European Journal of Sport Science, 2021
Vol. 21, No. 7, 944–957, https://doi.org/10.1080/17461391.2020.1793002

Forecasting the outcomes of sports events: A review

FABIAN WUNDERLICH & DANIEL MEMMERT

Institute of Exercise Training and Sport Informatics, German Sport University Cologne, Cologne, Germany

Abstract
In the scientific community a large literature on sports forecasting exists, covering a wide range of different sports, methods and research questions. At the same time a lack of general literature such as reviews or meta-analyses on aspects of sports forecasting can be attested, partly attributable to characteristics of forecasting in sports that make it difficult to present through systematic approaches. The present study contributes to filling this gap by providing a narrative review about forecasting related to the outcomes of sports events. An overview about relevant topics in forecasting the outcomes of sports events is presented, a basic methodology is discussed and a categorization of methods is introduced. Having a specific focus on forecasting from ratings, we shed light on the difference between systematic and unsystematic effects influencing the outcomes of sports events. Finally an outlook on the expected impact of the increasing amount and complexity of available data on future sports forecasting research is presented. The present review can serve as a valuable starting point for researchers aiming at the investigation of sports-related forecasts, both helping to find appropriate methods and classify their work in the context of the state of research.

Keywords: *Prediction, modelling, quantitative study, analysis*

Highlights

- Approaches general aspects of sports forecasting research by reviewing forecasting literature related to the outcomes of sports events.
- Both systematic and unsystematic (i.e. random) effects influencing the outcomes of sports events are discussed and common approaches in the literature to mathematically model these effects are presented.
- The article introduces a categorization of topics related to sports forecasting and sources of information on which the forecasts are based.
- The fundamental role af ratings as an intermediate step in forecasting models is emphasized while issues in assesing the quality of forecasts based on ratings are discussed.

1. Introduction

Sport is a common field of application for forecasting methods that attracts the interest of both companies and researchers. Consequently, there is extensive literature covering various predictive aspects in sports. The present paper provides a narrative review of the literature concerning forecasts that are directly related to the actual outcome of a sports event. The distinction between systematic and unsystematic effects influencing the outcomes of sports events is outlined and discussed in light of the reviewed literature. The fundamental role of ratings in forecasting is emphasized while theoretical limits in forecasting from ratings are discussed.

Several reasons contribute to the fact that forecasts for outcomes of sports events have been investigated comprehensively by researches from various disciplines. The global sports betting industry is a billion-dollar business (see European Gaming & Betting Association, 2014) and as accurate forecasting is a fundamental aspect of sports betting, appropriate forecasting models (Baker & McHale, 2013; Goddard, 2005; McHale & Morton, 2011) are required by bookmakers to set the betting odds as well as by ambitious bettors in an attempt to win money by exploiting inaccurate odds. Although the authors may have an incentive not to disclose models that systematically outperform the betting

Correspondence: Fabian Wunderlich, Institute of Exercise Training and Sport Informatics , German Sport University Cologne, Am Sportpark Müngersdorf 6, Cologne 50933, Germany. E-mail: f.wunderlich@dshs-koeln.de

market, studies claiming positive betting returns are regularly published (Lessmann, Sung, & Johnson, 2010; Peeters, 2018).

In the economic context forecasting in relation to betting odds is extensively studied, yet with a slightly different theoretical background. The main objectives are to test the efficiency of betting markets (Goddard & Asimakopoulos, 2004; Gray & Gray, 1997; Štrumbelj & Šikonja, 2010; Woodland & Woodland, 1994), identify and analyze aspects of inefficiency such as the favorite-longshot bias (Ottaviani & Sørensen, 2008; Snowberg & Wolfers, 2010) or inefficiencies based on sentiments (Braun & Kvasnicka, 2011) and analyze different market structures such as bookmakers and bet exchanges (Franck, Verbeek, & Nüesch, 2010; Smith, Paton, & Williams, 2006).

Moreover, researchers have been attracted by comparing the predictive power of forecasts based on various different sources of information such as betting odds, quantitative models, ratings and rankings or human forecasts (Song, Boulier, & Stekler, 2007; Štrumbelj & Vračar, 2012; Wunderlich & Memmert, 2016). Research on human forecasts has focused on the heuristics used when performing forecasts (Pachur & Biele, 2007; Scheibehenne & Bröder, 2007) and the differences between experts and novices (Andersson, Memmert, & Popowicz, 2009).

In sports science, research related to forecasts can help to gain insights to the nature of the underlying sport (Štrumbelj & Vračar, 2012), the performance analysis of teams (Wunderlich & Memmert, 2018) or the efficiency of official rankings, ratings and seedings (Boulier & Stekler, 1999; Lasek, Szlávik, & Bhulai, 2013; McHale & Davies, 2007).

Driven by the predictive power of betting odds, researchers have started to base forecasts on further sources of data in an attempt to profit from collaborative knowledge. These sources include prediction markets (Luckner, Schröder, & Slamka, 2008; Spann & Skiera, 2009), a website containing community-based market value estimations (Peeters, 2018) or data from the social media platform Twitter (Brown, Rambaccussing, Reade, & Rossi, 2017).

The vast majority of research focuses on a specific aspect of sports forecasting and although a lot of research has been done in this field, hardly any paper deals with sports forecasting in general (for an exception see Stekler, Sendor, & Verlander, 2010), its characteristics, methodology and categorization. The present paper contributes to filling this gap by evaluating and classifying general aspects of forecasting related to the outcomes of sports events.

It appears that the quality of a review is often judged by the degree of objectivity and the rigour of the systematic approach. However, highly systematic approaches do not come without disadvantages that especially apply to sports forecasting. A meta-analysis is intended to control for the degree of error in various studies concerning the same research question and aims at estimating the unknown true effect. While this is particularly necessary for experimental studies, in sports forecasting large real-world datasets are available reducing errors by randomness and giving results a relatively high power. At the same time datasets are highly heterogeneous making it difficult to compare results. For example, it would be unreasonable to compare forecasting models by resorting to the fraction of correctly predicted matches for different competitions or even across sports. Moreover, forecasting can rather be considered as data science than experimental research where no presence or absence of an effect, but the quality of a model is tested. Systematic reviews (besides undisputed advantages in objectivity, transparency, completeness and replicability) also possess weaknesses in general and with regard to sports forecasting. As already outlined above, sports forecasting methods are not always the main subject of investigation, but often accompanied by or even hidden by other motives (such as testing market efficiency, understanding human judgements, testing the quality of rankings or understanding the underlying processes of a sports game). Thus, an article search (even if systematic) must necessarily be based on subjective decisions and domain-specific knowledge by including or not including keywords like "market efficiency", "betting odds", "rankings", "human forecasts", "prediction markets", etc. which will have major impact on the literature found. A systematic search also conceals practical issues like solely using the term "sports forecasting" excluding studies strictly using sports- or league-specific wording like "football forecasting" and "forecasting NBA matches" or subjectively including a limited number of preselected sports as keywords. Another aspect is the trade-off between narrowing the scope omitting relevant aspects or including an unreasonable number of (potentially monothematic) studies. The lack of reviews and meta-analyses thus does not come without reasons as the use of systematic approaches is not unproblematic in this research field. This, however, is a dilemma as not reviewing the state of research at all will deprive researchers of the opportunity to find and make use of introductory overviews of the literature.

Driven by the above reasons, a narrative review style was chosen and an article search was performed using the Web of Science Core Collection and Google Scholar as two main starting points, but not aiming for completeness or full objectivity. Only articles related to forecasts of the actual outcome of a sports event were included which excludes forecasts of further

Table I. Topics and related research questions in forecasting the outcomes of sports events

Topic	Research question	References
Forecasting methods	Presenting and analyzing forecasting methods and models	See Table III
Market efficiency	Investigating the efficiency of sports betting markets	Dixon and Coles (1997) Goddard and Asimakopoulos (2004) Gray and Gray (1997) Leitner, Zeileis, and Hornik (2011) Štrumbelj and Šikonja (2010)
	Presenting evidence and explanations for the favorite-longshot bias	Ottaviani and Sørensen (2008) Snowberg and Wolfers (2010) Woodland and Woodland (1994)
	Presenting evidence and explanations for sentiments in betting markets	Braun and Kvasnicka (2011)
	Comparing different market structures (such as bookmaker and bet exchange)	Franck et al. (2010) Smith et al. (2006)
Comparing forecasts	Comparing forecasts from different sources of information such as betting odds, quantitative models, ratings/rankings or human forecasts	Forrest, Goddard, and Simmons (2005) Goddard (2005) Hvattum and Arntzen (2010) Kovalchik (2016) Lasek et al. (2013) Leitner, Zeileis, and Hornik (2010) Song et al. (2007) Spann and Skiera (2009) Wunderlich and Memmert (2016)
Human forecasts	Explaining heuristics used by humans when forecasting	Pachur and Biele (2007) Scheibehenne and Bröder (2007)
	Demonstrating differences between experts and novices in forecasting tasks	Andersson et al. (2009)
Sports science	Using forecasting methods to gain insights to the nature of the underlying sport	Heuer and Rubner (2009) Hill (1974) Štrumbelj and Vračar (2012)
	Using forecasting methods for performance analysis of teams	Wunderlich and Memmert (2018)
	Investigating the (predictive) value of official rankings, ratings and seedings	Barrow, Drayer, Elliott, Gaut, and Osting (2013) Boulier and Stekler (1999) Dyte and Clarke (2000) McHale and Davies (2007)
Sources of data	Analyzing forecasts based on prediction markets	Luckner et al. (2008)
	Analyzing forecasts based on a market value estimation website	Peeters (2018)
	Analyzing forecasts based on data from social media platform Twitter	Brown et al. (2017)

aspects related to sports, but not directly related to the outcome. These aspects include, but are not limited to forecasting the success of individual athletes in team sports, forecasting sports injuries, implications of predictive ratings in computer games also referred to as eSports, the inclusion of predictive models in detecting match-fixing or forecasting measures for consumption of professional sports such as stadium attendance, ticket sales or TV ratings. No limit in terms of publication years was set; however, due to the time of the search no literature after 2018 is included. Despite the narrative review style, a systematic overview on those articles reviewed is given in section 4.

Concerning terminology, both the terms *forecasting* and *predicting* are commonly used in the literature reviewed. In general, *predicting* is implicitly more associated with subjective methods, the choice of a specific outcome and in-sample valuation while *forecasting* is implicitly more associated with scientific methods, probabilistic estimation and out-of-sample methods. Although this might have played a role in the choice of the authors, none of them discusses the difference in terminology or their choice of wording and some even use *predicting* and *forecasting* as fully interchangeable. In consistency with the literature, we will not strictly distinguish between both terms, consistently using the slightly more common term *forecasting*, while maintaining established terms like *prediction markets* or terms not having a common equivalent with regard to forecasting such as *predictive*, *(un)predictable* or *predictability*.

Table I summarizes topics in forecasting sports outcomes and a selection of related research questions. Additionally, the references included in this review are assigned to the research questions as will be further explained in section 4.

2. Methodology and categorization

Besides choosing the appropriate method of forecasting, at least three issues need to be tackled when forecasting sports events: (a) The subject of forecasting, i.e. the aspect of sport that is supposed to be forecasted. (b) The source of information, i.e. the data source the forecast leans on. (c) The measure of predictive quality in order to determine the performance of the forecasting method. In the subsequent sections, these issues are evaluated and a categorization is introduced.

2.1. Subject of forecasting

Several aspects shall be mentioned that explain the diversity of forecasting tasks related to outcomes of sports events. First of all, different sports with completely different characteristics are studied. Frequently investigated sports include US sports such as American football, basketball and baseball; horse racing; soccer and tennis (cf. Stekler et al., 2010). We assume public interest, financial background, popularity in sports betting and data availability among the reasons why researchers choose these sports. Within the same sport, forecasting can focus on various forms of competitions with different structures such as national leagues (Goddard, 2005), international club competitions (Leitner et al., 2011) or competitions including national teams (Dyte & Clarke, 2000). Another aspect refers to different points of time and forecasting horizons: Forecasting a series of matches or the whole season using pre-season data generates different challenges in terms of data availability and complexity compared to forecasts of the subsequent match using within-season data. While short-term forecasts usually refer to single matches or races, longer time horizons also include forecasts for the outcome of major events such as European championships (Leitner et al., 2010) or Olympic Games (Forrest, Sanz, & Tena, 2010). Furthermore, the forecasting task can differ in the level of detail, i.e. only picking the presumable winner of a match (Luckner et al., 2008; Wigness, Williams, & Rowell, 2010) or calculating probabilities for various final scores of a match (Baker & McHale, 2013; Karlis & Ntzoufras, 2003) or various outcomes of a tournament (Clarke & Dyte, 2000).

2.2. Sources of information

A crucial step to categorize sports forecasting methods relates to the source of information used to obtain forecasts. Stekler et al. (2010) state three sources (namely *betting market*, *statistical models* and *experts*), though not really aiming at a classification. We roughly base our categorization on these three sources, but use finer distinctions and divide the statistical models into two subcategories.

Forecasts are divided into two main categories, namely human judgement and quantitative models and four subcategories. Human judgement includes all forecasts that are exclusively or predominantly driven by human decisions. The first subcategory covers individual human judgement, e.g. forecasts made by single persons, for example in an experimental environment (experts, novices) or published in the media (tipsters). The second subcategory of human judgement covers sources of collaborative human judgement where the forecast arises from an interaction between various persons. This is usually associated with the betting market, but also includes prediction markets, community based websites, or social media. Forecasts based on human judgement are usually characterized by a limited transparency as it is not apparent how the forecast is made and on which criteria the judgement is based.

Quantitative models include all forecasts exclusively or predominantly based on mathematical models or statistical methods and are divided into two subcategories. Forecasting based on external ratings/rankings refers to forecasting from ratings/rankings that are not part of the model itself and thus were not explicitly designed for this purpose. This, for example, covers the FIFA world ranking and the official ATP ranking. The other subcategory includes internal ratings/rankings that are part of the model itself and thus were explicitly designed for forecasting purposes like a single strength or multiple strength parameter of a team or player. Moreover, it includes models not based on ratings/rankings covering methods based on data that cannot be interpreted as a rating. Quantitative models are characterized by full transparency as the forecasting process is known and it is clear which criteria are applied to derive the forecast. The main advantage in (collaborative) human judgement can be seen in processing all available information and processing qualitative factors while quantitative models are limited by the availability of data and its usability. In turn, the advantage of quantitative models can be seen in not being subject to human misjudgement or biases except

Table II. Sources of information in forecasting the outcomes of sports events

Source of information	Detailed source of information	Example	Example reference
Human judgement	Individual human judgement	Tipsters	Spann and Skiera (2009)
		Experts	Andersson et al. (2009)
		Novices	Andersson et al. (2009)
	Collaborative human judgement	Betting markets	Forrest et al. (2005)
		Prediction markets	Luckner et al. (2008)
		Community based estimation of market value	Peeters (2018)
		Twitter	Brown et al. (2017)
Quantitative models	Based on external ratings/rankings	FIFA world ranking	Lasek et al. (2013)
		ATP ranking	Clarke and Dyte (2000)
		College basketball seedings	Boulier and Stekler (1999)
	Based on internal ratings/rankings or non-ratings/rankings based	Team strength	Manner (2016)
		Player strength	Newton and Aslam (2009)
		Various explanatory variables	Goddard and Asimakopoulos (2004)

in designing the model. Table II shows the categorization as well as examples and example references.

The table is limited to one exemplary reference that best covers the respective example presented. Although carefully developed, such a categorization cannot possibly make a sharp distinction. Obviously, humans can consciously or unconsciously make use of references to quantitative models when taking a decision (as especially bookmakers and professional gamblers will do). Moreover, quantitative models can be subject to human (mis)judgement when being designed. Approaches mixing various sources of information are also possible such as integrating human judgement into quantitative models (Forrest et al., 2010), integrating betting odds into quantitative models (Baker & McHale, 2013) or deriving ratings from betting odds (Leitner et al., 2010; Wunderlich & Memmert, 2018).

2.3. *Measures of predictive quality*

Regardless of the individual research question, the usage of an appropriate measure of predictive quality is an integral part of sports forecasting. A straightforward strategy is to simply count the number of "correct" and "incorrect" forecasts. If a binary forecast is made, i.e. the winner of a match is forecasted, this simplistic measure is sufficient and widely used (Song et al., 2007; Wigness et al., 2010). For evaluating more detailed probabilistic forecasts this measure is not satisfactory.

The most common way to measure predictive quality of probabilistic forecasts are squared forecasting errors (Lasek et al., 2013) which are denoted with different names such as Brier score (Cattelan, Varin, & Firth, 2013) and quadratic loss function (Štrumbelj & Vračar, 2012) or slightly adjusted versions such as the rank probability score (Cattelan et al., 2013). Other loss functions like the informational loss (Hvattum & Arntzen, 2010), the average of forecasted probabilities (McHale & Morton, 2011), the likelihood-function (Koopman & Lit, 2015) or pseudo likelihood (Rue & Salvesen, 2000) are further probabilistic measures. In general, all of these measures are intuitive and based on the principle of comparing forecasted probabilities to actual outcomes when backtesting the forecasting method using recent outcomes.

In some situations the Spearman correlation can serve as a useful measure for predictive quality, for example when comparing the forecasted outcome of a tournament (Leitner et al., 2010) or a medal table (Forrest et al., 2010) with the actual outcome.

Another common approach is to verify the economic value of a forecasting method by evaluating its financial benefit when using it as a basis for a betting strategy (Dixon & Coles, 1997; Franck et al., 2010). The forecast is used to construct a betting strategy, then the financial payoff for the strategy is calculated and used as a measure of predictive quality. One of the best-known methods in determining the stakes for profitable bets originates from the work of Kelly (Kelly, 1956), is referred to as *Kelly criterion* or *Kelly strategy* and frequently used when investigating betting strategies (Baker & McHale, 2013; Lessmann et al., 2010). Betting strategies are particularly suitable if the focus is on investigating the efficiency of betting markets (Goddard & Asimakopoulos, 2004; Gray & Gray, 1997; Woodland & Woodland, 1994).

As there seems to be no consensus on a most appropriate measure, the standard approach is to consider various of the measures concurrently (Koopman & Lit, 2015; McHale & Morton, 2011).

3. The role of ratings in forecasting

This section is focused on the role of ratings as an intermediate step in forecasting, therefore concentrating on quantitative models aiming at a probabilistic forecast and making use of ratings. The intention is to discuss the differences of systematic and unsystematic effects concerning the outcomes of sports events. No differences between parametric and non-parametric approaches or between in-sample and out-of-sample validation are analyzed in detail. The same applies to the granularity of methods, i.e. models being or not being able to determine within-match forecasts.

3.1. Definition of ratings

A large proportion of quantitative models explicitly or implicitly use two separate steps in the forecasting process: First, the ratings for all participants are determined and then the probabilities for the outcomes are derived from the given ratings. Regarding the definition of a rating (and ranking) we refer to Barrow et al. (2013):

> Let V be a set of n teams to be rated, which we enumerate $V = \{i\}_{i=1}^{n}$. A rating $\phi{:}A \to \mathbb{R}$ assigns each team a quantitative strength. A ranking is an ordering of the teams; a rank α team is stronger than n-α-other teams. A ranking may be obtained from a rating on V simply by sorting.

Obviously the same definition is applicable to the rating of any participant in a sports event (players, horses, etc.) instead of teams. As each rating implies a ranking, the expressions are not distinguished carefully and in practice the word ranking is sometimes used, even if the ranking as such is a rating as well (e.g. the FIFA world ranking in soccer or the ATP ranking in tennis). Within this chapter we will concentrate on forecasting from ratings rather than forecasting from rankings and explicitly allow the usage of various ratings for one participant such as separate attack and defence parameters or serve and return strengths.

3.2. Systematic and unsystematic effects in forecasting

Ratings are a way to express participant-specific effects that are relevant for forecasting and it seems reasonable to shed more light on the distinction of different effects in a forecasting process. There appears to be an implicit consensus in sports forecasting that outcomes of sports competitions are assessed as partly defined by systematic effects and partly by unsystematic (i.e. random) effects. Systematic effects include participant-specific factors such as the quality of a team or player (that can be expressed as a rating) and global factors that systematically influence the outcomes such as the home advantage. Unsystematic effects are factors that are neither participant-specific nor explainable by any global factor and thus can be described as random processes.

The degree of systematic and unsystematic effects reflected in the outcomes of a competition determines its degree of predictability. Determining a winner by tossing a coin is fully defined by unsystematic effects and thus unpredictable. Determining a winner by picking the oldest person to win (in knowledge of all dates of birth) is fully defined by systematic effects and thus predictable. Sports competitions are considered to be located somewhere in between these two extreme examples. Figure 1 illustrates the influence of systematic and unsystematic effects on the outcome of sports events.

The distinction between systematic and unsystematic effects concerns probabilistic forecasts. If the forecast is just a pick of the presumable winner, it is sufficient to model systematic effects and unsystematic effects are neglected. There exists little research specifically aiming at a distinction between systematic and unsystematic effects. Ben-Naim, Vazquez, and Redner (2006) have focused on evaluating the predictability of various sports and have presented a method to quantify the degree of predictability in American football, ice hockey, basketball, baseball and soccer. In the domain of soccer there is research explicitly stating that results are influenced partly by skill and partly by chance (Hill, 1974). A standard approach in soccer forecasting is therefore to model systematic effects in team specific strength parameters and outcomes by using a probability model while taking the home advantage into account, thus being completely in line with the above classification. A widely accepted method based on team specific attack and defence parameters and a Poisson model is often attributed to the work of Maher (Maher, 1982) and still commonly used (Koopman & Lit, 2015).

In the sports forecasting literature several methods have been used to reflect systematic participant-specific effects. Official ratings have been evaluated in soccer (Lasek et al., 2013) and tennis (Clarke & Dyte, 2000) as well as official seedings in College basketball (Boulier & Stekler, 1999). A straightforward method is to make use of previous results, which has been done in the form of moving weighted averages of matches won in basketball (Cattelan et al., 2013) or the average goal difference in soccer (Heuer & Rubner, 2009). A common approach is to define and model the quality of a team as a strength parameters, as has been done in basketball (Manner, 2016) and American football (Glickman & Stern,

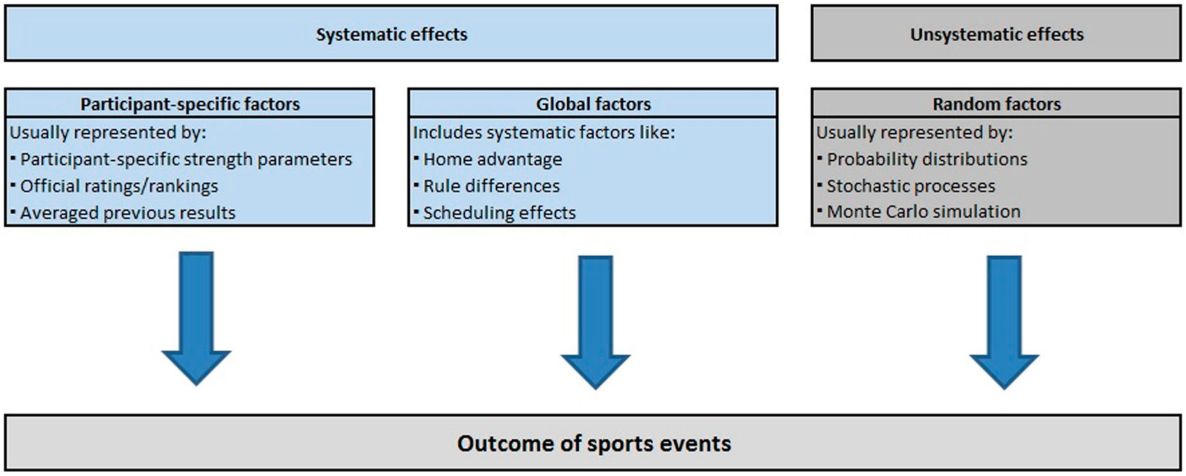

Figure 1. The influence of systematic and unsystematic effects on the outcome of sports events

2017). In some sports it is more common to introduce various strengths parameters, such as separate serve and return strengths in tennis (Newton & Aslam, 2009) or attack and defence strengths in soccer as described above. Moreover, variations of the ELO rating system, that was originally invented for chess (Glickman & Jones, 1999), have been applied to sports such as tennis (Kovalchik, 2016) and soccer (Hvattum & Arntzen, 2010).

The most relevant factor usually being modelled as a systematic global factor is the home advantage (see Nevill & Holder, 1999 for an overview). Further factors being modelled as systematic, but non-participant-specific are scheduling effects like back-to-back matches in basketball (cf. Manner, 2016), rule based differences like the number of winning sets in different tennis tournaments (cf. Clarke & Dyte, 2000) or motivational effects such as the significance of a match in a competition (cf. Goddard & Asimakopoulos, 2004).

The most common methods to assess unsystematic (i.e. random) effects are logit/probit regression models, probability models, stochastic processes and Monte Carlo simulation. Regression models, such as logit regression based on a rating as single covariate (Hvattum & Arntzen, 2010) and probit regression based on a variety of covariates (Goddard, 2005) have been used in soccer. Likewise, in tennis various regression models have been proposed (see Kovalchik, 2016 for an overview). Researchers have made use of probability models related to the simple Bradley–Terry model in tennis (McHale & Morton, 2011), soccer and basketball (Cattelan et al., 2013). Models based on the Poisson distribution are a common approach to model exact scores in soccer (Heuer, Müller, & Rubner, 2010; Karlis & Ntzoufras, 2003). Stochastic processes like birth processes have been proposed in American football (Baker & McHale, 2013) and soccer (Dixon & Robinson, 1998). In horse racing an adaptation of random forest classifiers has been used to forecast race outcomes (Lessmann et al., 2010). Monte Carlo simulation of single matches has been used in basketball (Štrumbelj & Vračar, 2012) and tennis (Newton & Aslam, 2009). Moreover full tournaments were simulated in tennis (Clarke & Dyte, 2000) and soccer (Leitner et al., 2010). In light of the reviewed literature a wide implicit agreement to the above classification of systematic and unsystematic effects can be attested.

3.3. *Limits in forecasting from ratings*

Assessing the predictive quality when forecasting from ratings can be difficult. An absolute measure of predictive quality only provides limited insights as the characteristics of a sport can hamper or support its predictability. While forecasting 70% of the winners correctly might be a good performance in one sport, it could be a poor performance in another sport or competition. Researchers therefore aim at finding the best possible method, validating it in comparison to other forecasting methods applied to the same data set, yet knowing that a perfect forecast remains unattainable. So even if finding a superior sports forecasting method, it remains unclear whether incomplete data, inaccurate methods or a limited predictability of the sport itself prevents an even more accurate forecast.

In forecasting from ratings three possible factors limiting the predictive quality can be identified, as illustrated in Figure 2.

Provided that the outcomes of sports events are influenced by truly random processes, the predictability is limited and there is a given (but unknown) threshold that will not be exceeded even in full

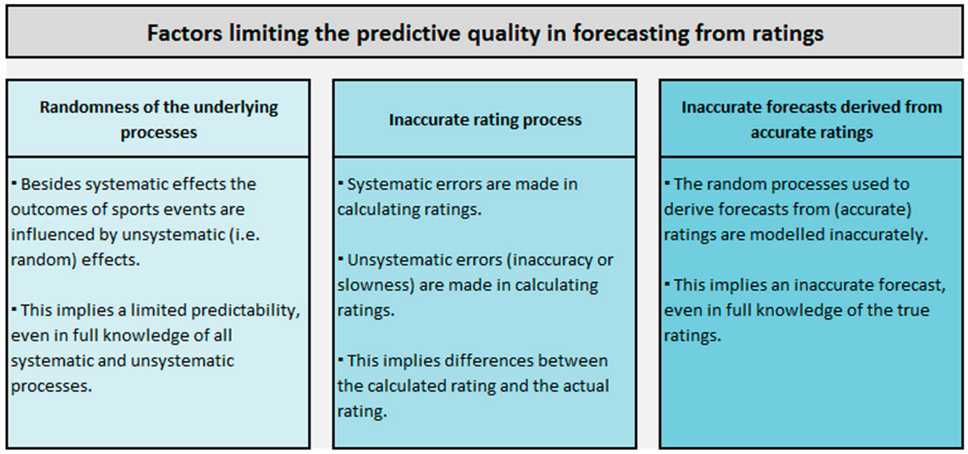

Figure 2. Factors limiting the predictive quality in forecasting from ratings

knowledge of all systematic and random processes. Two effects can be responsible for not reaching this threshold: Inaccurate estimation of ratings or inaccurate forecasting from ratings (occurring even in full knowledge of accurate ratings). This theoretical consideration illustrates a difficulty in evaluating the predictive qualities of a forecasting method. In contrast to theoretical set-ups the predictability of a sport is not known and cannot be observed exactly. It is also not possible to observe what the "actual" ratings are and how far they differ from the calculated ratings. While this difficulty is unavoidable in practical applications, theoretical approaches may help to gain a better understanding. In theoretical set-ups using simulated test data, both the actual ratings and the random processes (i.e. predictability) would be predefined and fully known. This approach could help to understand the weaknesses of rating processes and contribute to improvements in forecasting.

4. Systematic overview

Though this review does not pursue a strictly systematic approach, a structured overview of the reviewed literature is presented by systematically evaluating and categorizing all references in line with the structure of this review.

As a first step, references were classified based on the list of topics and research questions explained in section 1, the results of this classification have been included in Table I. Nine out of a total of sixty references mentioned in this review are either presenting an overview themselves (Stekler et al., 2010) or are not focused on forecasting and thus were excluded from the table. It can be argued that numerous references relate to more than one topic as for example references focused on testing market efficiency also need to use and present a sound forecasting model. Therefore, references were assigned to the topic best fitting the focus of the article and despite thorough examination by the authors this unavoidably implies a certain degree of subjectivity.

As a second step, all references focused on "Presenting and analyzing forecasting methods and models" were investigated in more detail and presented in Table III. For reasons of clarity and extent, we refrained from presenting a detailed analysis of all sixty references. The following information are included in Table III: Column *Subject of forecasting* illustrates the sport(s), the competition(s), the time interval of data (e.g. which seasons or tournaments) and the level of detail (e.g. match results or exact scores) in line with section 2.1. To avoid redundant data, the sources of information as in section 2.2 and Table II are not presented in an extra column as all references presented can be characterized as quantitative models and the detailed source of information (like external ratings/rankings) can be derived from the systematic effects. Column *Measures of predictive quality* summarizes the use of probabilistic measures or betting strategies as elaborated on in section 2.3. The modelling of *Systematic effects* and *Unsystematic effects* as discussed in section 3.2 and illustrated in Figure 1 are shown in the respective columns. Furthermore *Main goal* and *Main results* of each reference are presented in the respective columns. We refrain from presenting absolute numbers of forecasting accuracy as we consider a direct comparison across different sports, competitions and time intervals unreasonable.

5. Conclusion and outlook

This narrative review can contribute to a more comprehensive approach to sports forecasting literature. We would like to encourage evaluation of further cross-thematic aspects such as a bibliometric analysis of sports

Table III. Systematic overview of the references focused on presenting and analyzing forecasting methods and models

Reference	Main goal	Subject of forecasting	Measures of predictive quality	Systematic effects	Unsystematic effects	Main results
Baker and McHale (2013)	Developing a model to forecast exact scores in a sport with more than one type of score	American football (NFL, 8 seasons 2001–2008, exact scores)	Log-likelihood, Brier score, Mean absolute deviation of points, Percentage of correct forecasts, Kelly betting strategy	Several functions ("hazards" of scoring) for different types of scores in varying situations	Birth process model	The proposed model is marginally outperformed by the betting market with regard to match results, comparable to the betting market with regard to exact scores and slightly outperforms the betting market when incorporating betting odds into the model.
Cattelan et al. (2013)	Developing a forecasting model with team strengths dynamically changing over time	Basketball (NBA, season 2009/10, match results); Soccer (Serie A, season 2008/09, match results)	Brier score, Rank probability score	Home and away strengths, Home advantage	Basketball: Bradley-Terry model, Soccer: Extended Bradley-Terry model including draws	The predictive power of the proposed dynamic model is comparable to a static ability benchmark model, but superior to forecasts based solely on frequency of wins.
Clarke and Dyte (2000)	Modelling and simulating the results of tennis matches based on official ratings	Tennis (ATP, 3003 matches from 1997 and three Grand Slam tournaments from 1998/99, set winners, match results, tournament outcomes)	Comparison between expected and observed results	Difference in ATP rating points, Rule effects (number of sets needed to win a match)	Logistic regression model (with regard to set winning probability), Simulation (with regard to tournament outcomes)	Wins in straight sets were underestimated by the proposed model while close matches were overestimated.
Dixon and Robinson (1998)	Developing a method to model goals in soccer dependent on scoreline and matchtime	Soccer (four English divisions and cup matches, seasons 1993–1996, exact scores)	Comparison between simulated and observed results	Attack and defence strengths, Home advantage, Scoreline, Matchtime	Birth process model (including simulation)	In the context of the proposed model matchtime and scoreline have effects on scoring probabilities.
Forrest et al. (2010)	Improving an existing model to forecast medals at Olympic Games	Olympic Games (national team medals, 1992–2008)	Mean absolute error of medals, Spearman correlation between forecasted and observed results	Various explanatory variables (including recent performance and country-specific information related to politics and economy)	Tobit regression model	Forecasting performance could be improved by adding additional covariates to an existing model.
Glickman and Jones (1999)	Analyzing the ability of the ELO rating system in estimating strengths of chess players and obtaining match forecasts	Chess (United States Chess Federation, >220,000 tournament matches between Jan and Oct 1997, match winner)	Comparison between simulated and observed results	Player strengths	Logistic function	The ELO model used by USCF systematically overestimates the winning probability of the better rated player.

Glickman and Stern (2017)	Presenting methods to measure team strength and to forecast match outcomes in American football	American football (NFL, seasons 2006/07–2014/15, match winner and exact scores)	No forecast validated	Various methods are presented (mainly focused on team strengths including match location)	Various methods, e.g. Thurstone-Mosteller model and Bradley-Terry model (for match winner) and normal distribution (for exact scores)	Being focused on presenting various methods, rating procedures were demonstrated on a real-world data set, but no forecasts were analyzed.
Heuer et al. (2010)	Understanding and modelling the outcomes of soccer matches	Soccer (German Bundesliga, seasons 1987/88–1990/91 and 1992/93–2007/08, exact scores)	Comparison between theoretical and observed distribution of results	Average goal difference as a measure of team strength, Home advantage	Independent Poisson distributions	Except for draw effects, a very good fit of data can be achieved by independent Poisson distributions using an adequate distribution of strengths. It is demonstrated that random effects have a higher effect on results than team strength and that in-season or in-match fluctuation of team strengths is neglectable.
Karlis and Ntzoufras (2003)	Modelling the outcomes of soccer matches with a special focus on the number of draws	Soccer (Serie A, season 1991/92; Champions League, season 2000/01), Water-Polo (European Championships 1999)	Log-likelihood, Further likelihood based measures	Attack and defence strengths, Home advantage	Diagonal inflated bivariate Poisson model	A better fit of data can be achieved by including correlation and inflation of the draw probability into Poisson models.
Koopman and Lit (2015)	Developing a model to forecast soccer matches using dynamic team strengths	Soccer (Premier League, seasons 2003/04–2011/12, exact scores)	Log-likelihood, Sum of squared errors, Betting strategy	Attack and defence strengths, Home advantage	Bivariate Poisson model	A profitable betting strategy could be deduced from the proposed forecasting model.
Lessmann et al. (2010)	Making use of machine learning techniques to forecast horse racing	Horse racing (Hong Kong racetracks, 1000 races in 2005 and 2006, race winner)	Kelly betting strategy	A total of 41 explanatory variables (including general as well as horse- and jockey-specific information)	Random forest classifier	The proposed model was able to generate significant betting returns and to outperform conditional logit models.
Maher (1982)	Modelling the exact scores of soccer matches	Soccer (four English divisions, seasons 1973–1975, exact scores)	Comparison between simulated and observed results	Attack and defence strengths, Home advantage	Poisson model, Bivariate Poisson model	The proposed model shows a reasonably good fit to the data.
Manner (2016)	Developing a forecasting model for basketball, allowing for heteroscedasticity (team-specific volatility of performance) and dynamic team strengths	Basketball (NBA, seasons 2006/07–2013/14, score differences)	Mean squared error Mean absolute error (both with regard to score difference), Percentage of correct forecasts	Team strengths, Home advantage, Scheduling effects (back-to-back games)	Regression model	The proposed model shows weak evidence for heteroscedascity, but little evidence for time-varying strengths over the course of a season. Model forecasts are outperformed by the betting market, while combined forecasts are comparable to the market.

(*Continued*)

Table III. Continued.

Reference	Main goal	Subject of forecasting	Measures of predictive quality	Systematic effects	Unsystematic effects	Main results
McHale and Morton (2011)	Developing a model to forecast tennis matches based on previous results and surface	Tennis (ATP, seasons 2000–2008, match winner)	Log-likelihood, Average of forecasted probabilities, Betting strategies, Brier score, Percentage of correct forecasts	Player strengths, Surface	Bradley-Terry type model	The proposed model outperforms forecasts from two benchmark logistic regression models.
Newton and Aslam (2009)	Modelling tennis results including player consistency (i.e. random variation of strengths)	Tennis (ATP, season 2007, match winner, tournament outcomes)	No forecast validated	Serve and return strengths	Markov chain model using simulation for match and tournament results	The model's ability to simulate matches and tournament progress has been demonstrated.
Rue and Salvesen (2000)	Developing a model to forecast exact scores in soccer	Soccer (Premier League and Division 1, season 1997/98, exact results)	Pseudo likelihood, Betting strategy	Attack and defence strengths, Home advantage, Psychological effect	Markov chain model using simulation	The proposed model shows good results in comparison to betting odds and achieves positive betting returns in single bets.
Wigness et al. (2010)	Developing an improved ranking system in college football by using iterative methods	College football (NCAA, seasons 1998–2008, match winner)	Percentage of correct forecasts	Team rankings	None, as no percentage forecasts are made	The proposed ranking model outperforms BCS computer rankings in terms of predictive quality.

forecasting literature focusing on developments in sports forecasting over time or a network analysis of keywords related to sports forecasting. Moreover, we propose to put more effort in investigating theoretical frameworks of forecasting in the future. This could help to learn more about the question whether inaccurate forecasts are evoked by a limited predictability, by rating inaccurately or by forecasting incorrectly from correct ratings as discussed in section 3.3. In recent years, sports forecasting has taken advantage of the technological progress and especially of internet based data. The growth of online betting, simplified implementation of prediction markets, availability of community-based websites and social media data provide opportunities for research on sports forecasts that would not have been possible a few decades ago. Currently new sources of data have become available and will become available in near future (Memmert & Raabe, 2018). This includes match related data such as position data in soccer (Low et al., 2019; Memmert, Lemmink, & Sampaio, 2017) and basketball (Lucey, Bialkowski, Carr, Yue, & Matthews, 2014) or ball tracking data in tennis (Wei, Lucey, Morgan, & Sridharan, 2013). Moreover online data such as social media interaction on sports or sports-related search engine queries might be investigated more intensively in relation to sports forecasting. These data are characterized by high volume, limited veracity (position data and ball tracking data) as well as high velocity and variety including unstructured textual data and thus are increasingly associated with the field of Big Data (see Mauro, Greco, & Grimaldi, 2015 for a definition). We assume the growing data availability and complexity to have a significant impact and cause substantial changes in sports forecasting. Extensive and complex data analysis is bringing new technological challenges to researchers that will need to be able to raise, store and process large amounts of data. Both technological challenges (e.g. appropriate hardware, software and database systems) and methodological challenges (e.g. a stronger focus on artificial intelligence and machine learning) can be expected. In conclusion, it can be anticipated that the scientific community will preserve its curiosity in research on forecasting the outcomes of sports events.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Fabian Wunderlich http://orcid.org/0000-0002-7445-6858
Daniel Memmert http://orcid.org/0000-0002-3406-9175

References

Andersson, P., Memmert, D., & Popowicz, E. (2009). Forecasting outcomes of the World Cup 2006 in football: Performance and confidence of bettors and laypeople. *Psychology of Sport and Exercise*, *10*(1), 116–123. doi:10.1016/j.psychsport.2008.07.008

Baker, R. D., & McHale, I. G. (2013). Forecasting exact scores in National Football League games. *International Journal of Forecasting*, *29*(1), 122–130. doi:10.1016/j.ijforecast.2012.07.002

Barrow, D., Drayer, I., Elliott, P., Gaut, G., & Osting, B. (2013). Ranking rankings: An empirical comparison of the predictive power of sports ranking methods. *Journal of Quantitative Analysis in Sports*, *9*(2). doi:10.1515/jqas-2013-0013

Ben-Naim, E., Vazquez, F., & Redner, S. (2006). Parity and predictability of competitions. *Journal of Quantitative Analysis in Sports*, *2*(4). doi:10.2202/1559-0410.1034

Boulier, B. L., & Stekler, H. O. (1999). Are sports seedings good predictors? An evaluation. *International Journal of Forecasting*, *15*(1), 83–91. doi:10.1016/S0169-2070(98)00067-3

Braun, S., & Kvasnicka, M. (2011). National sentiment and economic behavior. *Journal of Sports Economics*, *14*(1), 45–64. doi:10.1177/1527002511414718

Brown, A., Rambaccussing, D., Reade, J. J., & Rossi, G. (2017). Forecasting with social media: Evidence from tweets on soccer matches. *Economic Inquiry*, *20*(3), 1363. doi:10.1111/ecin.12506

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. doi:10.1111/j.1467-9876.2012.01046.x

Clarke, S., & Dyte, D. (2000). Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research*, *7*(6), 585–594. doi:10.1016/S0969-6016(00)00036-8

Dixon, M. J., & Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *46*(2), 265–280.

Dixon, M. J., & Robinson, M. E. (1998). A birth process model for association football matches. *The Statistician*, *47*(3), 523–538.

Dyte, D., & Clarke, S. R. (2000). A ratings based Poisson model for World Cup soccer simulation. *Journal of the Operational Research Society*, *51*(8), 993–998. doi:10.1057/palgrave.jors.2600997

European Gaming & Betting Association. (2014). *Sports betting: Commercial and integrity issues*. Retrieved from http://www.egba.eu/media/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. doi:10.1016/j.ijforecast.2005.03.003

Forrest, D., Sanz, I., & Tena, J. D. (2010). Forecasting national team medal totals at the Summer Olympic Games. *International Journal of Forecasting*, *26*(3), 576–588. doi:10.1016/j.ijforecast.2009.12.007

Franck, E., Verbeek, E., & Nüesch, S. (2010). Prediction accuracy of different market structures — bookmakers versus a betting exchange. *International Journal of Forecasting*, *26*(3), 448–459. doi:10.1016/j.ijforecast.2010.01.004

Glickman, M. E., & Jones, A. C. (1999). Rating the chess rating system. *Chance*, (12), 21–28.

Glickman, M. E., & Stern, H. S. (2017). Estimating team strength in the NFL. *Handbook of Statistical Methods and Analyses in Sports*, 113–136.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of*

Forecasting, *21*(2), 331–340. doi:10.1016/j.ijforecast.2004.08.002

Goddard, J., & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, *23*(1), 51–66. doi:10.1002/for.877

Gray, P. K., & Gray, S. F. (1997). Testing market efficiency: Evidence from The NFL sports betting market. *The Journal of Finance*, *52*(4), 1725–1737. doi:10.1111/j.1540-6261.1997.tb01129.x

Heuer, A., Müller, C., & Rubner, O. (2010). Soccer: Is scoring goals a predictable Poissonian process? *EPL (Europhysics Letters)*, *89*(3), 38007. doi:10.1209/0295-5075/89/38007

Heuer, A., & Rubner, O. (2009). Fitness, chance, and myths: An objective view on soccer results. *The European Physical Journal B*, *67*(3), 445–458. doi:10.1140/epjb/e2009-00024-8

Hill, I. D. (1974). Association football and statistical inference. *Applied Statistics*, *23*(2), 203. doi:10.2307/2347001

Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, *26*(3), 460–470. doi:10.1016/j.ijforecast.2009.10.002

Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (the Statistician)*, *52*(3), 381–393. doi:10.1111/1467-9884.00366

Kelly, J. (1956). A new interpretation of information rate. *IEEE Transactions on Information Theory*, *2*(3), 185–189. doi:10.1109/TIT.1956.1056803

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. doi:10.1111/rssa.12042

Kovalchik, S. A. (2016). Searching for the GOAT of tennis win prediction. *Journal of Quantitative Analysis in Sports*, *12*(3), 311. doi:10.1515/jqas-2015-0059

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. doi:10.1504/IJAPR.2013.052339

Leitner, C., Zeileis, A., & Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008. *International Journal of Forecasting*, *26*(3), 471–481. doi:10.1016/j.ijforecast.2009.10.001

Leitner, C., Zeileis, A., & Hornik, K. (2011). Bookmaker consensus and agreement for the UEFA Champions League 2008/2009. *IMA Journal of Management Mathematics*, *22*(2), 183–194. doi:10.1093/imaman/dpq016

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. doi:10.1016/j.ijforecast.2009.12.013

Low, B., Coutinho, D., Gonçalves, B., Rein, R., & Memmert, D. (2019). A systematic review of collective tactical behaviours in football using positional data. *Sports Medicine*, 1–43.

Lucey, P., Bialkowski, A., Carr, P., Yue, Y., & Matthews, I. (2014). *How to get an open shot: Analyzing team movement in basketball using tracking data*. Proceedings of the 8th annual MIT SLOAN sports analytics conference.

Luckner, S., Schröder, J., & Slamka, C. (2008). On the forecast accuracy of sports prediction markets. In *Lecture notes in business information processing: Vol. 2. Negotiation, auctions, and market engineering* (Vol. 2, pp. 227–234). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-540-77554-6_17

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, *36*(3), 109–118. doi:10.1111/j.1467-9574.1982.tb00782.x

Manner, H. (2016). Modeling and forecasting the outcomes of NBA basketball games. *Journal of Quantitative Analysis in Sports*, *12*(1), 1128. doi:10.1515/jqas-2015-0088

Mauro, A. d., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. In *AIP conference proceedings* (pp. 97–104). AIP Publishing LLC. doi:10.1063/1.4907823

McHale, I., & Davies, S. (2007). Statistical analysis of the effectiveness of the FIFA world rankings. *Statistical Thinking in Sports*, 77–90.

McHale, I., & Morton, A. (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, *27*(2), 619–630. doi:10.1016/j.ijforecast.2010.04.004

Memmert, D., Lemmink, K. A. P. M., & Sampaio, J. (2017). Current approaches to tactical performance analyses in soccer using position data. *Sports Medicine*, *47*(1), 1–10. doi:10.1007/s40279-016-0562-5

Memmert, D., & Raabe, D. (2018). *Data Analytics in Football. Positional Data Collection, Modelling and Analysis*. Abingdon: Routledge.

Nevill, A. M., & Holder, R. L. (1999). Home advantage in sport. *Sports Medicine*, *28*(4), 221–236.

Newton, P. K., & Aslam, K. (2009). Monte Carlo tennis: A stochastic Markov chain model. *Journal of Quantitative Analysis in Sports*, *5*(3). doi:10.2202/1559-0410.1169

Ottaviani, M., & Sørensen, P. N. (2008). The favorite-longshot bias: An overview of the main explanations. *Handbook of Sports and Lottery Markets*, 83–101.

Pachur, T., & Biele, G. (2007). Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events. *Acta Psychologica*, *125*(1), 99–116. doi:10.1016/j.actpsy.2006.07.002

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. doi:10.1016/j.ijforecast.2017.08.002

Rue, H., & Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (the Statistician)*, *49*(3), 399–418.

Scheibehenne, B., & Bröder, A. (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting*, *23*(3), 415–426. doi:10.1016/j.ijforecast.2007.05.006

Smith, M. A., Paton, D., & Williams, L. V. (2006). Market efficiency in person-to-person betting. *Economica*, *73*(292), 673–689. doi:10.1111/j.1468-0335.2006.00518.x

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. doi:10.1086/655844

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. doi:10.1016/j.ijforecast.2007.05.003

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. doi:10.1002/for.1091

Stekler, H. O., Sendor, D., & Verlander, R. (2010). Issues in sports forecasting. *International Journal of Forecasting*, *26*(3), 606–621. doi:10.1016/j.ijforecast.2010.01.003

Štrumbelj, E., & Šikonja, M. R. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting*, *26*(3), 482–488. doi:10.1016/j.ijforecast.2009.10.005

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. doi:10.1016/j.ijforecast.2011.01.004

Wei, X., Lucey, P., Morgan, S., & Sridharan, S. (2013). *Sweet-spot: Using spatiotemporal data to discover and predict shots in tennis*. 7th Annual MIT Sloan sports Analytics Conference, Boston, MA.

Wigness, M. B., Williams, C. C., & Rowell, M. J. (2010). A new iterative method for ranking college football teams. *Journal of Quantitative Analysis in Sports*, *6*(2). doi:10.2202/1559-0410.1242

Woodland, L. M., & Woodland, B. M. (1994). Market efficiency and the favorite-longshot bias: The baseball betting market. *The Journal of Finance*, *49*(1), 269–279. doi:10.1111/j.1540-6261.1994.tb04429.x

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. doi:10.1080/02640414.2016.1218040

Wunderlich, F., & Memmert, D. (2018). The betting odds rating system: Using soccer forecasts to forecast soccer. *PloS One*, *13*(6), e0198668.