not be treated as a valid measure of model accuracy, even though they can be regarded as an adequate measure of profitability. Hence, an improved predictive performance might be achieved by carefully considering the roles of both accuracy and profitability when designing models, or, more specifically, when assessing the in-sample fit of data and evaluating out-of-sample forecasting performances.

© 2019 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

The huge and growing sports betting market (European Sport Security Association, 2014) can be considered an important factor in highlighting the relevance of sports forecasting. Bookmakers and professional gamblers are in need of powerful forecasting models (Goddard, 2005; Koopman & Lit, 2019; McHale & Morton, 2011), with the goal of gaining a competitive advantage over one another. Moreover, the sports betting market provides an appropriate environment in which for economists to test market efficiency (Angelini & De Angelis, 2019; Goddard & Asimakopoulou, 2004; Gray & Gray, 1997) in a rather easily observable real-world setup, and to analyse specific market inefficiencies (Braun & Kvasnicka, 2011; Snowberg & Wolfers, 2010).

While the financial impact of the sports betting industry highlights the relevance of research on sports

forecasting, the betting odds provided by bookmakers or betting exchanges also serve as a valuable benchmark for testing the predictive power of forecasting methods. There are several reasons for the popularity of betting odds as a benchmark. First, betting odds are easily observable, since they are displayed on any bookmaker's website. Second, betting odds can be assumed to have a high level of accuracy, as inaccurate odds could have a negative financial impact on bookmakers. Moreover, comprehensive scientific evidence shows a high predictive power of betting odds in real-world datasets across various sports (Forrest, Goddard, & Simmons, 2005; Kovalchik, 2016; Song, Boulier, & Stekler, 2007; Štrumbelj & Vračar, 2012). Finally, betting odds provide the possibility of earning money, and thus can be used to assess the financial benefit of a forecasting model.

The natural goal of forecasting is to achieve a high predictive power. Two sorts of measures are used commonly to assess the power of probabilistic sports forecasting methods: statistical measures and economic measures. Statistical measures are based on the idea of comparing the forecast (i.e., the probability of an outcome) to the

^{*} Corresponding author.

E-mail address: f.wunderlich@dshs-koeln.de (F. Wunderlich).

actual outcome. A high accuracy is achieved if the actual outcomes after the events show a substantial overlap with the probabilities of these outcomes that were forecast prior to the events. Economic measures are based on the idea of winning money by using the forecasts to construct profitable betting strategies. The betting odds offered by bookmakers or betting exchanges are judged considering the forecasts made and profitable bets are identified. Using the actual outcomes makes it possible to calculate the betting returns that would have been realised if these bets had been placed prior to the events.

A common approach that has been used in the previous literature is to make use of both statistical and economic measures to validate forecasting models with regard to separate aspects (Boshnakov, Kharrat, & McHale, 2017; Goddard & Asimakopoulou, 2004; Koopman & Lit, 2015; McHale & Morton, 2011). While the two types of measures are often used concurrently, authors do not always distinguish carefully between accuracy and profitability when evaluating the results. This poses the risk that positive betting returns may be equated – either explicitly or implicitly – with a high forecasting accuracy. Koopman and Lit (2015, p. 1) state that “the forecasts from this model are sufficiently accurate to gain a positive return over the bookmaker’s odds”. Lessmann, Sung, and Johnson (2010, p. 519), with reference to the work of Leitch and Tanner (1991), argue that “a model’s profitability is the primary indicator of forecasting accuracy” in the context of horse racing. As a consequence, economic measures are also referred to in studies where the main focus is on a forecasting method rather than on market efficiency (Lessmann et al., 2010; McHale & Morton, 2011).

Empirical results cast doubt on the supposedly unambiguous connection between positive betting returns and forecasting accuracy in terms of statistical measures. A forecasting model based on community-based market value estimation was shown to create positive betting returns even though the betting odds showed a superior forecasting accuracy in terms of statistical measures (Peeters, 2018). The same counterintuitive result was found in a forecasting model for NFL games, reporting positive betting returns despite the fact that betting odds were superior in picking the winner of a match (Baker & McHale, 2013). Different results with regard to distinct measures are usually accepted without analysing reasons or discussing implications in detail. This poses the question of how it is possible that betting odds frequently outperform forecasting models in terms of forecasting accuracy (as has already been discussed), while a wide variety of studies claim positive betting returns against the odds (Constantinou, Fenton, & Neil, 2012; Koopman & Lit, 2015).

Evidently, the results reviewed indicate that positive betting returns may be generated in the absence of a superior forecasting model. A detailed understanding of the reasons is complicated by the fact that the datasets used in the literature are rather heterogeneous, for example due to using a variety of different sports. Moreover, different time periods and bookmakers might complicate matters. In this respect, researchers report varying margins, such as a margin of around 25% for the German

state-owned bookmaker Oddset (Spann & Skiera, 2009), a margin of around 16% for the bookmaker Interwetten (Forrest & Simmons, 2008) or a margin of around 6% for betting odds averaged over several bookmakers (Wunderlich & Memmert, 2018). Likewise, there are massive differences in the numbers of bets considered in betting strategies, with some reporting betting returns from more than 4000 bets per model (Lessmann et al., 2010), while others investigate a few hundred bets (Peeters, 2018), and most strategies in some studies consisting of fewer than 100 bets (Forrest & Simmons, 2008; McHale & Morton, 2011).

Considering the reviewed literature, it is evident that the role of betting returns in judging forecasting models has not been sufficiently questioned and discussed. The present study therefore attempts to close this research gap by shedding light on the complex and counterintuitive connection between the forecasting profitability in terms of economic measures and the forecasting accuracy in terms of statistical measures. We contribute to the forecasting research by evaluating the following questions in relation to the characteristics of betting markets and profitable betting strategies.

- (a) Do positive betting returns imply a better forecasting accuracy in terms of statistical measures?
- (b) What systematic reasons may there be for positive betting returns in the absence of a superior forecasting model?
- (c) How can the role of randomness in the generation of positive betting returns be assessed adequately?

2. Materials and methods

One part of this study is based on theoretical considerations, mathematical derivation and theoretical examples, and thus does not require any specific data. Another part is based on simulation results, and makes use of three different real-world datasets that represent typical examples of the datasets that are used within sports forecasting. The terminology and notation used for the theoretical considerations, as well as the data used for the simulation results, are explained in the subsequent sections.

2.1. Terminology and notation

The terminology and notations used within this study are presented in Table 1 and explained below. There are several possible choices to bet on for each betting opportunity, e.g. *Player1* and *Player2* for a tennis match; *Home*, *Draw* and *Away* for a soccer match; or several possible winners for a horse race. We will refer to these choices as *outcomes* of the event. A common synonym for a betting opportunity with n outcomes is *n-way betting market*, and this term will also be used in this paper. Moreover, three different sets of probabilities need to be defined: the true but unknown probabilities with respect to the outcomes, and two sets of probability estimations by the bookmaker and the forecasting model, i.e., the bettor. The betting odds are calculated by inverting the bookmaker’s probability estimation and including a bookmaker margin

Table 1
Terminology and notation.

Terminology	Notation
Outcomes for a specific betting opportunity	$o_1 \dots o_n$
True probabilities	$p_1 \dots p_n$
Probabilities estimated by the bookmaker	$\bar{p}_1 \dots \bar{p}_n$
Fair betting odds	$\bar{b}_i = \frac{1}{\bar{p}_i}$ for $i \in \{1 \dots n\}$
Bookmaker margin	m
Betting odds	$b_i = \bar{b}_i (1 - m)$ for $i \in \{1 \dots n\}$
Probabilities estimated by the forecasting model	$\tilde{p}_1 \dots \tilde{p}_n$
Expected value of a bet	$e_i = b_i \cdot p_i - 1$ for $i \in \{1 \dots n\}$
Estimated expected value by bookmaker or forecasting model	$\bar{e}_i, \tilde{e}_i$ defined analogously

m , which is restricted to the interval (0,1) and commonly located between 0.05 and 0.1 in real-world applications related to high-class sports events. The margin is distributed equally across all outcomes, in line with the most common approach in linking betting odds to probabilities. It must be mentioned that there is reasonable criticism of this simplistic approach; a more detailed discussion and an advanced approach can be found in the work of Štrumbelj (2014, 2016). However, the basic findings of this study are not dependent on the exact definition of the margin. Each possible bet has an expected value (given in relation to a wager of one unit), and is called *profitable* if this value exceeds zero. Please note that the fair betting odds are not fair in the sense of having an expected value of zero, but fair in having an expected value of zero regarding the probability estimation by the bookmaker.

2.2. Real-world datasets

Three datasets have been analysed for the purpose of this study, covering different sports as well as varying sizes and characteristics. The first dataset includes all matches from seasons 2007/2008 until 2016/2017 of four major European soccer leagues (English Premier League, Spanish Primera Division, German Bundesliga, Italian Serie A). The data were obtained from <http://www.football-data.co.uk>. Discarding two matches that were decided by federation decision, the dataset consists of a total of 14,458 matches. The second dataset consists of professional tennis matches played in official matches of the WTA (Women's Tennis Association) tour 2018. Discarding 90 matches that were decided by walkover or retirement and one match without betting odds left a total of 2378 matches as part of the dataset. The data were obtained from <http://www.tennis-data.co.uk>. The third dataset consists of American football matches from the 2017/2018 NCAA College Football League, including regular season matches from all conferences and divisions, as well as bowl and playoff matches. The data were obtained from <http://www.oddsportal.com>. Discarding the 65 matches with odds missing for one or both teams, the dataset includes a total of 810 matches. The soccer dataset consists of three-way betting odds, while the tennis and American football datasets consist of two-way betting odds. A detailed overview of the characteristics of each dataset can be found in Table 2.

3. Theoretical considerations

From a theoretical point of view, three main reasons drive the ambiguity between statistical and economic measures and could enable positive betting returns to be obtained in the absence of a more accurate model. First, setting betting odds can be seen as a more difficult task than exploiting betting odds. Second, in betting strategies only a small number of bets are picked out and evaluated, while in statistical measures all possible bets are taken into account. Third, inaccurate probability estimation can still lead to a correct identification of profitable bets.

We substantiate these arguments in the remainder of this study by considering a slightly simplified model of the interaction between bookmakers and forecasters. We look at the betting odds as inverted probability estimation by the bookmaker, including a bookmaker margin. Simplification in terms of the margin calculation has been discussed in Section 2.1. Moreover, the emergence of betting odds will be more complicated in practice, and might be influenced by risk management decisions in relation to the bettor's behavior (i.e., the idea of balancing a book), marketing decisions (i.e., the idea of advertising particularly high odds) and the market situation (i.e., a reaction to the odds offered by other bookmakers). However, as a simplified model, we consider betting odds to be a pure reflection of the bookmaker's probability estimation, in contrast to the probability estimation of a bettor's forecasting model.

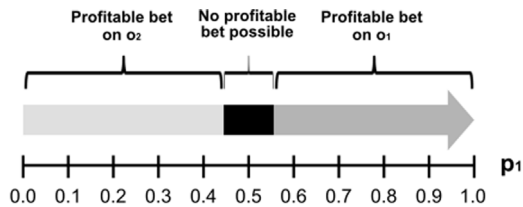
3.1. The asymmetry of the betting market

Estimation questions are a popular tool for avoiding ties in TV game shows. To pick a winner, the candidates are asked to estimate the answer to a question that none of them is expected to know exactly. Typical examples would be "How many e-mails are sent in one day?" or "How many cars are sold in France per year?". Usually, both candidates are asked to write down their estimates at the same time, in order to avoid one candidate being influenced by the other candidate's choice. This procedure follows from the intuitive idea that making the estimates one after the other would give the second candidate an unfair advantage. In fact, giving one candidate the opportunity to react to the other candidate's original estimate would simplify the task drastically by reducing it to the

Table 2

Overview of the real-world datasets and their characteristics.

Dataset	Number of matches	Outcomes	Average margin	Average odds			Percentage of clear longshots	
				Favourite	Overall	Outsider	>10	>20
Soccer	14,458	Home, Draw, Away	6.0%	1.90	3.73	5.53	3.4%	0.4%
Tennis	2,378	Player1, Player2	4.9%	1.48	2.34	3.20	0.5%	0.0%
American football	810	Team1, Team2	4.2%	1.32	4.09	6.86	7.5%	2.7%

**Fig. 1.** Profitable bets depending on the true probability of outcome 1, referring to estimated bookmaker probabilities of 50%, 50% and a 10% margin.

question: “Is the correct answer lower or higher than the first estimate?”.

A similar mechanism exists in sports betting markets where the bookmaker needs to offer betting odds first and the bettor can react afterwards by deciding whether or not to place a bet. The bookmaker is left with the rather complex task of estimating accurate betting odds, while the bettor need merely decide whether the betting odds have been set too high or too low. We will refer to this principle as *asymmetry of the betting market*. However, in contrast to the example of the estimation questions, the bookmaker has the advantage that he can include a margin in the betting odds.

Assuming that the bookmaker estimation for a betting opportunity with two outcomes is $\bar{p}_1 = \bar{p}_2 = 0.5$ and the margin is $m = 0.1$, resulting in betting odds of $b_1 = b_2 = 1.8$, a profitable bet for the bettor exists if the unknown (true) probability of outcome 1 (p_1) is smaller than 0.4 or greater than 0.6, giving the bookmaker quite a small error margin in order to avoid the presence of profitable bets. Fig. 1 illustrates this example graphically, but a more formal and detailed explanation of the asymmetry that includes bets with more than two outcomes can be found in Appendix.

3.2. The profitability paradox

The idea of betting strategies is to exploit betting odds that do not reflect the true underlying probability correctly. If a bettor's model is capable of identifying such bets, the accuracy of the model itself is of secondary importance. We demonstrate this by considering a three-way betting market, as usual when betting on the result of a soccer match ($o_1 = \text{Home}$, $o_2 = \text{Draw}$, $o_3 = \text{Away}$). Moreover, we assume the following true probabilities ($p_1 = p_2 = p_3 = \frac{1}{3}$), estimation by the bettor ($\tilde{p}_1 = \tilde{p}_2 = 0.4$, $\tilde{p}_3 = 0.2$), estimation by the bookmaker ($\bar{p}_1 = 0.2$, $\bar{p}_2 = \bar{p}_3 = 0.4$), and a margin of $m = 0.1$, resulting in the following betting odds ($b_1 = 4.5$, $b_2 = b_3 = 2.25$), as well as

the following expected betting returns ($\tilde{e}_1 = 0.8$, $\tilde{e}_2 = -0.1$, $\tilde{e}_3 = -0.55$). It is evident from the true probabilities that both sides made equal forecasting errors, and thus should be judged as having equal forecasting accuracies. However, the bettor will chose to bet solely on o_1 , yielding an expected win of

$$e_1 = b_1 \cdot p_1 - 1 = \frac{4.5}{3} - 1 = 0.5 > 0$$

per unit, and thus leading to systematic profits. But, what if the bettor and the bookmaker were to switch roles? The bettor – inspired by his or her profitable betting strategy – takes the role of the bookmaker him- or herself and offers betting odds, while the bookmaker takes the role of the bettor and tries to beat the (new) bookmaker. This leads to a completely analogous situation in which the bookmaker (as bettor) choses to bet on o_3 and not bet on the other two outcomes, resulting in an expected win of $0.5 > 0$.

While the estimation of the bettor can be used to obtain a profitable betting strategy against the estimation of the bookmaker, the same would be true the other way around, a phenomenon that we refer to as the *profitability paradox*. In other words, both models can be used successfully to exploit the weaknesses of the other model by identifying correctly the fraction of bets that are profitable.

4. Calculation and results

4.1. Differences between statistical and economic measures

The profitability paradox clearly shows that positive betting returns can be generated in the absence of a superior forecasting model. However, it is based on a simplistic example that would not occur in exactly the same way in a real-world scenario. We therefore develop a more complex theoretical model for simulating the estimation errors of bookmakers and bettors, in order to systematically compare statistical and economic measures. The simulation is not restricted to a specific type of betting opportunity with a specific number of outcomes. Instead, we model probabilities, betting odds and estimation errors for individual outcomes that could be part of a betting opportunity with any number of outcomes.

The simulation process requires the random generation of probabilities and errors. Unfortunately, the distribution of true probabilities depends strongly on the situation (i.e., sports and level of detail), and the estimation errors are not observable in real-world datasets, meaning that assumptions regarding the distributions of probabilities and errors are needed. Our approach draws

the true probabilities from a uniform distribution and uses a bivariate normal distribution model with logit transformation to include estimation errors in the probability estimates of bookmakers and bettors.

$$\begin{aligned}
 p &\sim U(0, 1) \\
 \bar{p} &= \frac{1}{1 + e^{-(p^* + \bar{X})}} \\
 \tilde{p} &= \frac{1}{1 + e^{-(p^* + \tilde{X})}} \\
 &\text{where} \\
 p^* &= \ln\left(\frac{p}{1-p}\right) \\
 (\bar{X}, \tilde{X}) &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \bar{\sigma}^2 & \bar{\sigma}\tilde{\sigma}\varrho \\ \bar{\sigma}\tilde{\sigma}\varrho & \tilde{\sigma}^2 \end{pmatrix}\right).
 \end{aligned}$$

The model is based on error variables for the bookmaker $\bar{X}$ and the bettor $\tilde{X}$ that follow bivariate normal distributions with standard deviations of $\bar{\sigma}$ and $\tilde{\sigma}$, and a correlation of ϱ . As normally-distributed error variables can be infinitely large, a logit transformation is performed in order to ensure that the estimated probabilities of both the bookmaker and the model are located in the interval $[0,1]$. The standard deviations $\bar{\sigma}$ and $\tilde{\sigma}$ can be seen as measures of estimation errors, and are referred to subsequently as *bookmaker error* and *bettor error*, or *forecasting errors* if referring to both.

The simulation procedure consists of the following steps:

(1) Choose the number of probabilities in each simulation (n), the bookmaker error ($\bar{\sigma}$), the bettor error ($\tilde{\sigma}$), the correlation of errors (ϱ) and the bookmaker margin (m).

Repeat steps (2) to (5) n times:

(2) Draw the true probability p , as well as the probability estimations by the bookmaker ($\bar{p}$) and the bettor ($\tilde{p}$), by using a random number generator and working in compliance with the pre-specified probability distributions.

(3) Derive the betting odds ($b = \frac{1}{\bar{p}} \cdot (1 - m)$) from the probability estimation of the bookmaker and the margin.

(4) Check whether the bettor ascribes a positive value to the betting opportunity ($\tilde{p} \cdot b - 1 > 0$), and place a bet accordingly.

(5) Calculate the true expected value of the bets placed by the bettor based on a wager of one unit ($p \cdot b - 1$).

(6) Return the mean relative (expected) betting returns, i.e., the average of the true expected values of all bets placed by the bettor.

The mean relative betting returns that result from the simulation are presented as an economic measure of the forecasting profitability. Moreover, a statistical measure is required, and the forecasting errors $\bar{\sigma}$ and $\tilde{\sigma}$ are presented as a statistical measure of the accuracy (where a lower error implies a higher accuracy). However, some caution is warranted, as these errors cannot be observed in real-world applications, and thus it could be argued that they do not apply to practical situations. We therefore tested whether the forecasting errors are in line with a common

statistical measure of the forecasting accuracy, the mean squared error, defined as

$$\overline{MSE} = \frac{1}{n} \sum_{i=1}^n [(p_i - \bar{p}_i)^2],$$

where n is the number of probabilities simulated (i.e., possible bets). For the forecasting model, the error MSE is defined analogously, considering $\tilde{p}_i$. Squared errors are a common accuracy measure for probabilistic forecasts, although they are typically used with slightly different definitions and labels, such as the mean squared error of Lasek, Szlávik, and Bhulai (2013), the Brier score (Cattelan, Varin, & Firth, 2013) or the quadratic loss function (Štrumbelj & Vračar, 2012). The simulation confirmed that squared errors are in fact just a reflection of the forecasting errors on a quadratic scale. We decided to show the more intuitive forecasting errors (i.e., standard deviations $\bar{\sigma}$ and $\tilde{\sigma}$) in the plots, instead of the mean squared errors.

Fig. 2 shows a contour plot of the mean relative betting returns for various (uncorrelated) bookmaker errors and model errors. The plot is based on a simulation of each combination of forecasting errors in steps of 0.005, generating $n = 10,000,000$ probabilities (i.e., possible bets) in each simulation and using a margin of $m = 0.05$.¹ With respect to Table 2, this margin seems a realistic choice when referring to betting odds in relation to high-class sports events.

The plot shows an indisputable connection between accuracy and profitability, as higher bettor errors imply lower betting returns if the bookmaker error is kept constant. However, the asymmetry of the betting market becomes visible when it is seen that bettor models with higher forecasting errors than the bookmaker can still be capable of generating positive returns. If the errors of the bookmaker and the bettor are comparable, the simulation shows that returns decrease with decreasing forecasting errors on both sides.

Another aspect that has to be taken into account is the interaction (i.e., correlation) of errors. Fig. 3 presents the mean relative betting returns for various forecasting error scenarios and degrees of correlation. Considering the scenarios with equal forecasting errors (Fig. 3a), decreasing the correlation increases the betting returns. This can be explained with reference to bettors specifically trying to exploit certain bookmaker errors. Thus, a search for profitable bets can be considered as a search for bets where the probability is undervalued by the bookmaker, and therefore the odds are overvalued. If the bookmaker's and the bettor's errors are highly correlated, such bets are probably not identified. However, profitable bets will be identified if the bettor overvalues exactly those probabilities that are undervalued by the bookmaker (even if the bettor's forecast is inaccurate as well). A similar pattern can be observed for scenarios where the bookmaker error is smaller than the bettor error (Fig. 3b). This reasoning is basically in line with the results of Hubáček,

¹ All simulations used within this study were implemented in the R system (version 3.4.2) for statistical computing (R Core Team, 2017). The source code will be made available on reasonable request.

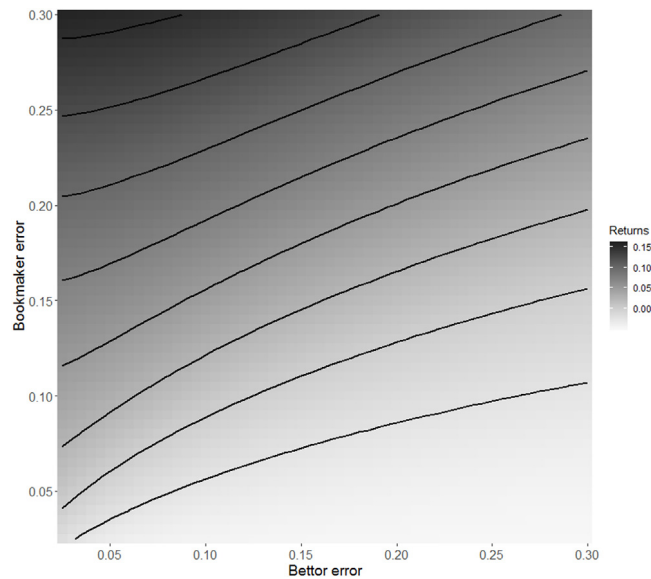


Fig. 2. Contour plot of the mean relative betting returns with regard to various uncorrelated forecasting errors of the bettor and bookmaker. The black lines illustrate levels of betting returns in steps of 2.5%, with the lowest line referring to returns of –2.5% and the highest line to returns of 15%.

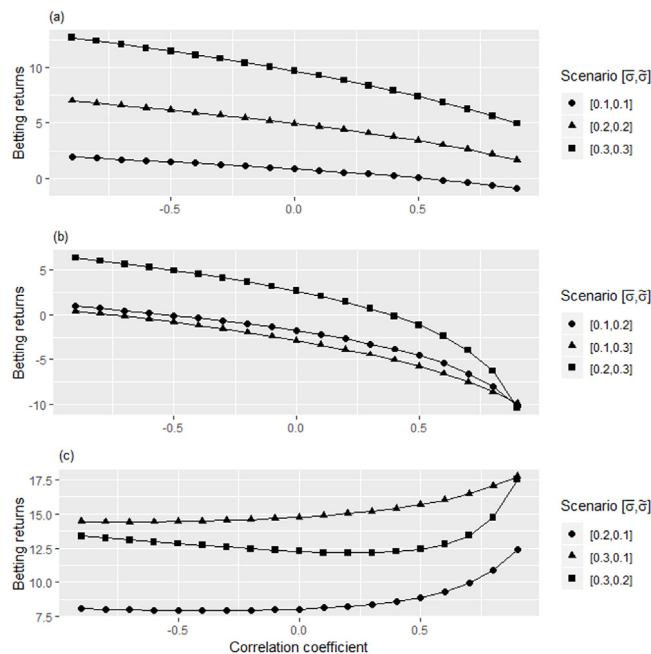


Fig. 3. Mean relative betting returns for various forecasting error scenarios and various degrees of correlation. The numbers in each scenario description refer to the bookmaker and bettor errors (in that order). Thus, (a) refers to scenarios with equal accuracies, (b) refers to scenarios where the bookmaker is more accurate and (c) refers to scenarios where the bettor is more accurate.

Šourek, and Železný (2019), who showed that decorrelation approaches increased the betting profits in a real-world forecasting model in basketball. While comparable, it should be mentioned that their model decorrelates the forecasted probabilities, whereas the plot in Fig. 3 is based on correlations of errors. The pattern of increasing returns for decreasing correlations breaks down for scenarios where the bettor errors are smaller than the

bookmaker errors (see Fig. 3c). However, such scenarios are rather unlikely to appear in real-world applications with reference to the sports forecasting literature.

In summary, the simulation results show that betting returns are influenced positively by a high accuracy of the bettor's model if other aspects (in particular bookmaker errors and the correlation) are kept constant. However, they are also influenced by the correlation between

bookmaker and bettor errors and show a kind of skewed pattern, as positive betting returns do not imply more accurate forecasts.

4.2. Significance of betting returns

Although we have argued that betting returns (by themselves) are problematic as a measure of the forecasting accuracy, their usefulness in evaluating the profitability of forecasts or judging market efficiency remains valid. However, as in a statistical test, researchers should also attend to the question of whether or not the returns reported could be a result of coincidence. Having constructed a betting strategy with a certain betting return, it seems useful to have a measure of how likely it is that the observed return could be explained by coincidence. In this respect, we follow the idea of a p -value in a statistical test and introduce p_{bs} , which we define as the probability that a random betting strategy would have produced the same or a higher return.

This probability can be expected to depend on the number of bets in the betting strategy, the margin included in the betting odds and the diversity of the betting odds (e.g. the existence of extreme longshots). Thus, it depends on the specific dataset, and no generally-valid value can be determined. When estimating p_{bs} , the full set of betting odds and observed results is needed. Let n_{bets} be the number of bets that are actually performed based on the betting strategy. Then, n_{bets} events are chosen randomly from the dataset and one outcome is chosen randomly to bet on. As a consequence, a random betting strategy with the same number of bets as the original strategy is constructed and the betting returns achieved by this strategy can be derived. Repeating the simulation a large number of times enables us to calculate the percentage of random betting strategies that result in returns that are equal to or larger than those of the original betting strategy (i.e., p_{bs}).

Fig. 4 illustrates the values of p_{bs} for each of the datasets described in Section 2.2, referring to different numbers of bets and different relative betting returns based on 100,000 simulations each. In addition, 95% confidence intervals are presented, based on a bootstrapping procedure that generates 10,000 resamples of the simulated returns for each sport and number of bets. The results of the simulation underline the high risk of generating positive betting returns by pure random selection. When applying a threshold of 5% (as is common in significance tests), even quite impressive relative betting returns of 0.1 in 100 bets in tennis, 0.15 in 100 bets in soccer, or 0.2 in 200 bets in American football could not exclude randomness as a reason with an acceptably high degree of certainty. The 95% confidence intervals are most pronounced for American football and for returns based on low numbers of bets. All in all, they are hardly noticeable in the plot, demonstrating that the results are robust.

The differences between the three figures show that the risk of generating returns randomly can depend heavily on the dataset. While the datasets from soccer and tennis show roughly comparable patterns, the American

football dataset is more prone to random positive betting returns. This can be explained by the greater heterogeneity of team skills in college football (on an NCAA level) compared to tennis (on a WTA tour level) and soccer (in the European top leagues), resulting in a larger number of clear longshots, as demonstrated in Table 2. Consequently, a single successful bet on such an event can lead to significant returns, influencing the overall result strongly, and thus also affecting the probability of gaining overall positive returns. While the datasets are heterogeneous with reference to the existence of longshots, all three contain fairly similar and low margins (see Table 2). Adding datasets with higher margins would be likely to result in a different pattern, and would probably lead to a decreased risk of randomly generating positive returns. Finally, we would like to mention that other approaches for assessing the random effects in betting returns do exist and have been used in the literature, such as bootstrapping methods (Boshnakov et al., 2017).

5. Discussion

It is reasonable to define financial success as the main focus of forecasting models in regard to the goals of both bookmakers and professional gamblers, or with the intention of investigating the market efficiency. However, many sports-related applications exist where accuracy rather than profitability should be the main concern of forecasts. Examples include the use of forecasts for understanding general concepts like crowd wisdom (Peeters, 2018), for validating official or unofficial ranking systems (Lasek et al., 2013; Wunderlich & Memmert, 2016), for gaining insights into the underlying sport (Štrumbelj & Vračar, 2012), for drawing conclusions on team strengths and the performance development of teams (Wunderlich & Memmert, 2018) or for fulfilling the interests of sports broadcasters (Barnett, O'Shaughnessy, & Bedford, 2011).

Although betting strategies are a standard measure in sports forecasting, positive betting returns are occasionally (mis-)used to prove the accuracy of forecasting models. From our own experience, we can say that the absence of economic measures is criticized regularly by the reviewers of journals, even if the focus of the investigation is on the forecasting accuracy. The results of this study show that there is a need to accept that forecasting accurately and forecasting profitably are different tasks, and should be treated as such. In particular, if a model's main focus is on maximizing betting profits, authors should state this clearly, to ensure that the model is not used erroneously in applications where accuracy is required. In this respect, the study adds to the growing body of literature attempting to understand this broad and complex topic. We would like to encourage subsequent studies to include further approaches in determining betting stakes, such as the much-noticed Kelly strategy, and to examine in more detail the role of correlations between the forecasting errors of bookmaker and forecasting models.

Both systematic and unsystematic influences (i.e., randomness in the betting returns) can drive the difference between profitability and accuracy. Systematic influences are characteristics of the betting market that make it possible to generate positive betting returns in the absence

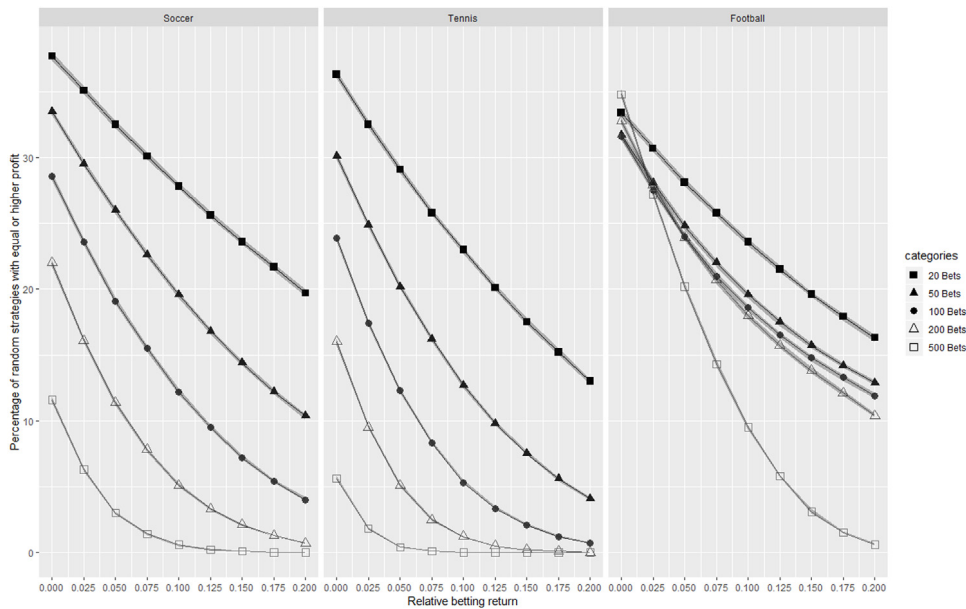


Fig. 4. The percentages of random betting strategies that reach or exceed a certain relative betting return (p_{bs}). We present results for three different datasets, as well as five different numbers of bets and nine different relative betting returns. The 95% confidence intervals are shown as grey areas.

of a superior forecasting model. One implication of this is that, when applying real-world datasets, it is easier to generate positive betting returns than to outperform betting odds in terms of the forecasting accuracy. In a more general sense, the study shows that negatively-correlated forecasting errors increase the profitability if the model (i.e., bettor) accuracy is comparable to the bookmaker accuracy. These results underline the fact that it is unreasonable to attribute superior accuracy to a model by merely referring to the existence of a profitable betting strategy.

In terms of statistical uncertainty, it is not surprising that positive returns can be generated by pure randomness for small numbers of bets. However, the fact that randomness cannot be excluded even for larger numbers of bets and fairly impressive betting returns might not be expected. We would encourage authors to calculate p_{bs} whenever evaluating the returns of a betting strategy, with the intention of giving the reader the possibility of assessing the significance of this result correctly, and to report it accordingly; for example, “*The betting strategy leads to relative betting returns of 0.12 in 347 bets ($p_{bs} < 0.01$). This means that in a simulation (100,000 repetitions), less than 1% of all random strategies generated equal or higher returns*”.

The findings of this study are not limited to the domain of sports forecasting, but can be transferred to any domain of forecasting in which betting odds are offered. The most relevant domain in this regard is betting on the outcome of political elections, which has been studied extensively (Erikson & Wlezien, 2012; Wolfers & Leigh, 2002; Wolfers & Zitzewitz, 2004), also including the investigation of betting strategies and betting returns (Rhode & Strumpf, 2004). Other examples of non-sports-related betting opportunities include further political developments like countries’ decisions to leave the European

Union, award ceremonies like Oscars and Golden Globes, or international cultural events such as the Eurovision Song Contest. Finally, although they are not directly comparable to betting markets, it should be mentioned that the differences between statistical and economic measures have been discussed in several fields of economics, such as interest rates (Leitch & Tanner, 1991) and stock markets (Granger & Pesaran, 2000).

Another effect of this study is to stimulate discussion on new approaches for the in-sample fitting of data. Fitting the data (in-sample) using statistical measures and assessing the forecasting accuracy (out-of-sample) by means of economic measures does not seem consistent. If the focus of a model is on profitability, wouldn’t it be beneficial to fit the model initially using betting odds and economic measures? The work of Hubáček et al. (2019) points in this direction and shows highly promising results. The authors clearly focus on profitability as the goal of their forecasting model, make use of strategies to reduce the correlation between the bookmaker and model forecasts in an attempt to exploit inaccurate odds, and demonstrate that this leads to improved betting returns. We would like to encourage the investigation of further strategies for using betting odds to fit forecasting models, such as maximizing the in-sample betting returns instead of maximizing the in-sample accuracy in terms of the likelihood or related measures.

6. Conclusions

In sports and other domains related to betting markets, researchers should indicate clearly whether the main purpose of a forecasting model is accuracy or profitability, and choose their measures accordingly. When designing a model to maximize profitability or evaluating market efficiency, an economic measure (like betting returns)

would be the appropriate choice. However, the informative value of economic measures in terms of forecasting accuracy should not be overestimated, and statistical measures would be a better choice when aiming for optimal accuracy.

Further, the present study has shown that forecasters need to take both systematic and unsystematic influences into account when assessing betting returns. Systematic influences refer to the fact that positive betting returns can be generated systematically without a superior forecasting method, with the betting returns depending on the correlation between the bookmaker and model errors. Unsystematic influences relate to the fact that betting returns can be the result of randomness, and an undersized number of bets may be investigated.

Finally, regarding efforts to increase the profitability of forecasting models further, it could be a promising approach to emphasize the use of betting odds in combination with economic measures for the in-sample fitting of models.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank two anonymous reviewers and the Associate Editor for their highly valuable feedback, helpful comments and useful suggestions that helped to improve an earlier version of this paper. Moreover, special thanks goes to Alessandro Seck for his support in performing simulation runs and in visualizing results.

Appendix

Section 3.1 discussed the asymmetry of the betting market. We put this idea in a more formal context and illustrate the difficulty faced by the bookmaker by defining the *area of no profitable bet*. Regarding a betting opportunity with n outcomes, the bookmaker has estimated probabilities $\bar{p}_1, \dots, \bar{p}_n$, resulting in betting odds of $b_1, \dots, b_n$. The area of no profitable bet P is defined as the set of tuples of probabilities that give the bettor no possibility of placing a profitable bet; i.e., all combinations of true probabilities that imply the absence of any profitable bet under the given betting odds:

$$P = \left\{ p \in \mathbb{R}^n \mid 0 \leq p_i \leq 1, \sum p_i = 1, p_i \leq \frac{1}{b_i} \right\}.$$

Taking a slightly simplified view of the betting market, the task of the bookmaker is to estimate the probabilities accurately enough to ensure that the true (unknown) probabilities are part of the area of no profitable bet. The area of no profitable bet and its consequences can be illustrated very clearly using a betting opportunity with only

two possible outcomes. Given such a betting opportunity with outcomes o_1 and o_2 , the area of no profitable bet is

$$P = \left\{ p \in \mathbb{R}^2 \mid 0 \leq p_i \leq 1, p_1 + p_2 = 1, p_1 \leq \frac{1}{b_1}, p_2 \leq \frac{1}{b_2} \right\},$$

which can be transformed as

$$P = \left\{ p \in \mathbb{R}^2 \mid 0 \leq p_i \leq 1, 1 - \frac{1}{b_2} \leq p_1 \leq \frac{1}{b_1} \right\},$$

and thus no profitable bet is possible if

$$p_1 \in \left[1 - \frac{1}{b_2}, \frac{1}{b_1} \right].$$

Assuming that the bookmaker's estimation is $\bar{p}_1 = \bar{p}_2 = 0.5$ and a margin of $m = 0.1$ is used, resulting in betting odds of $b_1 = b_2 = 1.8$, the area of no profitable bet is $[0.4, 0.5]$. A different margin calculation by the bookmaker would yield a slightly different interval, but would not affect the general principle. Generalizing this idea to bets with any number of possible outcomes, it can be shown that the area of no profitable bet is empty for a negative margin, consists only of one tuple of probabilities (the probability estimation of the bookmaker) for a margin of zero, and includes further tuples of probabilities and can be thought of as a polytope in n -dimensional space for a positive margin. [Theorem 1](#) states that the area of no profitable bet is empty if the margin is negative, while [Theorem 2](#) states that the area of no profitable bet consists of exactly one tuple of probabilities if the margin is zero. This tuple equals the probability estimation of the bookmaker, meaning that a profitable bet is possible if the bookmaker does not estimate the correct probabilities exactly. [Theorem 3](#) states that the area of no profitable bet includes additional tuples of probabilities if the margin is positive.

Theorem 1.

$$m < 0 \Rightarrow P = \emptyset$$

Proof.

$$\begin{aligned} m < 0 &\Rightarrow \frac{1}{b_1} + \dots + \frac{1}{b_n} \\ &= \frac{\bar{p}_1}{1-m} + \dots + \frac{\bar{p}_n}{1-m} < \bar{p}_1 + \dots + \bar{p}_n = 1 \end{aligned}$$

Assuming $\exists p \in P \Rightarrow p_1 \leq \frac{1}{b_1}, \dots, p_n \leq \frac{1}{b_n}$ and $p_1 + \dots + p_n = 1$
 $\Rightarrow \frac{1}{b_1} + \dots + \frac{1}{b_n} \geq p_1 + \dots + p_n = 1$ (in contradiction to the first column of the proof)
 $\Rightarrow P = \emptyset$.

Theorem 2.

$$m = 0 \Rightarrow P = \{\bar{p} = (\bar{p}_1, \dots, \bar{p}_n)\}$$

Proof. As a set of probabilities, the probability estimation of the bookmaker fulfils the following characteristics: $\sum \bar{p}_i = 1$ and $0 \leq \bar{p}_i \leq 1$.

The absence of a margin implies: $m = 0 \Rightarrow \bar{p}_i = \frac{1}{b_i} \Rightarrow \bar{p} \in P$.

The assumption that P contains another tuple of probabilities leads to a contradiction:

Assuming $\exists p \neq \bar{p}$ with $p \in P \Rightarrow \exists i$ with $p_i > \frac{1}{b_i}$ (in contradiction to $p \in P$)
or $\exists i$ with $p_i < \frac{1}{b_i}$ and $p_j \leq \frac{1}{b_j}$ for $j \neq i \Rightarrow \sum p_i < 1$ (in contradiction to $p \in P$).

Theorem 3.

$m > 0 \Rightarrow \exists p \neq \bar{p}$ with $p \in P$

Proof.

$$m > 0 \Rightarrow \frac{1}{b_1} + \dots + \frac{1}{b_n} = \frac{\bar{p}_1}{1-m} + \dots + \frac{\bar{p}_n}{1-m} > \bar{p}_1 + \dots + \bar{p}_n = 1$$

$$\Rightarrow \exists j \leq n \text{ and } 0 \leq \varepsilon < \frac{1}{b_j} \text{ with } \sum_{i=1}^{j-1} \frac{1}{b_i} + \varepsilon = 1$$

$$\Rightarrow \left(\frac{1}{b_1}, \dots, \frac{1}{b_{j-1}}, \varepsilon, 0, \dots, 0 \right) \in P.$$

Formally, **Theorem 3** only states that at least one additional tuple of probabilities is included in P . However, it can be seen easily that in fact an infinite number of tuples can be constructed, meaning that P can be thought of as a polytope in n -dimensional space. Although this is not demanded formally by the theorem, it will be shown that obviously $\bar{p}$ is always included in P . As a set of probabilities, the probability estimation of the bookmaker fulfils the following characteristics: $\sum \bar{p}_i = 1$ and $0 \leq \bar{p}_i \leq 1$.

The positive margin implies: $m > 0 \Rightarrow \bar{p}_i = \frac{1}{b_i} (1 - m) < \frac{1}{b_i} \Rightarrow \bar{p} \in P$.

References

- Angelini, G., & De Angelis, L. (2019). Efficiency of online football betting markets. *International Journal of Forecasting*, 35(2), 712–721.
- Baker, R. D., & McHale, I. G. (2013). Forecasting exact scores in National Football League games. *International Journal of Forecasting*, 29(1), 122–130.
- Barnett, T., O'Shaughnessy, D., & Bedford, A. (2011). Predicting a tennis match in progress for sports multimedia. *OR Insight*, 24(3), 190–204.
- Boshnakov, G., Kharrat, T., & McHale, I. G. (2017). A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, 33(2), 458–466.
- Braun, S., & Kvasnicka, M. (2011). National sentiment and economic behavior. *Journal of Sports Economics*, 14(1), 45–64.
- Cattelan, M., Varin, C., & Firth, D. (2013). Dynamic Bradley–Terry modelling of sports tournaments. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 62(1), 135–150.
- Constantinou, A. C., Fenton, N. E., & Neil, M. (2012). pi-football: A Bayesian network model for forecasting association football match outcomes. *Knowledge-Based Systems*, 36, 322–339.
- Erikson, R. S., & Wlezien, C. (2012). Markets vs. polls as election predictors: An historical assessment. *Electoral Studies*, 31(3), 532–539.
- European Sport Security Association (2014). Sports betting: commercial and integrity issues. Retrieved from <http://www.eu-ssa.org/wp-content/uploads/Sports-Betting-Report-FINAL.pdf>.
- Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: the case of English football. *International Journal of Forecasting*, 21(3), 551–564.
- Forrest, D., & Simmons, R. (2008). Sentiment in the betting market on Spanish football. *Applied Economics*, 40(1), 119–126.
- Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21(2), 331–340.
- Goddard, J., & Asimakopoulou, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, 23(1), 51–66.
- Granger, C. W. J., & Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19(7), 537–560.
- Gray, P., & Gray, S. (1997). Testing market efficiency: evidence from the NFL sports betting market. *The Journal of Finance*, 52(4), 1725–1737.
- Hubáček, O., Šourek, G., & Železný, F. (2019). Exploiting sports-betting market using machine learning. *International Journal of Forecasting*, 35(2), 783–796.
- Koopman, S. J., & Lit, R. (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 178(1), 167–186.
- Koopman, S. J., & Lit, R. (2019). Forecasting football match results in national league competitions using score-driven time series models. *International Journal of Forecasting*, 35(2), 797–809.
- Kovalchik, S. A. (2016). Searching for the GOAT of tennis win prediction. *Journal of Quantitative Analysis in Sports*, 12(3), 127–138.
- Lasek, J., Szlávik, Z., & Bhulai, S. (2013). The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition*, 1(1), 27–46.
- Leitch, G., & Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *The American Economic Review*, 81, 580–590.
- Lessmann, S., Sung, M.-C., & Johnson, J. E. V. (2010). Alternative methods of predicting competitive events: An application in horserace betting markets. *International Journal of Forecasting*, 26(3), 518–536.
- McHale, I., & Morton, A. (2011). A Bradley–Terry type model for forecasting tennis match results. *International Journal of Forecasting*, 27(2), 619–630.
- Peeters, T. (2018). Testing the wisdom of crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, 34(1), 17–29.
- R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>.
- Rhode, P. W., & Strumpf, K. S. (2004). Historical presidential betting markets. *Journal of Economic Perspectives*, 18(2), 127–142.
- Snowberg, E., & Wolfers, J. (2010). Explaining the favorite-long shot bias: is it risk-love or misperceptions? *Journal of Political Economy*, 118(4), 723–746.
- Song, C., Boulier, B. L., & Stekler, H. O. (2007). The comparative accuracy of judgmental and model forecasts of American football games. *International Journal of Forecasting*, 23(3), 405–413.
- Spann, M., & Skiera, B. (2009). Sports forecasting: A comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, 28(1), 55–72.
- Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30(4), 934–943.
- Štrumbelj, E. (2016). A comment on the bias of probabilities derived from betting odds and their use in measuring outcome uncertainty. *Journal of Sports Economics*, 17(1), 12–26.
- Štrumbelj, E., & Vračar, P. (2012). Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. *International Journal of Forecasting*, 28(2), 532–542.
- Wolfers, J., & Leigh, A. (2002). Three tools for forecasting federal elections: lessons from 2001. *Australian Journal of Political Science*, 37(2), 223–240.
- Wolfers, J., & Zitzewitz, E. (2004). Experimental political betting markets and the 2004 election. *The Economists' Voice*, 1(2).
- Wunderlich, F., & Memmert, D. (2016). Analysis of the predictive qualities of betting odds and fifa world ranking: evidence from the 2006, 2010 and 2014 football world cups. *Journal of sports sciences*, 34(24), 2176–2184.
- Wunderlich, F., & Memmert, D. (2018). The betting odds rating system: using soccer forecasts to forecast soccer. *PLoS One*, 13(6), e0198668.