Institute of Exercise Training and Sport Informatics
German Sport University Cologne
Executive Head of Institute: Prof. Dr. Daniel Memmert

Sports forecasting – Current applications in sports science and moving towards Big Data

Doctoral thesis accepted for the degree
Dr. rer. nat.

by
Fabian Wunderlich

Cologne 2022

First reviewer: Prof. Dr. Daniel Memmert, Institute of Exercise Training and Sport Informatics, German Sport University Cologne, Germany
Second reviewer: Prof. Dr. Andreas Heuer, Institute for Physical Chemistry, University of Münster, Münster, Germany
Chair of the doctoral committee: Prof. Dr. Mario Thevis, Institute of Biochemistry, German Sport University Cologne, Germany.
Thesis defended on: 15.03.2022

Affidavits following § 7 section 2 No. 4 and 5 of the doctoral regulations from the German Sport University Cologne, February 20th 2013:

Hereby I declare:

The work presented in this thesis is the original work of the author except where acknowledged in the text. This material has not been submitted either in whole or in part for a degree at this or any other institution. Those parts or single sentences, which have been taken verbatim from other sources, are identified as citations.

I further declare that I complied with the actual "guidelines of qualified scientific work" of the German Sport University Cologne.

Fabian Wunderlich, Cologne, 18.03.2022

Acknowledgements

First of all, I would like to thank my supervisor Daniel for fully trusting me, giving me all the freedom I needed and giving me the opportunity to pursue my PhD at this very special place.

Caro, you are my biggest hero. You’ve made me the person I am, you have always been there, you’ve never doubted me, you make me smile and then you gave the cutest present ever. Mats, you are my smallest hero, it is so wonderful to see you grow up.

I would like to thank my parents, my sister, my aunts and uncles for loving me, believing in me, giving me all opportunities in the world and always supporting me.

Thank you Anne for becoming my work wife in #115. Office life (and other life) wouldn’t be half as fun without you.

I am grateful to all PhDs, colleagues and friends, who inspired me. Giulia, I like your positive aggressiveness and Dennis, I like your aggressive positivity. Ben, you offended me politely. Robert, you demotivated me friendly. Marius, you showed me I don’t understand football, but you don’t understand team choice. Niels you made me jealous, stop sending holiday pictures. Alex, Ann-Kathrin, Ash, Carina, Dominik aka Fabian, Doro, Laura, Marc, Max, Michel, Philip, Sebastian, you made work not feel like working.

This work would not have been possible without the help from my student assistants Alessandro Seck and Joris Wittkugel, as well as several further students and interns. Thank you for your dedicated work.

Finally, thanks to all my friends, you are the best! I won't start naming you or it will be hard to stop.

"It is very difficult to predict – especially the future."

Niels Bohr

Zusammenfassung

Der Wunsch zukünftige Ereignisse vorhersehen zu können, steht im Mittelpunkt prädiktiver Modelle, die - insbesondere im Bereich des Sports - interdisziplinäre Ansätze erfordern. Die vorliegende Dissertation fußt daher auf einer ganzheitlichen Betrachtung prädiktiver Modelle im Sport, die ökonomische und mathematische, sportwissenschaftliche sowie datenanalytische und informatische Aspekte einschließt.

Bei ökonomischer Betrachtung wird in der Regel die Profitabilität eines prädiktiven Modells genutzt, um seine prädiktive Qualität zu belegen. In der vorliegenden Dissertation wird allerdings anhand von theoretischen Überlegungen sowie simulierten und realen Datensätzen der Nachweis erbracht, dass auch mit ungenauen prädiktiven Modellen Wettgewinne erzielt werden können und somit Profitabilität und Genauigkeit bei prädiktiven Modellen im Sport als getrennte Konzepte angesehen werden müssen.

Im Sportbereich gibt es bereits vielfältige Evidenz für die Genauigkeit von kollektiven menschlichen Prognosen, was sich insbesondere in den im Wettmarkt angebotenen Wettquoten widerspiegelt. Auf Grundlage von Daten aus dem Fußballbereich konnte die hohe Genauigkeit von Wettquoten als Prädiktor in mehreren zusätzlichen Zusammenhängen bestätigt werden: Wettquoten vor dem Spiel liefern mehr Informationen über die Spielstärke einer Mannschaft als die tatsächlichen Spielergebnisse. Zudem sind Wettquoten ein wertvoller Prädiktor für den Erfolg von Mannschaften im Elfmeterschießen. Bezogen auf das Ergebnis der zweiten Halbzeit, verbessert sogar die Information über in der ersten Halbzeit erzielte Tore nicht die prädiktive Qualität von vor dem Spiel bekannten Wettquoten.

Theoretische Grundlagen für die hohe Genauigkeit von Wettquoten lassen sich aus Überlegungen zur Markteffizienz und zu kollektiven menschlichen Einschätzungen (Crowd Wisdom) ableiten. Dieses Wissen kann auch auf sportwissenschaftliche Fragen transferiert werden, indem aus Wettquoten Maße für die relative oder absolute Mannschaftsstärke sowie für (un)ausgeglichene Spiele und den Heimvorteil extrahiert werden.

Darüberhinaus trägt die vorliegende Dissertation dazu bei, das immer aktueller werdende Thema Big-Data-Analyse auf prädiktive Modelle im Sport zu beziehen. Die Charakteristika von textuellen Daten des Kurznachrichtendienstes Twitter sowie von Event- und Positionsdaten wurden in Bezug auf die Definition von Big Data diskutiert. Zudem wurde getestet, ob Datensätze aus beiden Bereichen die Qualität von prädiktiven In-Play-Modellen steigern. Für beide Datenquellen konnten keine Hinweise auf derartige Verbesserungen der In-Play-Prognosequalität gefunden werden. Dieses Ergebnis erlaubt allerdings nur begrenzte Rückschlüsse auf die Datenquellen, da gezeigt werden konnte, dass es im Allgemeinen schwierig ist, Prognosen für den weiteren Spielverlauf anhand von innerhalb des bisherigen Spielverlaufs verfügbar werdenden Informationen zu verbessern. Dies ist konsistent mit der Annahme, dass Torerfolge im Fußball einem relative stabilen statistischen Prozess entspringen, der im Wesentlichen durch die vor dem Spiel schätzbaren Spielstärken der beiden Mannschaften sowie zufällige Prozesse beeinflusst wird.

Die vorliegenden Ergebnisse haben theoretische Implikationen für die Leistungsanalyse im Sport sowie praktische Implikationen für Buchmacher, professionelle Sportwetter und Spielanalysten. Im Bereich der Leistungsanalyse im Fußball sollten Wettquoten standardmäßig als situative Variable genutzt werden. Spielanalysten, die während eines Spiels Schlussfolgerungen ziehen, sollten dabei aufpassen die im Laufe dieses Spiels verfügbar werdenden Informationen nicht überzubewerten. Gleiches gilt für Buchmacher bei der Entwicklung von In-Play-Quotenmodellen. Professionelle Sportwetter sollten zudem bei der Entwicklung prädiktiver Modelle die Auswirkungen der Unterschiede zwischen Genauigkeit und Profitabilität beachten.

Abstract

The desire to know what will happen, before it actually happens is at the heart of forecasting, which – in particular in the domain of sports – demands for interdisciplinary approaches. The present dissertation thus takes a holistic view on sports forecasting, including economic and mathematical aspects, aspects of sports science as well as aspects of data analysis and computer science.

While economically, the profitability of a forecasting model is commonly used to substantiate its predictive quality, the present dissertation uses theoretical considerations, as well as simulated and real-world data to prove that positive betting returns can be generated with inaccurate forecasting models and as such, profitability and accuracy need to be assessed separately in sports forecasting.

With regard to predictive accuracy, the high quality of collaborative forecasts in general and betting odds in particular is already well evidenced. In accordance with this, I have presented evidence for the high accuracy of betting odds based on football data in several further contexts: Pre-game betting odds provide more information about team strength than the actual match results, betting odds are a valuable predictor of success in penalty shootouts and first half goals are clearly outperformed by betting odds when forecasting second half goals in football.

The theories of market efficiency and crowd wisdom provide a theoretical framework for the forecasting accuracy of betting odds, which are also beneficial in the domain of sports science. Accordingly, the use of betting odds to obtain measures of relative or absolute team strength, an indicator for balanced or unbalanced matches and a measure of home advantage are highlighted methodologically.

Moreover, the present dissertation contributes to moving sports forecasting towards Big Data analysis. The characteristics of Twitter data as well as event and positional data as sources of Big Data in sports have been outlined and tested in the context of forecasts performed during the course of football matches (so-called in-play forecasting). For both data sources, no

evidence for improvements on in-play forecasting were found. However, this can be considered to be only partly, if at all, driven by the data itself. In-play forecasting in general has been evidenced to be a highly difficult task, which supports the notion that goal scoring in football, if controlling for pre-game expectation, is a highly stable process.

The present results have theoretical implications for performance analysts as well as practical implications for bookmakers, professional gamblers and match analysts. Performance analysts in football should standardly use betting odds as a situational variable. The value of in-play information should not be overvalued by match analysts when drawing conclusions during a match or bookmakers when compiling in-play betting odds. Moreover, professional gamblers should be aware of the differences in profitability and accuracy when designing forecasting models.

Overview of the Articles

Table 1 gives an overview of the scientific output, in which the author has been involved during the period of this doctoral programme. A total of nine articles (nine first authorships including one shared first authorship) are part of the synopsis of this cumulative dissertation and will be discussed in detail in the subsequent chapters. Two additional articles (co-authorship) are part of the synopsis, will be referenced in the subsequent chapters, but not discussed in detail. Two additional articles are not part of this synopsis. By the time of acceptance of this dissertation, all articles have been published in international peer-reviewed journals. Impact factors (IF) of the journals refer to the year of publication (wherever possible), or to the most recently published impact factor from 2019.

Table 1: List of Articles

	Articles, which are part of this synopsis and will be discussed in detail
I	**Wunderlich, F.**, & Memmert, D. (2021). Forecasting the outcomes of sports events: A review. *European Journal of Sport Science*, 21(7), 944-957. https://doi.org/10.1080/17461391.2020.1793002. [IF 2019: 2.6, 24/85 in Sports Sciences, 72th percentile, Q2]
II	**Wunderlich, F.**, & Memmert, D. (2020). Are betting returns a useful measure of accuracy in (sports) forecasting?. *International Journal of Forecasting*, 36(2), 713-722. https://doi.org/10.1016/j.ijforecast.2019.08.009. [IF 2019: 2.8, 66/373 in Economics, 83th percentile, Q1]
III	**Wunderlich, F.**, & Memmert, D. (2018). The Betting Odds Rating System: Using soccer forecasts to forecast soccer. *PloS one*, *13*(6). https://doi.org/10.1371/journal.pone.0198668. [IF 2018: 2.8, 24/69 in Multidisciplinary Sciences, 66th percentile, Q2]
IV	**Wunderlich, F.**, Berge, F., Memmert, D., & Rein, R. (2020). Almost a lottery: the influence of team strength on success in penalty shootouts. *International Journal of Performance Analysis in Sport*, *20*(5), 857-869. https://doi.org/10.1080/24748668.2020.1799171. [IF 2019: 1.5, 60/85 in Sports Sciences, 30th percentile, Q3]
V	**Wunderlich, F.**, Weigelt, M., Rein, R., & Memmert, D. (2021). How does spectator presence affect football? Home advantage remains in European top-class football matches played without spectators during the COVID-19 pandemic. *Plos one*, 16(3). https://doi.org/10.1371/journal.pone.0248590 [IF 2019: 2.7, 27/71 in Multidisciplinary Sciences, 62th percentile, Q2]

VI **Wunderlich, F.**, Seck, A., & Memmert, D. (2021): The influence of randomness on goals in football decreases over time. An empirical analysis of randomness involved in goal scoring in the English Premier League. *Journal of Sports Sciences.* 39(20), 2322-2337. https://doi.org/10.1080/02640414.2021.1930685. [IF 2019: 2.6, 27/85 in Sports Sciences, 69th percentile, Q2]

VII **Wunderlich, F**., & Memmert, D. (2020). Innovative Approaches in Sports Science - Lexicon-Based Sentiment Analysis as a Tool to Analyze Sports-Related Twitter Communication. *Applied Sciences, 1*0(2), 431. https://doi.org/ 10.3390/app10020431. [IF 2019: 2.5, 32/91 in Engineering, Multidisciplinary, 65th percentile, Q2]

VIII **Wunderlich, F.**, & Memmert, D. (2022). A big data analysis of Twitter data during premier league matches: do tweets contain information valuable for in-play forecasting of goals in football?. *Social Network Analysis and Mining*, 12(1), 1-15. [Impact 2020: 3.9, Q1 in Media Technology according to Scimago]

IX Klemp, M.*, **Wunderlich, F.***, & Memmert, D. (2021). In-play forecasting in football using event and positional data. *Scientific reports*, 11(1), 1-10. [IF 2020: 4.4, 17/72 in Multidisciplinary Sciences, 77th percentile, Q1]

*both authors contributed equally to this manuscript

Articles, which are part of this synopsis, but will not be discussed in detail.

X Yi, Q., Gómez-Ruano, M. Á., Liu, H., Zhang, S., Gao, B., **Wunderlich, F.,** & Memmert, D. (2020). Evaluation of the Technical Performance of Football Players in the UEFA Champions League. *International Journal of Environmental Research and Public Health*, 17(2), 604, https://doi.org/10.3390/ijerph17020604. [IF 2019: 2.8, 58/193 in Public, Environmental & Occupational Health, 70th percentile, Q2]

XI Yi, Q., Gómez-Ruano, M. Á., Liu, H., Gao, B., **Wunderlich, F.,** & Memmert, D. (2020). Situational and Positional Effects on the Technical Variation of Players in the UEFA Champions League. *Frontiers in Psychology*, 11, 1201, https://doi.org/10.3389/fpsyg.2020.01201. [IF 2019: 2.1, 45/138 in Psychology, Multidisciplinary, 68th percentile, Q2]

Further articles, which are not part of this synopsis

XII **Wunderlich, F**., Heuer, H., Furley, P., & Memmert, D. (2019). A serial-position curve in high-performance darts: The effect of visuomotor calibration on throwing accuracy. *Psychological research*, 1-8, https://doi.org/10.1007/s00426-019-01205-2. [IF 2019: 2.4, 31/89 in Psychology, Experimental, 66th percentile, Q2]

XIII Phatak, A., Mujumdar, U., Rein, R., **Wunderlich, F.,** Garnica, M., & Memmert, D. (2020). Better with each throw—a study on calibration and warm-up decrement of real-time consecutive basketball free throws in elite NBA athletes.

German Journal of Exercise and Sport Research, 1-7, https://doi.org/10.1007/s12662-020-00646-x.

Table of Contents

List of Tables

List of Figures

1. Introduction

Although each of us may be fascinated by the idea of living in the present moment, most people spend a considerable amount of time reflecting on the past or planning the future. The desire to anticipate and thus better control future events, seems to be part of the human nature. At this point, a variety of reasons could be mentioned that contribute to the human interest in forecasting the future. While one important reason is certainly pure curiosity, others are based on very practical incentives. The need for safety, which leads us to want to anticipate regional natural disasters or dangerous global developments such as climate change in order to react adequately and ultimately ensure our own survival. The importance of security, which makes it necessary to deal with future dangers ranging from everyday crime to threats of national security including terrorist attacks. The individual desire for well-being, that makes us look at the weather forecast every day, in order to choose the right clothing and to have an umbrella with us if necessary. But certainly, also economic interests, since accurate forecasts for stock prices, interest rates as well as trends and consumer behaviour are equivalent to financial success for the stakeholders involved. Driven by these and a multitude of further reasons, great effort has been and will be made to develop and test accurate forecasting models in business and research.

In the scientific area, journals such as the *International Journal of Forecasting* and the *Journal of Forecasting* deal exclusively with this topic and thus, in a way, regard forecasting as a separate discipline of science. In general, forecasting is most closely associated with the field of economics, although it includes a wide variety of different (economic and non-economic) application cases. For example, the everyday aspect of weather forecasts (Taylor & Buizza, 2004) and the current, highly socially relevant aspect of climate forecasting (Green, Armstrong, & Soon, 2009). In politics, researchers are interested in forecasting the results of elections (Wolfers & Leigh, 2002) and the development of political conflicts (Brandt, Freeman, & Schrodt, 2014). Other relevant aspects are criminality (Gorr, Olligschlaeger, & Thompson, 2003) and demographic developments due to ageing societies or migration (Booth, 2006). By far the most versatile field of application is economics, including forecasts related to financial markets focusing on stock returns, interest rates and prices for options or treasury bonds; macroeconomic forecasts related to output growth and unemployment rates (Timmermann, 2000) or energy forecasting „which includes but is not limited to the forecasting of the supply, demand and price of electricity, gas, water, and renewable energy resources." (Hong et al., 2016, p. 896). Another application of forecasting, which has attracted notable interest from researchers is the field of sports forecasting.

1.1. Sports forecasting

There is a large number of applications for sports-related forecasts, naturally including forecasts for the outcomes of events (Koopman & Lit, 2019), but also additional aspects such as injuries (Rossi et al., 2018), stadium attendance (Mueller, 2020), TV audiences (van Reeth, 2019), player movement on the pitch (Le, Carr, Yue, & Lucey, 2017) or talent development (Boulier, Stekler, Coburn, & Rankins, 2010). The International Journal of Forecasting has already dedicated a Special Issue twice to the topic of sports forecasting (McHale & Swartz, 2019; Vaughan Williams & Stekler, 2010), which highlights the relevance of sports-related applications in the domain of forecasting. In the context of this dissertation, the term sports forecasting shall only comprise forecasts related to intermediate or final outcomes of professional sports events, thus excluding the additional aspects mentioned above. A significant number of articles covering outcome-related sports forecasting are published mainly in journals related to forecasting, economics or statistics. In the literature both the terms *forecasting/forecast* and the terms *predicting/prediction* are common (see e.g. Andersson, Memmert, & Popowicz, 2009; Hvattum & Arntzen, 2010; Kovalchik, 2016; Leitner, Zeileis, & Hornik, 2010), however, without addressing the choice of each term or discussing differences in wording. In the context of this dissertation, the terms *forecasting/forecast* are consistently used, but common wording with relation to *predicting/prediction* such as *prediction markets*, *predictive*, *(un)predictable*, or *predictability* is maintained (cf. Article I).

Research on sports forecasting can be considered to relate both to basic and applied research. This is particularly true for the present dissertation, which takes an interdisciplinary approach including strongly application-based aspects, while touching theoretical aspects with mathematical, economic and sociological background. For this reason, the general practical and theoretical relevance of research on sports forecasting will be discussed first, before outlining the research questions and putting them into the context of existing literature. Theory-based aspects with regard to the individual articles, will be elaborated on in the further course of this dissertation (e.g. market efficiency and crowd wisdom with regard to using betting odds in sports science in section 3.1; theoretical explanations for the home advantage in section 3.4 and Article V; or dynamic systems theory related to event and positional data in section 4.7 and Article IX).

The application-based relevance of the present topic is strongly connected with the sports betting industry, which can be considered the practical side of forecasting in sports. The large and growing sport betting market (Nederlandse Online Gambling Associatie, 2015) provides significant financial incentives for bookmakers and professional gamblers. Both are driven by the need to gain an economic advantage over the other side by being in possession of sophisticated forecasting models (Baker & McHale, 2013; Koopman & Lit, 2019; Kovalchik, 2016; Manner, 2016).

Another factor is the high availability of data for sports events, including outcomes, further performance-related statistics and betting odds, that is driven by the massive public and media interest in these events. Sports data thus gain relevance as a fruitful real-world application to test methods from mathematics, statistics and increasingly computer science. This is evidenced by the decent number of sports forecasting methods published in statistical journals (e.g. Karlis & Ntzoufras, 2003; Koopman & Lit, 2015; Maher, 1982) or applying machine learning-based methods to sports data (Baboota & Kaur, 2019; Horvat & Job, 2020; Hubáček, Šourek, & Železný, 2019; Lessmann, Sung, & Johnson, 2010)

Theory-based relevance is associated with studies, where the forecast itself is only a means to an end, namely to verify respective theories originating, for example, from economics, sociology or psychology.

In this context, economic research is interested in forecasting models in sports as sports offers an excellent environment to empirically test efficient-market hypotheses (Angelini & Angelis, 2019; Direr, 2011; Goddard & Asimakopoulos, 2004). Violations of the idea of market efficiency can be represented by the identification of specific market inefficiencies (Braun & Kvasnicka, 2013) including, but not limited to the presumably best studied example of the favourite-longshot bias (Direr, 2011; Ottaviani & Sorensen, 2008). Such findings are not only of theoretical interest, as inefficiencies are evidence against the efficiency of sports betting markets, but also of practical financial interest by deducing profitable betting strategies.

A phenomenon with sociological and statistical background that is related to sports forecasting, is the idea of the wisdom of the crowd that goes back to the early work of Galton (1907) and assumes that groups of individuals are able to collaboratively make very precise estimations. By investigating collaborative forecasts for sports events, it has been demonstrated that this idea can be reinforced in various set-ups including the betting market (Forrest, Goddard, & Simmons, 2005; Štrumbelj & Šikonja, 2010), so-called prediction markets (Spann & Skiera, 2009) or community-based estimations of market values (Peeters, 2018).

Studies with psychological background are particularly driven by the interest to better understand human forecasts, considering forecasting heuristics (Pachur & Biele, 2007), as well as the ability and confidence of groups with different levels of knowledge in forecasting sports events (Andersson et al., 2009).

1.2. Research gaps, objectives and research questions

Despite focussing on a relatively specific and easily definable field of application, this dissertation can be considered to be highly interdisciplinary and combines aspects from the research areas of economics, sports science and computer science. This is a consequence of the fact that a useful predictive model in sports requires an (often economic) relevance, domain-specific knowledge in sports as well as useful methods of data analysis that usually stem from statistics or computer science. Due to the interdisciplinarity and the diversity of reasons for considering sports forecasting models, a number of different research gaps can be identified and will be divided into three main aspects, that will be considered in detail in chapters 2 to 4.

1) Economic and mathematical aspects (addressed in Article I and II)

2) Aspects of sports science (addressed in Article III-VI)

3) Aspects of data analysis and computer science (addressed in Article VII-IX)

Traditionally, publications in sports forecasting are largely based on an economic and mathematical approach. On the one hand, this can be attributed to the high economic relevance from both an application-based and a theory-based standpoint that has already been described. On the other hand, it can be explained by the fact that reasonable probabilistic forecasts cannot be obtained without the use of sound mathematical models (Baker & McHale, 2013; Cattelan, Varin, & Firth, 2013; Hvattum & Arntzen, 2010; Koopman & Lit, 2019, among many others).

Despite the considerable number of publications in this area, two aspects can be identified, which have not been given sufficient attention in the literature so far: Firstly, the absence of generalised approaches to sports forecasting, resulting in a lack of review articles or meta-analyses that deal with cross-thematic aspects or issues relevant in a variety of different sports. The article by Stekler, Sendor, and Verlander (2010) giving an overview on relevant topics and results in sports forecasting as well as a review focusing on sports forecasting as an application of machine learning (Horvat & Job, 2020) are very rare exceptions in this regard. Secondly, the very limited attention paid to the differences between accuracy and profitability in sports forecasting. Both measures of accuracy often denoted as score or loss functions (see Cattelan et al., 2013; Constantinou & Fenton, 2012; Hvattum & Arntzen, 2010) and measures of profitability, i.e. betting returns (see Constantinou, Fenton, & Neil, 2012; Koopman & Lit, 2015; Lessmann et al., 2010) are a standard tool to assess the quality of forecasting models. However, the differences between accuracy and profitability have only been explicitly discussed in other economic areas (Boothe & Glassman, 1987; Ertimur, Sunder, & Sunder, 2007; Fuertes, Kalotychou, & Todorovic, 2015; Leitch & Tanner, 1991), but not in the domain of sports forecasting so far.

Consequently, the following objectives and related research questions can be derived from these two research gaps. A literature review on general aspects of forecasting models in sports related to the question of how to categorise subjects of forecasting, sources of information and measures of predictive accuracy (Article I). This includes an overview of how systematic and unsystematic influences on the outcomes of sports events can be reasonably separated and modelled. Moreover, an investigation of differences between accuracy and profitability of forecasting models related to betting markets (Article II). This analysis addresses in particular the question of whether returns obtained from betting strategies (and used as a measure of profitability) are likewise a useful measure of accuracy.

Due to the economic and mathematical focus, forecasting in sports is hardly associated with sports scientific questions. Sports science can therefore be considered a neglected facet of sports forecasting, although forecasting models naturally benefit from a profound domain-specific knowledge of the underlying processes in a specific sports event. At the same time, predictive statistical modelling can help to gain insights to the inherent processes in sports games. For example, this has been demonstrated in football by investigating the degree of influence of randomness, team strength and other external factors on the results (Heuer, Müller, & Rubner, 2010), the influence of match time and recent goals on goal scoring patterns (Heuer & Rubner, 2012) and by confirming or falsifying football-related myths (Heuer & Rubner, 2009). Despite the potential to answer sports-related questions, such research is more likely to be found in statistical or multidisciplinary journals than in journals related to the domain of sports science. In this respect, sports science should be open to contribute to predictive modelling by transferring existing knowledge to the forecasting domain or by deducing relevant sports-specific knowledge from successful forecasting methods.

Although predictive approaches are increasingly used to rate situations and player behaviour in sports games (Cervone, D’Amour, Bornn, & Goldsberry, 2014; Dick & Brefeld, 2019; Seidl, Cherukumudi, Hartnett, Carr, & Lucey, 2018; Wei, Lucey, Morgan, & Sridharan, 2013), the lack of interdisciplinary transfer between forecasting and sports science can still be considered a notable research gap. One objective of the present dissertation is therefore to transfer a central finding from the sports forecasting domain and use it to improve sports scientific research. There is plenty of evidence that betting odds (i.e. forecasts directly derived from betting odds) possess an excellent predictive quality (Forrest et al., 2005; Hvattum & Arntzen, 2010; Kovalchik, 2016). Within this dissertation, this knowledge is used to answer sports scientific questions for the first time or more accurately. While transferring insights from the forecasting domain, the four research questions tackled within the present dissertation have, if at all, only a minor connection to forecasting models and thus concern a wide range of aspects. All studies relate to the domain of football, although the basic idea of using betting odds in sports scientific research can be applied to a wide variety of other sports as well. The first is concerned with the assessment of team strengths and answers the question of how to estimate these strengths as accurately as possible over the course of time (Article III). The second aims at answering the question of how the strength of a team influences its success in penalty shootouts (Article IV). The third study is related to the phenomenon of home advantage and analyses which effect the absence of spectators has on the degree of home advantage (Article V). The fourth study quantifies and analyses to what extent random processes contribute to goal scoring processes and in which match situations random goals are more likely to occur (Article VI).

The arguably most obvious research gap, for various reasons, can be identified in the domain of data analysis and computer science. These domains are associated with the topics of Big Data (Mauro, Greco, & Grimaldi, 2016) and Machine Learning (Dey, 2016), which have undoubtedly had a major impact on economy, society and science in recent years. Although some of the basic methods have been known for decades (Ho, 1995; Hochreiter & Schmidhuber, 1997), their usage has just recently gained importance and is primarily driven by the rapidly increasing availability of data and computational power. Machine learning methods have already been applied to sports forecasting (Baboota & Kaur, 2019; Horvat & Job, 2020; Hubáček et al., 2019), but due to the relative novelty of these approaches and the still increasing availability of data, it can be assumed that the exploitation of this branch of research has just begun. The present dissertation contributes to tackling this research gap by testing whether Big Data approaches can help to improve in-play forecasting models in the domain of football. To do this, two possible sources of Big Data are considered: Short textual messages published on Twitter (so-called tweets) as well as event and positional data. Both data sources can be attested a highly unstructured character. While numerous studies have already dealt with the extraction of relevant information from event and positional data in football (Grunz, Memmert, & Perl, 2012; Lago-Ballesteros & Lago-Peñas, 2010; Lepschy, Wäsche, & Woll, 2020; Memmert, Lemmink, & Sampaio, 2017; Rein, Raabe, & Memmert, 2017), the question of whether valid information can be extracted algorithmically from football-related textual data is answered in this dissertation (Article VII). The primary research question tackled, however, is to what extent information extracted from such textual data (Article VIII) as well as event and positional data (Article IX) can be useful to improve in-play forecasting models.

In summary, the present dissertation answers a multitude of research questions that address the topic of sports forecasting from three different perspectives. The objectives and research questions as well as the related articles are summarized in Table 2.

Table 2: Overview about the objectives and research questions tackled within this dissertation

	Objectives and research questions	Addressed in
1	***Economic and Mathematical Aspects*** General aspects of sports forecasting	

	a)	How can subjects of forecasting, sources of information as well as measures of predictive accuracy be categorised in sports forecasting and how can systematic and unsystematic influences on the outcomes of sports events be separated?	Article I
	b)	Are betting returns a useful measure of accuracy in sports forecasting?	Article II
2		***Aspects of Sports Science*** Using betting odds in the domain of sports science	
	a)	How can the strength of a football team be determined accurately over time?	Article III
	b)	How does the strength of a football team influence its success in penalty shootouts?	Article IV
	c)	Which effects does the absence of spectators have on the home advantage in football?	Article V
	d)	To what extent do random processes influence the scoring of goals in football and in which match situations do random goals occur most frequently?	Article VI
3		***Aspects of Data Analysis and Computer Science*** Using Big Data in sports forecasting	
	a)	Are algorithmic methods of sentiment analysis accurate enough to correctly classify the sentiment in football-specific textual messages?	Article VII
	b)	Can information extracted from Twitter data help to improve in-play forecasting models in football?	Article VIII
	c)	Can information extracted from event and positional data help to improve in-play forecasting models in football?	Article IX

2. Economic and mathematical aspects

The majority of studies in the field of forecasting models in sport are based on an economic and mathematical approach. On the one hand, this refers to the economically driven relevance of these studies, which can include the identification of profitable betting strategies (Lessmann et al., 2010), the verification of market efficiency hypotheses (Angelini & Angelis, 2019), the investigation of various market structures (Franck, Verbeek, & Nüesch, 2010) or the presentation of inefficiencies in the market (Ottaviani & Sorensen, 2008). On the other hand, it refers to the need for statistical methods and mathematical models in order to come up with accurate forecasts. Poisson models (Koopman & Lit, 2015), regression models (Goddard & Asimakopoulos, 2004), so-called birth-process models (Baker & McHale, 2013), simulation based on Markov models (Štrumbelj & Vračar, 2012) as well as ELO-ratings (Hvattum & Arntzen, 2010) and all sorts of adaptive stochastic processes for modelling strength parameters (Cattelan et al., 2013) represent only a selection of the most frequently used methods and models.

2.1. General aspects of sports forecasting

2.1.1. Previous research

As explained in the Introduction, there exists a broad and multi-faceted literature on sports forecasting. The high diversity is also reflected in the variety of sports that have been used as use cases, including football (Koopman & Lit, 2019), tennis (Kovalchik, 2016), American football (Baker & McHale, 2013), baseball (Soto Valero, 2016), basketball (Štrumbelj & Vračar, 2012), horse racing (Lessmann et al., 2010), handball (Groll, Heiner, Schauberger, & Uhrmeister, 2020), cricket (Asif & McHale, 2016) and many more. The focus of the individual publications, however, is almost exclusively on a single sport and only in very few studies the approaches are applied to more than one sport (for exceptions, see Cattelan et al., 2013; Karlis & Ntzoufras, 2003). Attempts to go beyond the scope of individual aspects are even more scarce, although it seems reasonable to approach reviews of the sports forecasting literature in order to tackle general aspects of sports forecasting and putting them into a broader context. Some of the very few cross-thematic papers in this domain, that have been published so far, are the article of Stekler et al. (2010) summarising relevant topics in sports forecasting, or the very recent review of Horvat and Job (2020) on machine learning in sports forecasting. Williams (1999) also pursues the strategy of putting existing literature into a broader context, however, he focuses on the narrow topic of market efficiency in sports betting. Motivated by this paucity of research, Article I contributes to answer the question of how sports forecasting methods can be categorised and how systematic and unsystematic influences on the outcomes of sports events can be separated.

2.1.2. Answer to Research Question 1a)

Given that Article I represents a literature review, the answer to this research question consists of synthesised and structured information from the sports forecasting literature on the following aspects: A summary of topics and research questions; a categorization with regard to the subject of forecasting, source of information and measures of predictive quality as well as the separation of modelling systematic and unsystematic influences on the outcomes of sports events. Two main results of the article shall be highlighted in particular.

First, the article introduces a detailed categorization for the source of information that forecasts are based on. The first category includes forecasts by human judgement and is subdivided into those forecasts made individually or collaboratively. The second category includes quantitative models and is subdivided into those models making use of external ratings or rankings and those based on internal ratings/rankings or not ratings/rankings-based at all.

In the future, this categorisation will make it possible to easily classify methods from existing or new studies into one of the four categories. Interviewing participants with questionnaires including predictive tasks (see Andersson et al., 2009), for example, falls into the category of individual human judgement. Prediction markets (Luckner, Schröder, & Slamka, 2008) are specifically designed to profit from collaborative human judgement and are therefore a prime example for this category. The first group of quantitative models refers to models making use of an external rating or ranking like the benchmark model of McHale and Morton (2011) based on the official ATP ranking in tennis. Such rankings are predominantly designed to reward for past performance and are used as seedings for tournament draws, but usually not optimised for predictive purposes. The second group of quantitative models, in contrast, deduces the forecasts directly from the given data, or by first estimating one or more rating parameters, which are specifically designed for forecasting purposes (Hvattum & Arntzen, 2010; Koopman & Lit, 2015; Manner, 2016).

This categorization also includes the opportunity of examining broader groups of studies to take a comparative view on the superiority or inferiority of different sources of information. For example, whether collaborative human judgement is always superior to individual human judgement and whether there are certain set-ups (e.g. sports or competitions) in which statistical models are inferior or superior to human judgement.

Second, a distinction between different factors influencing the outcomes of sports events is introduced (see Figure 1). This can be considered a very general concept of sports forecasting, indicating the most important aspects that need to be tackled in the process of developing a mathematical forecasting model. The first factor comprises all systematic participant-specific factors and thus represents the quality of a participant. The second factor includes global factors like the home advantage (Courneya & Carron, 1992; Nevill & Holder, 1999, see also Article V) that systematically affect the outcomes, but are not attributable to a participant. The third factor refers to the inherently random processes, that are not systematically predictable, but result in a probability distribution of results.

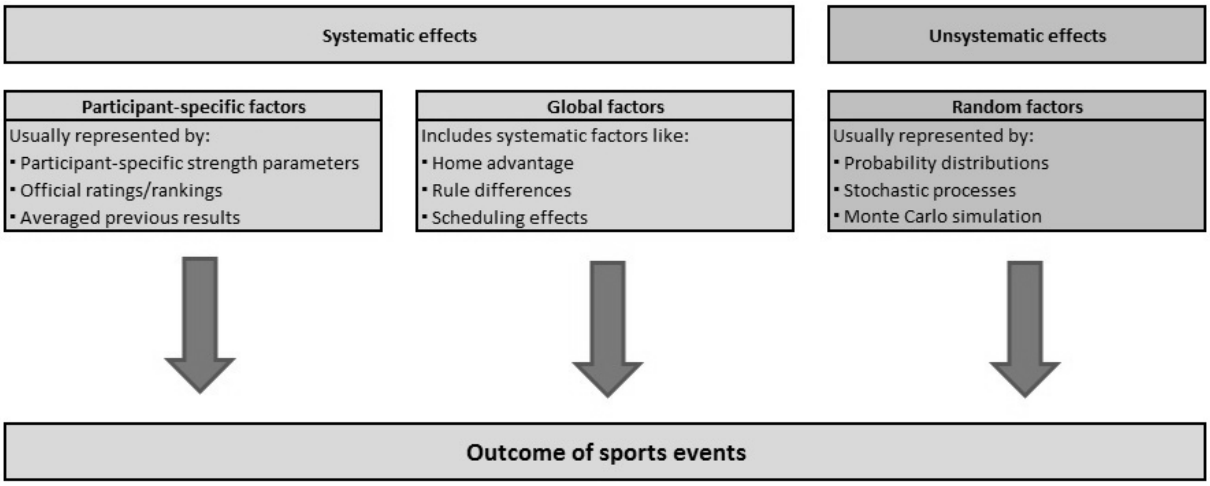

Figure 1: Conceptual illustration of the factors determining the outcomes of sports events

This concept illustrates that sports forecasting goes far beyond an attempt to foresee the outcome in advance, but rather requires domain-specific knowledge in order to accurately and correctly model the inherent processes of a sport. It is also a reference to the important impact of randomness on sports outcomes that is well-accepted in forecasting, but does not always seem to be acknowledged in sports practice.

2.2. Differences between accuracy and profitability

2.2.1. Previous research

The evaluation of forecasting methods requires adequate measures, which are designed to assess the quality of forecasts. Two different strands of literature need to be outlined before approaching this very central aspect of forecasting. On the one hand, the state of art in measuring predictive power in the sports forecasting literature and, on the other hand, the discussion on differences between accuracy and profitability in other economic fields of forecasting.

In the sports domain, the arguably most simplistic approaches to measure predictive power are reporting the proportion of ‘correct’ forecasts (Song, Boulier, & Stekler, 2007) or reporting the correlation between a forecasted ranking and the observed ranking (Leitner et al., 2010). Such measures have advantages in terms of comprehensibility and comparability, but are not sufficient to take the complexity of probabilistic forecasts into account in a meaningful way. For this reason, more complex measures, which can be divided into statistical measures and economic measures, have become the standard for probabilistic forecasting models.

Statistical measures are based on comparing probabilities forecasted prior to an event and the actual outcome observed after the event. This is either done by only considering the probability of the outcome that actually occurred (e.g. log-likelihood, Koopman & Lit, 2015) or by considering the (quadratic) differences between forecasted and actual outcomes often denoted as quadratic loss or Brier score (Štrumbelj & Vračar, 2012). Although a decent number of different measures and inconsistent names for highly related measures are used, the general approach of statistical measures is largely similar. The question which statistical measure is the most sensible for the domain of sports forecasting has already been explicitly addressed (Constantinou & Fenton, 2012) and is still subject to discussion (Wheatcroft, 2019).

Economic measures aim at assessing the financial profit of forecasting models with reference to the sports betting market. Knowing the estimated model probabilities and the bookmakers' betting odds, it is straightforward to decide which outcomes to bet on before an event. Once the observed outcomes are known, the financial profit or loss (i.e. the betting returns) of these bets can be calculated. Different betting strategies are conceivable, for example, betting one unit if the expected value is positive, standardising the stakes in a way that each winning bet is worth one unit or determining the stakes with reference to the so-called Kelly criterion (see e.g. Hvattum & Arntzen, 2010).

It has become quite common to concurrently report statistical and economic measures in the domain of sports forecasting (Hvattum & Arntzen, 2010; Koopman & Lit, 2015; McHale & Morton, 2011), but there is no literature that specifically addresses the relationship between the statistical accuracy of a forecasting model and the economic profitability of the associated betting strategy so far. This is different from other economic fields, where differences between accuracy and profitability have been acknowledged and discussed for a long time reporting mixed results across different applications. Ertimur et al. (2007) investigated whether analysts being capable of performing accurate earnings forecasts are able to translate these forecasts to profitable stock recommendations and found a “strong positive association between accuracy and profitability” (p. 568). Boothe and Glassman (1987) compared accuracy and profitability in the domain of exchange rate forecasting and concluded that “models with the most accurate forecasts are not always the most profitable” (p. 66). Leitch and Tanner (1991) argue that economic forecasts can be financially valuable even if they do not outperform naïve models in terms of statistical accuracy measures. They substantiate their claim by analysing interest rate forecasts finding only very weak relationship between accuracy and profitability. Likewise, Fuertes et al. (2015) investigated models to forecast equity volatility and concluded “that statistical significance does not have a direct mapping onto economic value” (p. 251). Although the results and conclusions from economic domains are inconsistent, there is awareness that accuracy and profitability are two different goals of forecasting models and need to be assessed separately.

2.2.2. Answer to Research Question 1b)

With regard to research question 1b), Article II provides strong evidence that betting returns are not a useful measure of accuracy in sports forecasting or other domains of forecasting related to betting markets. This answer is based on theoretical considerations, the results of a simulation model and the analysis of real-world datasets from three different sports.

From a theoretical standpoint, the betting market can be described as asymmetric in the sense that the bettor has a systamtic advantage over the bookmaker. While the bookmaker needs to set the betting odds for each outcome, it is sufficient for the bettor to identify outcomes where the betting odds have been set to high in order to bet profitably. To do this, the bettor does not necessarily require an accurate probability estimation, which is a central reason why profitable forecasts in absence of an accurate forecasting model are possible. Article II even constructs an example, denoted as the profitability paradox, in which one actor's model can be profitably used to exploit the weaknesses of another actor's model, while the reverse is true as well.

A simulation model was used to examine the relationship between accuracy and profitability in a more generally applicable way. The model makes it possible to control for the degree of forecasting errors made by bookmaker and bettor (i.e. their accuracy) as well as the correlation between these errors. It simulates the (true) probabilities of outcomes as well as the estimated probability by the bettor and the betting odds offered by the bookmaker and consequently the betting returns can be calculated for a variety of different predefined set-ups. Figure 2 gives an example of results obtained from this simulation and illustrates how this approach can help to understand factors influencing the profitability of a model. The figure contains three scenarios with different forecasting errors, where the errors of bookmaker and bettor are equal, i.e. the models of both actors do not differ in terms of accuracy. Moreover, the returns for each scenario were calculated based on a variety of different correlations.

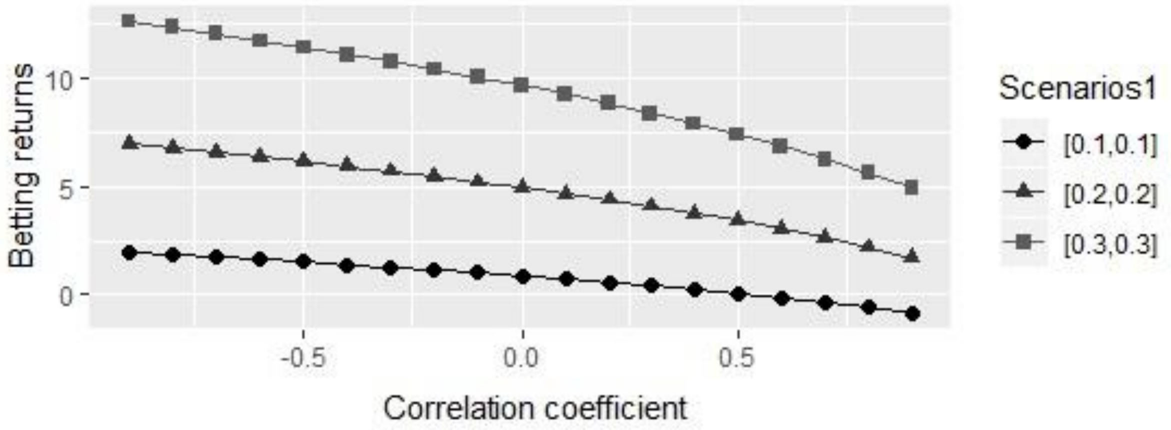

Figure 2: Mean relative betting returns for various combinations of forecasting errors of bookmaker and bettor as well as correlation of errors based on a simulation model.

The central result of the simulation is that the betting returns decrease if the correlation between the errors increases, which is explained by the fact that profitable bets are better identifiable if the error of the models is in the opposite direction. Moreover, the almost entirely positive betting returns are further evidence that a models' forecast can be profitable, although it is not superior to the bookmaker model in terms of accuracy. Finally, it can be seen that the betting returns increase if the errors of the models increase, which means that larger bookmaker errors imply an increased opportunity to bet profitably even if the bettor model is likewise subject to increased errors.

Using the simulation model, a whole series of other error combinations were systematically investigated, and some generally valid results were identified. First, there is a connection between accuracy and profitability in a sense that – if all other aspects are constant – a higher accuracy implies a higher profitability. Second, positive betting returns do not imply that the bettor is in possession of a superior model in terms of forecasting accuracy. Third, if the bettor model possesses at most an equal accuracy compared to the bookmaker, the betting returns decrease with an increasing correlation of errors.

Another reason why profitable betting strategies need to be assessed with caution is the inherent randomness in betting returns. To investigate this, random betting strategies were applied to real-world datasets, which means that a certain number of outcomes was randomly chosen to be bet on and the profits were calculated. Again, Figure 3 illustrates an example of the results that were obtained from this analysis for a women's tennis dataset. The results can be interpreted in a similar way to the p-value of a statistical test. If obtaining a certain betting return from a strategy with a certain number of bets, the figure shows how many random strategies were able to generate equal or even higher returns. Referring to a probability of 5% as commonly chosen in significance testing, even for relative betting returns of 0.1 (10%) in 100 bets or 0.15 (15%) in 50 bets, randomness cannot be excluded as a reason for the profit generation.

This example is clear evidence that caution is needed, as not only systematic reasons, but also randomness is complicating the interpretation of betting returns. Analysis of similar datasets from soccer and American football are reported in Article II and suggest that the danger of random positive returns depends on the sport and the specific dataset. In particular, the bookmaker margin and the existence of longshots in the odds can be argued to influence the risk of randomly profitable betting strategies. As a consequence of Article II, it would be desirable that researchers specify the probability that such a return is generated by randomness, whenever reporting betting returns to highlight the power of a forecasting model.

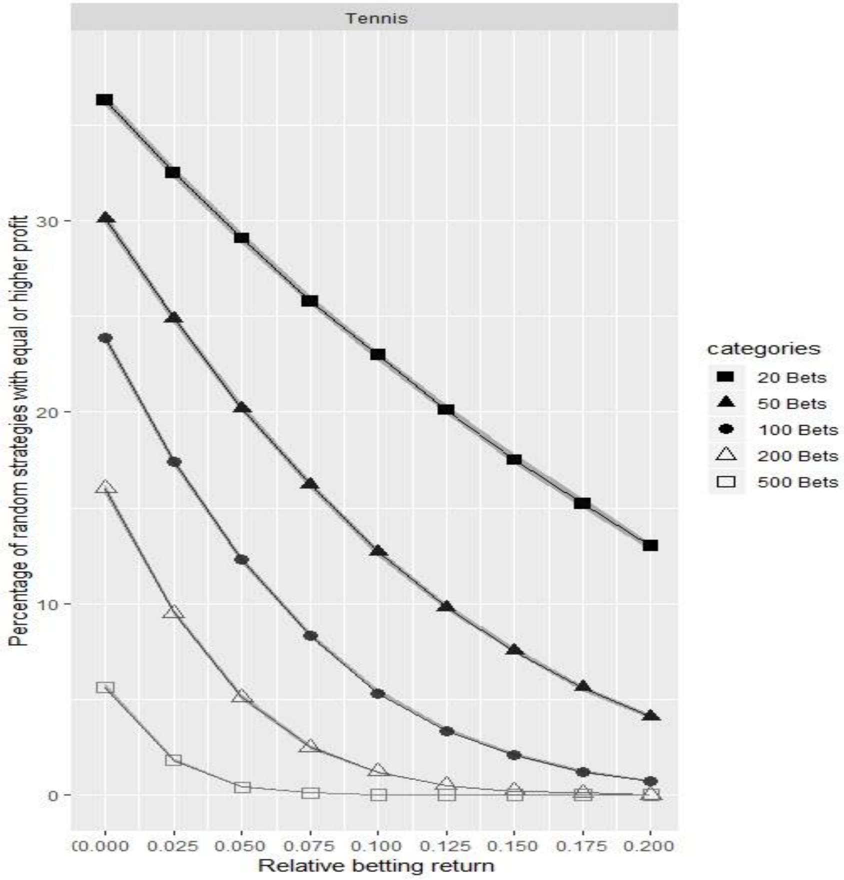

Figure 3: Percentage of random betting strategies whose returns exceed a specific threshold based on a real-world dataset of WTA (Women's Tennis Association) tennis matches.

In summary, the connection between model accuracy and profitability is by far not as unambiguous as might be assumed. In particular, it is easier possible to develop a profitable forecasting model than one that outperforms the betting market in terms of accuracy. This has two important implications if insights for sports science are to be derived from forecasting models, like in the next chapter. First, forecasting models should reflect the exact processes of the sport and therefore be selected on the basis of statistical measures. Second, the mere existence of profitable betting strategies does not imply that the predictive accuracy of betting odds should be doubted.

3. Aspects of sports science

A purely economic and mathematical approach to sports forecasting can certainly be criticised as one-dimensional. After all, the actual subject of forecasting, i.e. a sports event should not be neglected and the creation of a forecasting model always requires a profound knowledge of the underlying sport. The present chapter, however, is not focused on transferring knowledge from sports science to be used in the modelling process, but specifically on the question, whether knowledge from the sports forecasting domain can be valuably transferred to tackle sports scientific research questions. This refers in particular to the idea of using betting odds as a rich source of information for evaluations in the game of football.

3.1. Betting odds

3.1.1. Theoretical considerations

The business model of bookmakers is to give bettors the opportunity to bet on several outcomes related to professional sports events. To take account of the different likelihood of these outcomes, bookmakers set and publish betting odds, which determine the financial payoff, if the bettor wins a bet. Once bettors start to bet on an event, bookmakers can adjust the odds due to the betting behaviour in order to minimise their financial risk or to account for inaccurate odds setting. While the exact process of odds creation and risk management can be considered a trade secret of each bookmaker, the betting odds themselves are published and therefore openly available. Accordingly, betting odds can be used as an easily accessible implicit forecast of the betting market. From a theoretical standpoint, two mechanisms give rise to the assumption that betting odds should have a high predictive quality: Information efficiency in markets and wisdom of the crowd.

It appears highly plausible that in a market situation with strong financial incentives for the parties involved (such as the sports betting market), market prices (i.e. betting odds) are able to accurately reflect available information. If some relevant information were not reflected in market prices, bettors would be able to use this information and financially exploit bookmakers and other bettors. This idea is at the heart of research on market efficiency in economics, that has been heavily influenced by the work of Fama (1970), who stated that “A market in which prices always ‘fully reflect’ available information is called ‘efficient’.” (p. 383).

With reference to the considerations of Fama (1970), modern efficient market theory assesses three different degrees of efficiency, based on increasing levels of information. Weak form efficiency, where the information basis considered is just historical prices; semi-strong form efficiency that includes all further publicly available information and strong form efficiency that even includes insider information exclusively available to certain groups or individual persons. Investigations on the efficiency of sports betting markets have been argued to even possess an advantage over stock markets or comparable financial markets, as the sports betting market is “characterized by a well-defined termination point at which each asset (or bet) possesses a definite value” (Williams, 1999, p. 1).

Regardless of formal tests for the efficient market hypothesis, it can be assumed that the inherent principles of markets are predestined to allow for a relatively high degree of information efficiency. Moreover, as shown in Article II, even a violation of market efficiency by generating profits with a forecasting model does not necessarily imply that the forecasting accuracy of this model is superior to the market.

A further mechanism that theoretically supports the notion of accuracy in betting odds is the idea that a group of individuals is collaboratively able to outperform individual judgement in estimation tasks including forecasts and to come up with highly accurate results. It is often associated with Galton (1907) who observed this phenomenon when a group of individuals was asked to estimate the weight of an ox and the median of these estimates was remarkably accurate in approximating the actual weight. Starting from this early anecdotal evidence, the phenomenon has been investigated in the scientific literature, widely popularised by the book of Surowiecki (2005) and is denoted as crowd wisdom or wisdom of the crowd. Surowiecki (2005) underpins the idea of crowd wisdom by giving illustrative examples from stock markets, political elections, quiz shows and also sports betting.

The notion that any crowd is always estimating accurately in every situation, however, is hardly imaginable, which is why crowds are assumed to require diversity of opinion, independence, decentralisation and aggregation to be wise (Hosseini et al., 2015; Surowiecki, 2005). The phenomenon that aggregated judgements are superior to individual judgements, can be evidenced statistically. Davis-Stober, Budescu, Dana, and Broomell (2014) define a crowd to be "wise if a linear aggregate, for example a mean, of its members' judgments is closer to the target value than a randomly, but not necessarily uniformly, sampled member of the crowd" (p. 79) and evaluated crowd wisdom by using a mathematical framework. In line with the idea of independence and diversity, the authors state that crowd wisdom benefits from uncorrelated or even negatively correlated individual judgements. Moreover, crowd wisdom was found to be a robust effect and highly likely to be proven true even if the individual judgements are correlated and biased. Lorenz, Rauhut, Schweitzer, and Helbing (2011), while not denying that crowds are wise if judging independently, present experimental evidence that this effect can be negatively affected by social influence, i.e. if individuals are made aware of the estimation of others.

Independence is evidently not given in sports betting, as the odds are observable at any time by the bettors and other bookmakers in the market. However, the other three requirements can be considered to be fulfilled. In line with the idea of diversity and decentralisation, the betting market can be assumed to involve a large number of individuals with a different degree of knowledge, as the odds can be influenced by professional odds compilers working for the bookmakers and professional gamblers as well as by a large crowd of recreational bettors with a strongly varying degree of expertise in sports and forecasting. Moreover, the market represents a viable mechanism to aggregate the explicit or implicit estimations of these individuals in the betting odds.

In summary, theoretical considerations give rise to the assumption that the structure of betting markets leads to a high degree of information efficiency and aggregation of collaborative knowledge resulting in a high forecasting accuracy. These theoretical considerations are not limited to the sports domain and have led to the idea of intentionally creating predictive market structures, so-called prediction markets (Wolfers & Zitzewitz, 2004), in order to deduce accurate forecasts. Prediction markets are strongly related to sports betting, as payoffs are directly dependent on the outcome of events and can be focused on domains as diverse as geopolitical risk, presidential elections, economic statistics or the success of movies (Wolfers & Zitzewitz, 2004).

3.1.2. Empirical evidence

If these theoretical considerations hold true, there should be evidence for the high predictive accuracy of betting odds independent of the specific situation considered. Indeed, this is a very robust finding across different sports (e.g. football: Hvattum & Arntzen, 2010; tennis: Kovalchik, 2016; American football: Baker & McHale, 2013; basketball: Štrumbelj & Vračar, 2012). Moreover, it has been argued that the predictive quality of betting odds has further increased over time, most probably due to a more intensive competition in the market (Forrest et al., 2005; Štrumbelj & Šikonja, 2010).

Researchers have compared betting odds to a range of other forecasting methods and - despite the difficulty to compare forecasting results across studies - it can be summarised that other methods are consistently found to be inferior or at most comparable to the accuracy of betting odds. This particularly includes sophisticated statistical models specifically designed for forecasting purposes (Baker & McHale, 2013; Forrest et al., 2005; Hvattum & Arntzen, 2010; Kovalchik, 2016), that are often capable of coming close to the accuracy of betting odds, yet without outperforming them. As models are optimized to make the best use of the information available to them, this supports the notion that the market is equally efficient in processing this information or even capable of reflecting additional information. However, it remains unknown whether the main driver of the accuracy in betting odds is the financial incentives to forecast accurately, the market situation, or the crowd wisdom.

In line with the idea of crowd wisdom, individual human judgement (Spann & Skiera, 2009) or heuristics based on human recognition (Herzog & Hertwig, 2011) tend to be clearly outperformed by betting odds, while sources of collaborative knowledge, including, but not limited to the betting market are capable of providing high accuracy (Forrest et al., 2005; Peeters, 2018; Spann & Skiera, 2009). Interestingly, there are indications that real financial incentives are not necessarily required for accurate forecasts, as prediction markets (Spann & Skiera, 2009) show good results, despite only offering play money or low-value prizes to the participants. Moreover, the results of Peeters (2018) suggest that even a market situation with indirect financial incentives is not required to profit from collaborative knowledge, as accurate forecasts were deduced from the values of football players collaboratively estimated by users of a website.

Explicit tests for betting market efficiency with regard to to weak form efficiency report mixed results (Angelini & Angelis, 2019; Direr, 2011). While the favourite-longshot bias, representing a slight mispricing of favourites and longshots, is strongly evidenced and discussed (Direr, 2011; Ottaviani & Sorensen, 2008), other large-scale evaluations support weak form efficiency with regard to average betting odds (Angelini & Angelis, 2019). Moreover, semi-strong efficiency has been tested by Bernardo, Ruberti, and Verona (2019) who report that betting odds inefficiently reflect the impact of head coach replacements. Goddard and Asimakopoulos (2004) report to test weak form efficiency, although including information like involvement in cup competitions or significance of matches that can rather be associated with semi-strong efficiency. The authors find indications of market inefficiency, in particular for the final weeks of a season. Even without focusing on market efficiency tests, any published evidence of profitable betting strategies (Boshnakov, Kharrat, & McHale, 2017; Constantinou et al., 2012; Hubáček et al., 2019; Koopman & Lit, 2015), if considered to be of a systematic nature, constitutes a violation of market inefficiency.

In a real-world market situation, however, it is hardly imaginable to not come across any evidence contradicting efficiency at all, which is why Fama (1970) already stated that for the efficient market hypotheses “like any other extreme null hypothesis, we do not expect it to be literally true” (p. 388). In that sense, existing findings of inefficiencies are not yet in full contradiction with the assumption that markets conceptionally possess a high degree of information efficiency.

Moreover, within this dissertation, the forecasting accuracy of betting odds is of central interest, while the focus of efficient market hypotheses is on the (im)possibility of generating profits. As generating profits does not imply a superior forecasting accuracy (see Article II), inefficiencies in the market are not necessarily evidence to the contrary of a markets' high forecasting accuracy.

3.1.3. Usage of betting odds in sports science

Despite the ongoing discussion on formal efficiency, the excellent predictive accuracy of betting odds is a very robust and well-established result from the domain of sports forecasting. In this respect, betting odds can be considered to reflect a high degree of information on sports events, in particular publicly available information on systematic and unsystematic influences as specified in Article I. In view of the theoretical considerations and the empirical evidence, the non-usage of betting odds in sports science appears to be an unexploited potential. While this idea could arguably be applied to every sport that is covered by the betting market, the articles presented subsequently are limited to the game of football.

In football, three main aspects that should conceptionally be reflected in the betting odds are the strengths of both teams, the home advantage and the inherent randomness of football matches. The following articles will provide application cases for extracting such information from betting odds and using them to gain an improved understanding of the game.

3.2. Estimation of team strength

3.2.1. Previous research

The estimation of team strength is an integral part of forecasting models as has been discussed in Article I. Such ratings represent a quantification of all skills of a team that are assumed to contribute to the probability of winning a match. Usually this is operationalised as a single number (Hvattum & Arntzen, 2010), or separate parameters for offensive and defensive abilities (Koopman & Lit, 2015) that might additionally be differentiated between home and away abilities (Maher, 1982). Likewise, external sources can be used as a proxy for team quality, e.g. world rankings (Lasek, Szlávik, & Bhulai, 2013) or seedings (Boulier & Stekler, 1999).

Several studies have addressed the question of whether ratings or rankings have predictive value at all or have compared the predictive quality of different approaches. Boulier and Stekler (1999) conclude that seedings (which are primarily based on previous success) are useful predictors in basketball and tennis. McHale and Morton (2011) demonstrate that forecasts based on a self-designed method for strength estimation are superior to forecasts based on the official ATP rankings in tennis. In a comprehensive comparative study, Lasek et al. (2013) even compared the predictive accuracy of fifteen specifications of eight different rating models in football. The authors demonstrated that the majority of these models was able to outperform the official FIFA World Ranking.

One of the most established and widely applicable predictive rating systems is the so-called ELO rating, which has been used for many years to determine the playing strength of chess players (Glickman & Jones, 1999). Kovalchik (2016) compared forecast from four different classes of models including a total of fifteen different model specifications in tennis and concluded that “a predictive method based on Elo ratings was the closest competitor to bookmaker predictions” (p. 135). Hvattum and Arntzen (2010) presented a model transferring the ELO rating to the domain of football. In particular, they made use of two slightly different versions of the ELO rating, based on the results of a match in terms of win, draw or lose and on the exact score of a match in terms of goals respectively. The ratings have proven their quality by outperforming forecasting models introduced by Goddard (2005), when using ordered logistic regression models to transfer ELO ratings to forecasts.

In line with the results presented in section 3.1, even the successful ELO based models investigated by Kovalchik (2016) as well as Hvattum and Arntzen (2010) have been shown to be inferior to betting odds. It can therefore be considered a bit surprising, that no approaches have attempted to use the information reflected in betting odds as a basis to develop improved ratings so far. The only study known to the author, that points in this direction, is the study of Leitner et al. (2010). With regard to the European Championships 2008 in football, the authors use a reverse tournament simulation to obtain team-specific ratings from the betting odds to win the tournament. This enables the authors to directly compare ratings derived from betting odds to ratings derived from the ELO rating and the FIFA World Ranking. Moreover, it sheds light on the interaction between team strength and winning probability that can be confounded by the tournament draw. However, no betting odds on a match-level are considered and the database (one tournament) is highly limited. Moreover, the analysis of predictive quality can be considered as rather undetailed, as it is limited to calculating Spearman's correlation between the actual tournament ranking and the forecasted ranking deduced from the ratings. Compared to rankings based on ELO and the FIFA World Ranking, the ranking based on betting odds was reported to have a higher correlation with the final tournament result, which suggests that the idea of transferring betting odds to ratings is indeed a highly promising approach.

More accurate team ratings could be a valuable tool in performance analysis in football, where the importance of situational variables is generally acknowledged (Fernandez-Navarro, Fradua, Zubillaga, & McRobert, 2018; Lago, 2009; Lago-Peñas, 2012; see also Article X & Article XI). To account for the quality of teams, however, no consensus on a standard method seems to exist. Standard approaches to operationalise team strength are based on differences in end-of-season rankings (Fernandez-Navarro et al., 2018; Lago, 2009), differences in the number of points (O'donoghue & Robinson, 2016) or the fact whether teams did or did not qualify for the knockout stage of a tournament (see Article X & Article XI). These measures should be assessed very critically, as all of them are actually measures of success, not measures of quality. Moreover, in such approaches the quality of teams is not estimated a priori, i.e. before the matches, but in fact partly determined by results of the matches analysed.

3.2.2. Answer to Research Question 2a)

The answer to research question 2a) is that a team rating making use of the information reflected in betting odds is a very accurate way to determine the individual strength of a football team over time. This conclusion is based on the investigation of a newly introduced model transferring the information in betting odds into a global rating of individual team strength. The novel rating has been shown to clearly outperform comparable ratings based on the results of football matches in terms of goals or wins.

As explained before, betting odds can be considered a useful measure of team strength, however, betting odds reflect the strength of a team in the specific context of a match. Therefore, they should be assessed as a measure of relative team strength providing information about the strength when controlling for the game location, the strength of the opponent and further match-specific factors. Depending on the application, it can be more relevant to have an accurate estimation of the global team strength over time.

The novel rating is based on the well-established ELO rating and the validation method is largely adopted from the ELO rating in football introduced by Hvattum and Arntzen (2010), that has been discussed in the previous section. In particular, two rating versions based on results and goals have been adopted and are denoted as ELO-Result and ELO-Goals. The novelty of the so-called ELO-Odds model presented in Article III is to transfer the information reflected in betting odds into such a rating. This is achieved by replacing the information about the actual result after a match with the information about the probability of winning derived from the betting odds prior to a match.

A dataset of almost 15.000 football matches from ten seasons in four European leagues as well as the UEFA Champions League and UEFA Europa League was used to validate the quality of the novel rating. Table 3 summarizes the main results of Article III by illustrating the predictive accuracy of each method. ELO-Odds is able to significantly outperform both ELO-Goals and ELO-Result, which underlines the value of the novel approach incorporating betting odds into ratings. Moreover, it is evidence that the information in betting odds is more valuable in determining the strength of a team than the actual outcome of a match. These results also support the notion that goals in football have a very limited informative value with regard to the quality of teams, which indicates a high influence of randomness on the process of goal scoring. Unsurprisingly, ELO-Odds falls short of the betting odds themselves in terms of forecasting accuracy. Certainly, it is unreasonable from the outset to expect to outperform the betting odds by solely using information from the betting odds. Additionally, both measures are not completely comparable in terms of information processing. While the ‘last’ information reflected in ELO-Odds is the market expectation prior to the prior match, the betting odds can additionally reflect all information becoming available during the last match (e.g. the result and the course of the game) as well as all information becoming available in the time between the last match and the currently forecasted match.

Table 3: Accuracy of three ELO ratings based on betting odds, goals and results as well as the betting odds themselves. The informational loss Li is used as measure of forecasting accuracy. P-values refer to paired t-tests comparing each model to the model in the subsequent column.

Forecasting Model	Average L_i	Standard deviation L_i	p-value
Betting Odds	1.380	0.674	< 0.0001
ELO-Odds	1.391	0.706	< 0.0001
ELO-Goals	1.401	0.714	0.0202
ELO-Result	1.403	0.715	-

While forecasting accuracy was used as a measure to validate the quality of the ratings as a strength estimation, the purpose of introducing ELO-Odds is not purely focused on forecasting. On the contrary, if it were about forecasting the next game, the betting odds themselves would be the easier and more accurate choice. The practical value of ELO-Odds lies in its ability to harness the information in betting odds to generate a highly accurate estimation of team strength over time. Figure 4 is intended to illustrate this idea by presenting ELO-Odds and ELO-Result for the team Borussia Dortmund during seasons 2013/14 and 2014/15.

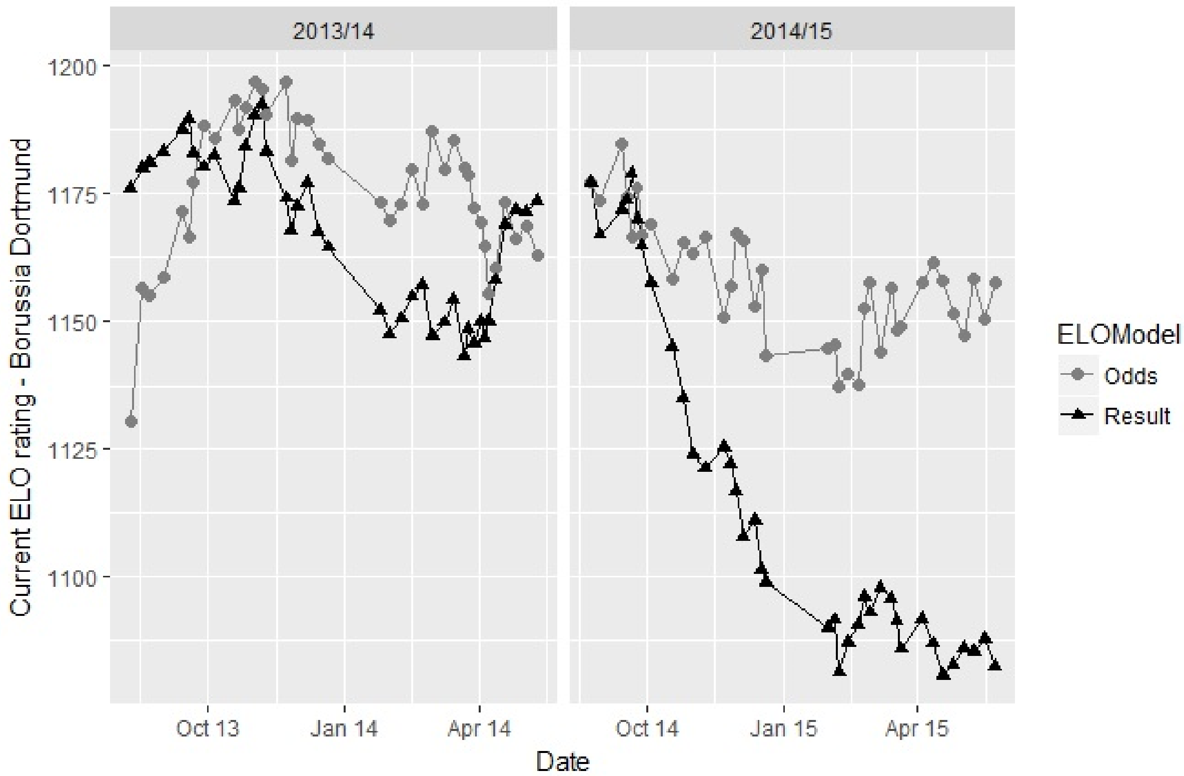

Figure 4: The evolvement of ELO-Odds (circles) and ELO-Result (triangles) for the German football team Borussia Dortmund during seasons 2013/14 and 2014/15.

The reasons why this example (i.e. this team and this period) was chosen and the interpretation on why the ratings reacted in precisely this way are explained in detail in Article III. At this point, it should primarily be illustrated that the ratings allow valuable conclusions to be drawn about the medium and long-term strength development of a team. This can be a valuable indicator to evaluate the performance of coaches or club officials. The figure illustrates how strongly the assessment of the team strength depends on whether it is based on results or betting odds. As the extensive analysis of the ratings has shown, ELO-Odds is clearly preferable to ELO-Rating. In this respect, ELO-Odds – if used for decision making – could prevent club officials from being subject to outcome biases (cf. Brechot & Flepp, 2020; Gauriot & Page, 2019).

3.3. Influence of team strength on success in penalty shootouts

3.3.1. Previous research

Penalty kicks in football are naturally associated with a very high chance of scoring a goal (Dalton, Guillon, & Naroo, 2015). Moreover, the penalty shootout is used as a method to decide drawn matches in tournaments including World Cups, continental championships or domestic cup competitions. Thus, the importance of penalty kicks can be cited as a central reason for their popularity in scientific research. The other central reason is arguably the simplicity and standardised character of the penalty situation consisting solely of the penalty-taker, the goalkeeper, one shot and a binary result. Consequently, the penalty allows for simple mathematical modelling of strategies, straightforward statistical analysis of empirical results and even the possibility for experimental investigation with reasonable effort (for an overview on penalty research see Dicks, Uehara, & Lima, 2011; Memmert, Hüttermann, Hagemann, Loffing, & Strauss, 2013). The present section is limited to research with regard to the specific aspects of penalties that are addressed in Article IV.

The first and central aspect is the influence of team strength on penalty success. It is conceivable to assume that - apart from chance - psychological, technical and physiological aspects could affect penalty success (Jordet, Hartman, Visscher, & Lemmink, 2007). The search for success-enhancing strategies applicable by penalty-takers (Jordet, Hartman, & Sigmundstad, 2009; Noël, Furley, van der Kamp, Dicks, & Memmert, 2015) and goalkeepers (Furley, Noël, & Memmert, 2017; Savelsbergh, van der Kamp, Williams, & Ward, 2005) implicitly assumes that elements of task-specific penalty skill exist, but does not answer the question, whether a generally high skill level of players contributes to penalty success. Attempts to establish a link between general player skill and success are complicated by the difficulty to find accurate measures to operationalise the skill of a player. Jordet et al. (2007), for example, used position and age as a proxy for player skill and experience, finding no significant relationship, but a tendency of forwards and young players to be slightly more successful penalty takers.

While not being applicable to individual players, the use of betting odds as a measure for general team strength opens up the possibility to gain a better understanding on the relationship between skill and success in penalty shootouts. This analysis is accompanied by the introduction of a model to forecast success in penalty shootouts. Despite the broad literature on forecasting in sports and penalty shots, no existing study has focused on this aspect so far, to the best of the authors knowledge.

The second major aspect is home advantage, i.e. the question whether the game location influences the results of penalty shootouts. From a theoretical standpoint, it is not unreasonable to suppose that the well-documented and robust home advantage (Courneya & Carron, 1992; Nevill & Holder, 1999; Pollard & Gomez, 2014; Pollard & Pollard, 2005) transfers to penalty situations. At the same time, it is possible that the home team is subject to a disadvantage caused by mechanisms such as choking under pressure (cf. Dohmen, 2008). There is clear evidence that more penalties are awarded to home teams (Sutter & Kocher, 2004) and home teams score more penalty goals (Boyko, Boyko, & Boyko, 2007; Nevill, Newell, & Gale, 1996). However, this can be explained by tactical behaviour, fouling behaviour or referee bias and does not imply any information about a home advantage in the relative conversion rate of penalties. Kocher, Lenz, and Sutter (2008) did not find evidence for such a home advantage in penalty shootouts. The analysis, however, was highly limited in terms of sample size and competitions investigated (less than 100 shootouts from a single competition, the German DFB-Pokal).

3.3.2. Answer to Research Question 2b)

With regard to research question 2b), the results of Article IV indicate that the team strength influences success in penalty shootouts in a way that the team with a higher general team strength is also more likely to win. The effect, however, is relatively small, as even very weak teams have a roughly 40% chance of winning a shootout against a very strong team. Besides the influence of strength, Article IV also shows that there is no evidence for any effect of the game location on penalty success and demonstrates the usefulness of a model forecasting the success in penalty shootouts.

The conclusions are based on an analysis of 1,067 penalty shootouts from 14 seasons across 12 national and two international cup competitions. The database thus can be considered to be large enough to obtain significant results and diverse enough to obtain results generalisable to high-class football. Only including the winner of the penalty shootout, the dataset does not possess a particularly high granularity, which means that no individual shots or the contributions of penalty-taker and goalkeeper were analysed.

The benefit of using betting odds as a measure of team strength has already been explained at the beginning of the chapter. In the case of cup competitions, using this method is even more important, as alternatives based on table ranks or points are not applicable for matches with teams from different divisions or domestic leagues. The results of penalty shootouts are analysed on a match-level (i.e. shootout-level) and as such the analysis is based on the relative team strength of teams compared to each other. Betting odds for the normal match time thus can be directly used and do not need to be transferred to global strengths as in Article III. However, besides the relative team strengths, the betting odds reflect the home advantage. Article IV accounts for this by adjusting the strength difference by excluding the home advantage in order to exclusively account for the inherent playing strength of the teams.

The team strength obtained from betting odds was used to identify the relatively stronger and the relatively weaker team for every match. Moreover, teams were divided into home and away teams. Results show that stronger teams win more penalty shootouts (53.6%) than weaker teams (46.4%), which is a significant difference ($p < .05$) with reference to a two-sided binomial test. There was no evidence for an influence of the game location on the result of penalty shootouts, neither if considering a reduced dataset due to possible rule effects (home: 49.5%, away: 50.5%, $p = .944$)., nor if considering all matches (home: 48.6%, away: 51.4%, $p = .391$). By using a logistic regression with the binary shootout result as a dependent variable and the difference in team strengths as an independent variable, it was possible to derive a formula directly linking the team strengths to the probability of success, which is illustrated in Figure 5. In addition, the value of forecasts based on a logistic regression was tested in an out-of-sample approach and the superiority to a benchmark model was demonstrated ($p < .05$ based on a Wilxocon test of Brier scores).

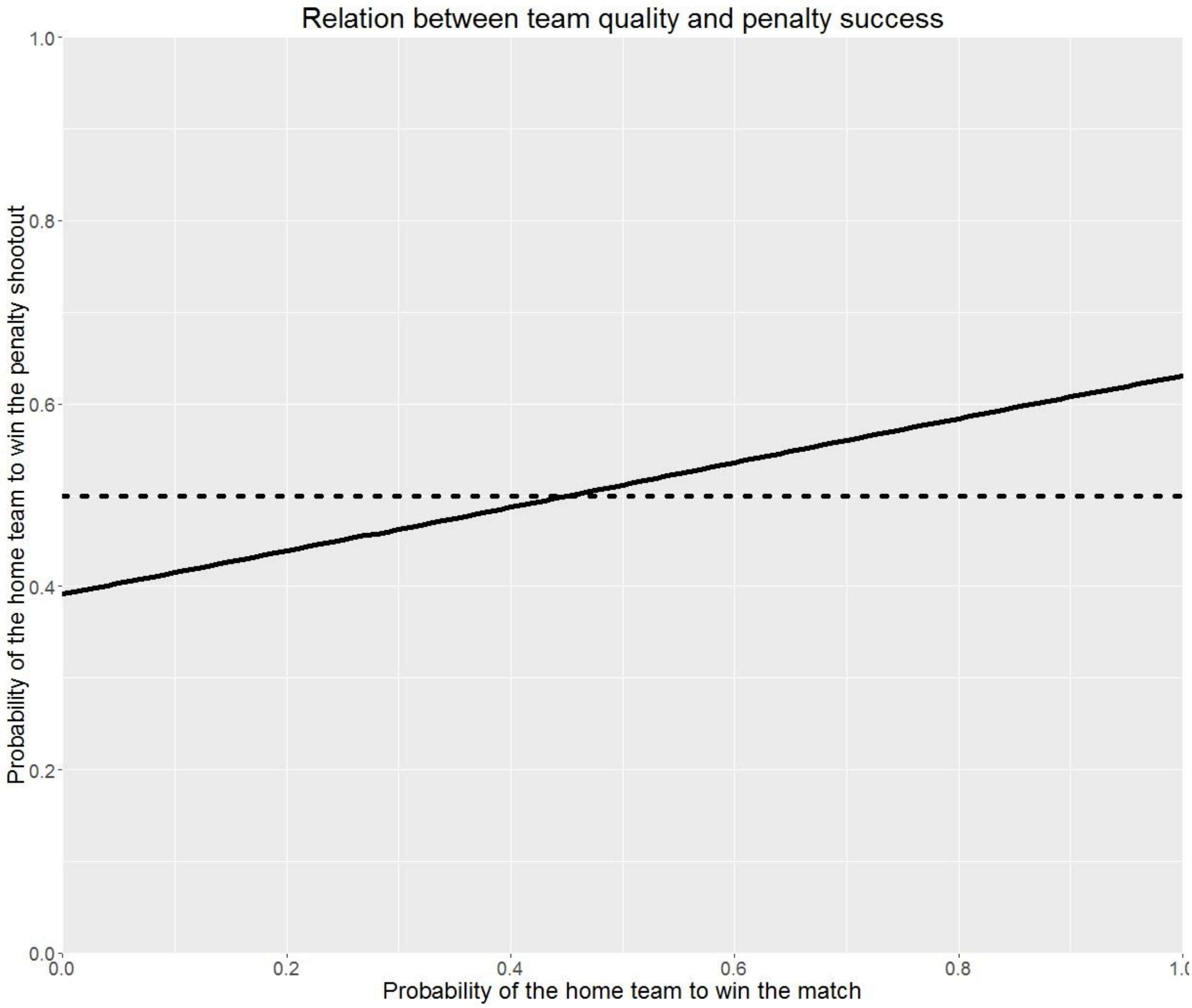

Figure 5: The relation between team strength and penalty success given by the probability of winning a penalty shootout as a function of the probability to win the match in regular time.

In summary, Article IV – for the first time – answered the question whether the general strength of football teams translates significantly to penalty shootouts. Visually, Figure 5 indicates that this is the case. This observation is statistically substantiated by the fact that significantly more shootouts are won by the stronger team and that the difference in team strength is a significant factor in a logistic regression model. As an additional result, this gives insights to the question whether skill or luck is the decisive factor in penalties and whether it is justified to denote penalty shootouts as ‘lottery’, which is often done by the media. With respect to the general team strength, penalty shootouts cannot be considered a pure lottery, but it needs to be acknowledged that team strength only translates loosely to success as a very weak team still has a roughly 40% chance of winning a shootout against a very strong team.

More research is certainly needed to better understand how the mechanisms of general playing strength and penalty success work. In particular, it is interesting whether the goalkeeper, the penalty-taker or both contribute to the enhanced success for strong teams. Moreover, it is worth investigating whether this effect is also true for penalties taken within the match and whether the inferiority of stronger teams can be attributed to technical, physiological or psychological skills.

No evidence for a home advantage or disadvantage in penalty shootouts was found, which can be considered surprising in light of the robust existence of home advantage in football matches in regular playing time. This result is particularly interesting in view of the fact that there is still no consensus on the factors underlying the home advantage, which will be discussed in more detail in Article V. It is not directly evident, why mechanisms leading to the home advantage should be present in regular time, but not in a penalty shootout. Theoretically, an important difference could be referees, that have been shown to be subject to biased behaviour in deciding on disciplinary sanctions (Boyko et al., 2007; Goumas, 2014a; Sutter & Kocher, 2004). In penalty shootouts, however, it can be assumed that outcomes are not or only very marginally influenced by referee decisions. Other factors, that could play a role are psychological factors, including the presence of spectators that might have negative effects on the home team (cf. choking under pressure, Dohmen, 2008).

3.4. The effect of spectator absence on the home advantage

3.4.1. Previous research

The term home advantage or home field advantage refers to the socio-psychological phenomenon that sports teams or athletes tend to be more successful at home than away. It can be argued that the existence of the home advantage is one of the most robust and well-studied phenomena in sports (Courneya & Carron, 1992; Jamieson, 2010; Jones, 2013; Nevill & Holder, 1999; Pollard & Pollard, 2005) with plenty of evidence from the domain of football (Goumas, 2014b; Pollard, 1986; Pollard & Gomez, 2014). Researchers have tried to identify reasons for its existence and several factors that have been hypothesized to contribute to the home advantage are summarized in the literature. The most commonly discussed ones include the familiarity of the home team with its own sports facility, fatigue caused by the travel burden of the away team, effects of the crowd supporting the home team, an unequal treatment of home and away teams due to biased referee behaviour, an increased hormonal reaction of home players who want to defend the own territory denoted as territoriality, specific rules giving a systematic advantage to home teams, as well as effects of different psychological expectations or tactical behaviour in home and away matches (cf. Courneya & Carron, 1992; Neave & Wolfson, 2003; Nevill & Holder, 1999; Pollard, 2008).

The fact that the interest in home advantage research has remained over decades, is a clear indication that the identification of reasons is difficult and that still no final consensus exists. The present section is focused only on those two factors that are clearly associated with the presence of spectators, namely crowd support and referee biases.

Crowd support refers to the idea that the behaviour of spectators can directly contribute to the home advantage by motivating the home team (e.g. through cheering and singing) or intimidating the away team (e.g. through booing and insulting). Fans, but also players and referees perceive crowd support as an important factor for the home advantage (Anderson, Wolfson, Neave, & Moss, 2012; Wolfson, Wakelin, & Lewis, 2005). Empirical results, however, fail to reveal clear evidence for the notion that crowd support is a major contributor to the home advantage. First, investigations of indicators such as crowd size or crowd density show mixed results. Comparisons across divisions or countries have demonstrated that the home advantage is increased in higher divisions (Nevill et al., 1996) and countries with stronger leagues (Pollard & Gomez, 2014), which is assumed to be related to the higher mean attendance. Analysis on a match level revealed that crowd size, but not crowd density increased home advantage in the English Premier League (Boyko et al., 2007). Similar results were found for the Australian A-League, however, crowd size was only related to home advantage up to 20,000 spectators (Goumas, 2014b). Second, negative consequences of spectators on home team performance such as the so-called choking under pressure phenomenon are also plausible and have been found in sports competitions (Baumeister & Steinhilber, 1984; Dohmen, 2008). Third, and most importantly, studies indicating an association of crowd size and home advantage, cannot be considered direct evidence for crowd support, as it is not clear whether the differences are really caused by the players reaction to the crowd or whether it is rather the referees that react to the crowd.

It would be a plausible explanation, that referees exhibit a biased behaviour and unconsciously give an advantage to the home teams, which is induced by the social pressure of spectators in the stadium. In fact, there is robust evidence from real-world football matches, showing that away teams receive more yellow and red cards (Buraimo, Forrest, & Simmons, 2010; Goumas, 2014a; Nevill et al., 1996), while more penalties are awarded to the home teams (Nevill et al., 1996; Sutter & Kocher, 2004). Moreover, referees are subject to a biased behaviour when deciding on the extra time at the end of a match (Garicano, Palacios-Huerta, & Prendergast, 2005; Riedl, Strauss, Heuer, & Rubner, 2015; Scoppa, 2008; Sutter & Kocher, 2004).

Researchers have also tackled the question, whether these results are attributable to a spectator-induced referee bias, or rather a reflection of different tactical or fouling behaviour of the teams. Results support the notion of biased referee behaviour, as unequal treatment is found even if controlling for possibly confounding variables such as scoreline and team strength (Buraimo et al., 2010), referee decisions were found to depend on crowd noise in an experimental study (Unkelbach & Memmert, 2010) and there are first indications that referee bias disappears in matches without spectators (Pettersson-Lidbom & Priks, 2010).

Despite these results, it remains difficult to directly assess the impact of spectators on home advantage and referee bias. Natural experiments, where some professional football matches are played without spectators for specific reasons, are required to accurately estimate the effect of spectator presence or absence. Pettersson-Lidbom and Priks (2010) make use of such a natural experiment evoked by spectator exclusion due to hooligan violence in Italian football. While they find evidence for a disappearing referee bias in such matches, they do not quantify the contribution to the home advantage and the sample size of only 24 matches without spectators can be considered highly limited. Making use of matches played without spectators due to the COVID-19 pandemic in 2020, Article V will tackle the question of how spectator absence influences the processes within a football match and how these changes are reflected in the home advantage.

3.4.2. Answer to Research Question 2c)

The short answer to research question 2c) is that spectator absence has major influence on the processes in professional football matches, while a substantial degree of home advantage remains in empty football stadiums. This conclusion is based on a statistical analysis of eight measures of disciplinary sanctions, match dominance, market expectation and home advantage from more than 35.000 professional matches and two measures of home advantage in more than 5.000 amateur matches.

Article V benefits from an unprecedented situation caused by the COVID-19 pandemic in 2020, that forced professional football leagues to terminate seasons or continue seasons in absence of spectators. Thus, is was possible to include more than 1.000 professional matches in total absence of spectators to the study, that can be considered a large-scaled natural experiment. The effects of spectator absence on the differences between home and away teams with regard to these eight measures were investigated using linear mixed effects regression models with two time horizons and controlling for season and league effects.

With regard to the notion of a spectator-induced referee bias, results revealed that away teams receive more disciplinary sanctions in matches with spectators, but this effect disappears or is even slightly reversed in absence of spectators (fouls: $p < .001$; yellow cards: $p < .001$; red cards: $p < .05$). While this is generally consistent with the literature on referee bias presented in the previous subchapter, the clear connection between spectator presence and referee bias has not been demonstrated that clearly in a natural experiment with a comparable database before. Analysis of match dominance revealed that the absence of spectators is clearly reflected in the number of shots and shots on goal. The difference for both measures is roughly halved, which confirms a highly significant influence of spectators on the match dominance (shots: $p < .001$; shots on target: $p < .01$).

The results with regard to the influence of the spectators on the actual home advantage in terms of points and goals can be considered less conclusive. If just considering matches from the full 2019/20 season, the home advantage is non-significantly reduced by about one sixth (goals: $p = .45$; points: $p = .50$). If including matches from the full last 10 seasons, it is reduced by one third, which is closely failing to reach significance at a five percent level (goals: $p = .06$; points: $p = .07$). The different results can be attributed to the fact that the home advantage was already found to be reduced at the begin of the 2019/20 season, even before the pandemic. This drop between seasons can be considered surprising and is not fully explainable by the general tendency of a slightly decreasing home advantage over the last years.

In consistency with the idea of this chapter, the way in which betting odds can contribute to better interpret these results, should be underlined. As has been illustrated in Article I, useful forecasting models need to reflect global systematic aspects such as the home advantage. Consequently, bookmakers and bettors need to react to the information of spectator exclusion by adjusting their expectation on the home advantage and this will be reflected in the betting odds. Thus, a measure of the market expectation on the home advantage, can be obtained by transferring odds into forecasted probabilities and calculating the difference between expected points for the home and the away team. Figure 6 illustrates the difference in expected points as well as the difference in actually observed points on a monthly basis over ten seasons.

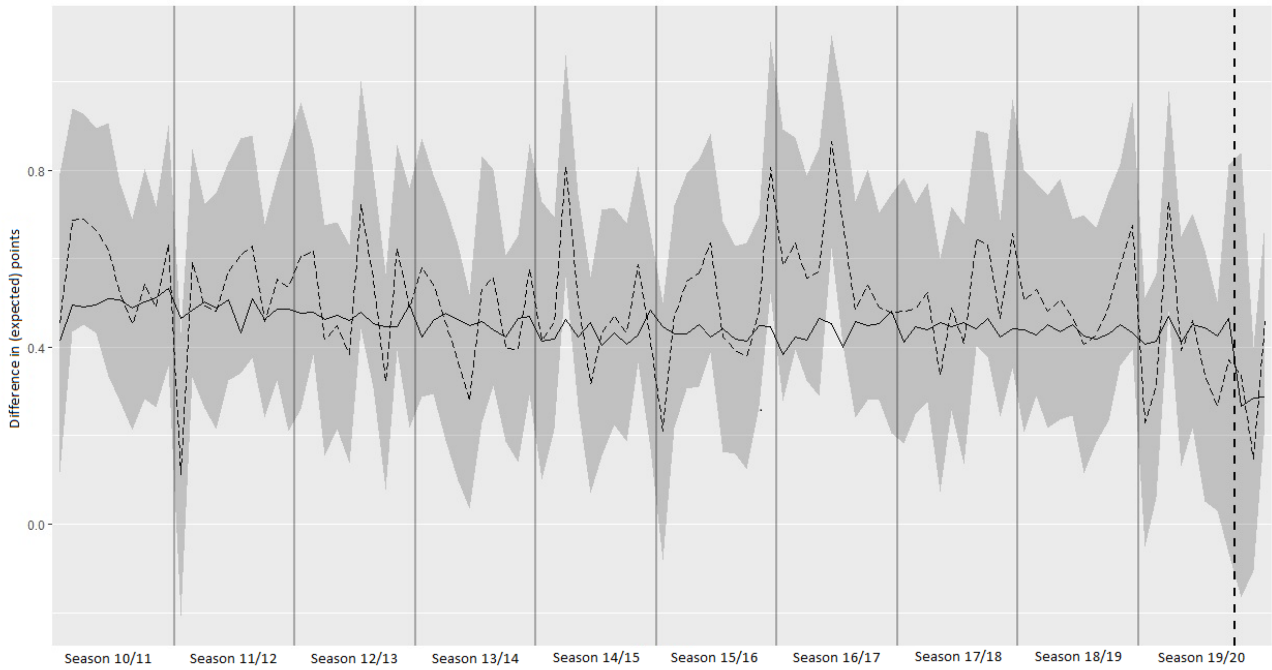

Figure 6: Home advantage as an average difference in points (dashed line) and market expectation on home advantage as an average difference in expected points (solid line) over ten seasons. The grey area refers to 95% confidence intervals for the difference in points. The dashed vertical line refers to the COVID-19 based interruption of the 19/20 season.

As described at the beginning of this chapter, betting odds have a very high ability to reflect actual probabilities and as such can be assumed to also reflect unusual information like spectator exclusion accurately. The advantage of betting odds is that these are not subject to the large random variation that is inherent in the outcomes (i.e. points) of football matches. From Figure 6 this becomes apparent as the actual home advantage in terms of points fluctuates strongly while the expected home advantage shows a pretty stable development.

The advantage of betting odds is that the market should react to systematic changes in football, but not to random influences. This helps to shed light on the market view on two important aspects of the analysis. First, did the drop at the begin of season 2019/20 have systematic or random reasons? Second, what was the actual effect of spectator absence on the home advantage? From Figure 6 and the results of the regression models, it can be concluded that the market estimation of the home advantage did not change as a reaction to the drop at the begin of season 2019/20, but reduced by about one third as a consequence of the spectator exclusion. In contrast to the home advantage in terms of points, the effect of spectators on the market expectation of the home advantage is highly significant (p < .001), which is a consequence of the lower variation in betting odds. Although the betting odds only constitute an estimation of the market, these results are an indication that the home advantage actually dropped systematically by one third due to the spectator exclusion.

Given the relevance of home advantage research and the unprecedented database, becoming available during the COVID-19 pandemic, this research question attracted the interest of several researchers. Consequently, a large number of highly related studies were published during submission of Article V and submission of this dissertation. None of the results are in full contradiction to the results of Article V, however the conclusions across studies can depend drastically on the choice of leagues, seasons and the statistical method used. Results with regard to a disappearing referee bias are consistent (Scoppa, 2021; Sors, Grassi, Agostini, & Murgia, 2020; Tilp & Thaller, 2020) and in line with Article V. Results on the home advantage, however, are mixed and include no effect of spectators on the home advantage reported for the Portuguese Primeira Liga (Matos et al., 2021), halved home advantage during the pandemic reported across five European countries (Scoppa, 2021) or even a home disadvantage in matches without spectators reported for the German Bundesliga (Tilp & Thaller, 2020). Given the large database and the statistical model chosen in Article V, it can be argued that other studies largely overstate the effect of spectators on home advantage due to the choice of leagues (Tilp & Thaller, 2020), choice of seasons (Sors et al., 2020) or models not adequately controlling for long-term decreasing effects of the home advantage (Scoppa, 2021). At this time, more data already got available and will be reflected in further studies soon. Once spectators return into the stadium, the effects of spectators on home advantage will be quantifiable with an even higher level of certainty.

3.5. The extent of random influence on goal scoring

3.5.1. Previous research

Success in football is exclusively defined by the number of goals scored and conceded. Although being rare events in football matches, goals thus are a naturally central object of investigation in performance analysis in football. Researchers have therefore extensively studied how, when and why goals are scored in football (for an overview see Pratas, Volossovitch, & Carita, 2018). Some aspects can be considered to have a descriptive character such as the number of goals in various match periods that show an increasing tendency over the course of matches (Alberti, Iaia, Arcelli, Cavaggioni, & Rampinini, 2013). Others are clearly connected to the ideas of identifying success-enhancing strategies. For example, the relatively large number of goals scored following set plays (Armatas & Yiannakos, 2010), the clear majority of goals scored from within the penalty area (Michailidis, Michailidis, & Primpa, 2013), the increased scoring frequency if the shooting opportunity occurred behind the defensive line (Gonzalez-Rodenas, Lopez-Bondia, Calabuig, Pérez-Turpin, & Aranda, 2017), or the effectiveness of different offensive tactical approaches (Tenga, Ronglan, & Bahr, 2010).

While modelling of random influences plays an important role in sports forecasting (see Article I), performance analysis in general focuses on systematic aspects of success rather than on random aspects of success. On the one hand, this is reasonable, as coaches and players can work on systematically improving skills, but cannot train to have more luck. On the other hand, it can be argued that a very important contribution to success thus remains largely unexplored. After all, the notion that randomness is involved in football can hardly be denied. Hill (1974) even stated to "find it difficult to imagine that anyone, who had ever watched a football match, could reach the conclusion that the game was either all skill or all chance" (p. 203), which is in line with the idea of having both systematic and unsystematic influences as described in Article I. Heuer et al. (2010), taking a strictly statistical look at football matches, even concluded that "scoring goals is a highly random process" (p. 4). Such conclusions, however, are predominantly based on statistical analyses of match results, while not considering the actual sequence of events on the pitch. Consequently, there is a lack of studies that focus on random contributions being directly visible and quantifiable as uncontrollable events on the pitch.

Performance analysis does partly account for the inherent randomness involved in goal scoring, for example by considering score box possessions instead of goals (Sgrò, Aiello, Casella, &

Lipoma, 2017; Tenga, Holme, Ronglan, & Bahr, 2010; Tenga, Ronglan, & Bahr, 2010) or by identifying other key performance indicators that are linked to success, but based on more frequently occurring events than goals (Brechot & Flepp, 2020; Castellano, Casamichana, & Lago, 2012; Rein et al., 2017). However, very little research has been conducted so far, classifying and analysing random events on the pitch. Besides the above-mentioned bias in favour of systematic effects, this might also be complicated by the difficulty to define objective criteria of randomness. An example is the work of Gauriot and Page (2019) who investigated shots touching the goal posts and going in, in comparison to shots touching the goal post and not going in. These situations either end with a goal or with no reward at all, although being caused by an almost equally accurate shot. The authors thus consider the actual outcome of the event to be mainly driven by luck. Using real-world data from top-class European leagues, they were able to find evidence that luck in football is overrewarded by coaches and journalists.

The only study that, to the best of the authors knowledge, has pursued a direct approach to define random contributions to goals in football, is the study of Lames (2018), who defined six variables of random influence. Article VI can be considered a replication and extension of the above-mentioned study.

3.5.2. Answer to Research Question 2d)

With regard to research question 2d), Article VI demonstrates that random processes have a substantial influence on goal scoring in football. Moreover, it can be stated that the degree of randomness has decreased over seasons, while random influences are more pronounced for weaker teams and in situations where the current scoreline is a draw.

These conclusions are based on an analysis of 2,660 matches from seven full seasons of the English Premier League, including a total of 7,263 goals. The video material for each goal was analysed and manually annotated by one of seven observers. The observers were asked to evaluate the occurrence of six different variables of random influence for each of the goals. Figure 7 shows the prevalence of goals that were subject to at least one variable of random influence, as well as the prevalence of goals fulfilling each of the variables individually.

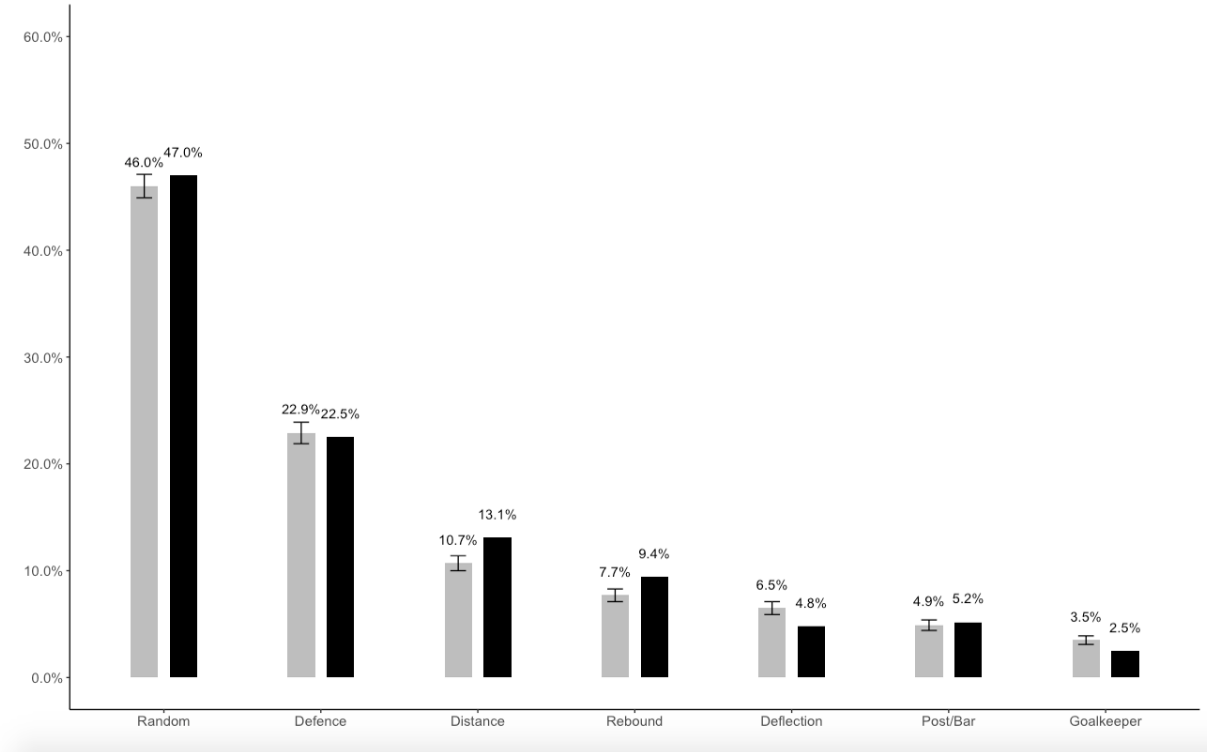

Figure 7: Prevalence of goals with random influences. The bars refer to the proportion of goals, that were subject to six different variables of random influence. The additional variable Random refers to the proportion of goals that were subject to at least one of these six variables. Grey bars refer to Article VI, while black bars refer to the results of Lames (2018). Error indicators refer to 95% confidence intervals.

Despite minor deviations, the results of Lames (2018) were successfully replicated, which includes the generally high prevalence of random goals in football. The fact that at least one random influence was identified for roughly every second goal, suggests that randomness is a key aspect in the realisation of goals and consequently results in football.

Moreover, Article VI helps to gain a better understanding of match situations, in which random goals are more likely to occur. In line with the general approach of this dissertation, the role of betting odds as a key element to investigate differences between stronger and weaker teams shall be discussed in particular. On the basis of betting odds for each match, it is possible to decide which team can be considered the stronger and which team can be considered the weaker team. This information was used in a logistic regression model (additionally considering a total of eight further situational variables) in order to decide whether there is a meaningful difference between the proportion of random goals scored by stronger and weaker teams respectively. The logistic regression suggests that random goals are more pronounced for weaker teams ($p < .05$), who are subject to random influence for 47.6% of goals, compared to 44.8% for stronger teams. If teams are divided in four equally sized groups of strength levels, the pattern that stronger teams are associated with a smaller proportion of random goals, is confirmed, as illustrated in Figure 8.

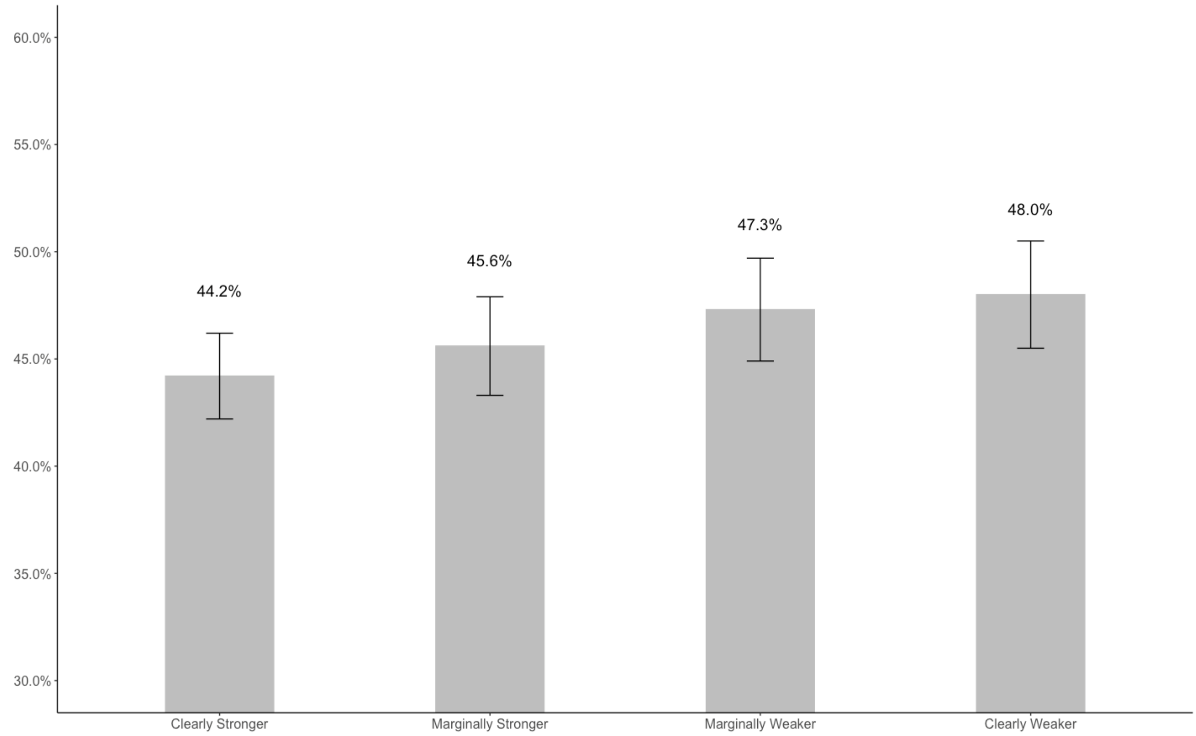

Figure 8: Prevalence of random goals dependent on four different strength levels with regard to the difference in relative team strengths as measured by betting odds.

As Article VI shows, the use of betting odds as an accurate measure of strength in a specific match, can help to directly analyse differences between stronger and weaker football teams. The pattern of goal scoring is only one example for this, but the general approach can be used for analysing any aspect of team behaviour and performance on the pitch.

Due to the large database and the fact that datasets for several human annotators were randomised across seasons, it was moreover possible to investigate timely trends in random contributions to goals for the first time. Results reveal that the degree of randomness in goal scoring decreased over seven seasons ($p < .001$). Furthermore, random goals occurred more frequently if the current scoreline of the match was a draw than when one team was leading ($p < .05$). and were dependent on the match situation, where penalties are almost unaffected by random influences while goals following a freekick, a corner or scored from open play were all subject to substantial random influence.

No significant influence on the proportion of random goals could be found for the time period in the match, the goal number, the matchday, the match location as well as between balanced and unbalanced matches in terms of team strength.

4. Aspects of data analysis and computer science

4.1. Big Data

Research, business and society have been subject to substantial changes in recent years, driven by the increased volume and complexity of available data. More data promise more insights, more economic advantages and more socially relevant improvements, while it poses enormous challenges to efficiently store, access and analyse large datasets. The term Big Data has been established to describe this development as well as the associated opportunities and challenges (Ekbia et al., 2015). As the term seems to be used inflatedly and inconsistently in public debate, researchers have tried to find and establish a consensual definition of what Big Data actually is (Alwan & Ku-Mahamud, 2020; Mauro et al., 2016; Sagiroglu & Sinanc, 2013).

While the word ‘big’ suggests that it is the sheer amount that qualifies data as Big Data, there are actually different aspects of complexity, which lay the foundation for Big Data. In defining what qualifies data as Big Data, the use of different Vs has been established. These are Volume, Velocity and Variety (Mauro et al., 2016; Sagiroglu & Sinanc, 2013), partly accompanied by Veracity (Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012) and Value (Alwan & Ku-Mahamud, 2020). With regard to the aforementioned articles, these Vs are attributed to the following characteristics of data: Volume refers to the quantity of the data, that is associated with Big Data if the amount of available data is large enough to impose the need for specific technology and methodology to store and analyse it. Velocity describes the speed at which new data becomes available, referring to the issue of data becoming outdated quickly and a resulting pressure for solutions to analyse data in real-time. Variety refers to the different types of data (with a structured, semi structured and unstructured character) stemming from potentially different data sources, that are included in the datasets and add complexity to the analysis of data. Veracity points to the limited reliability of data, potentially including inaccurate and missing values as well as inherent unpredictability, which needs to be considered and handled efficiently. Value refers to the question whether data is important and includes information that can be actually transferred into value, which, however, can only conclusively be assessed after analysing the data.

Within this dissertation, the consideration of Big Data is delimited to the analysis and extraction of insights from complex data by adapting and adjusting useful methods, while no technological advances or inventions of completely novel methods are aspired. Despite the applications of Big Data growing in many research fields, the identification and conception of the main challenges emerge from the domain of computer science, including methods and strategies to manage, store and analyse massive datasets. In terms of theory, methodological advances in this field can be driven by mathematics and algorithms, while applications to real-world data, like in the present dissertation, are strongly application-based. The analysis of Big Data is therefore strongly connected to the fields of Data Mining (Raval, 2012) and Machine Learning (Dey, 2016).

4.2. (Big) Data in Sports

Technological development and digitalisation have also had a major impact on sports, resulting in an increased availability of data for analysis. Applications of complex data analysis in sports are therefore more and more associated with the field of Big Data (Morgulev, Azar, & Lidor, 2018; Rein & Memmert, 2016). Investigation of sports data increasingly requires interdisciplinary approaches, combining knowledge from the sport scientific domain with expertise in information technology with a focus on aspects of data mining (Ofoghi, Zeleznikow, MacMahon, & Raab, 2013), artificial intelligence and machine learning (Beal, Norman, & Ramchurn, 2019).

Research related to the collection, processing and analysis of large and complex datasets in sports, often referred to as sports analytics (Morgulev et al., 2018), has developed into a flourishing field of research. This development is evidenced by a decent number of recent special issues, e.g. on sports analytics in *DataMining and Knowledge Discovery* (Brefeld & Zimmermann, 2017), on sports analytics in *Big Data* (Assunção & Pelechrinis, 2018), on machine learning for soccer in *Machine Learning* (Berrar, Lopes, Davis, & Dubitzky, 2019), on computational intelligence and data mining in sports in *Applied Sciences* (Fister, 2021), as well as the well-known MIT Sloan Sports Analytics Conference, which provide a wide collection of articles and a good insight into the state of art in sports-related (Big) Data analysis.

Within this dissertation two sources of Big Data related to football matches are tackled. First, textual data from messages posted at the microblogging-platform Twitter over the course of football matches. Second, event and positional data from football matches being a detailed representation of all events as well as all movements of the players and the ball on the pitch.

4.3. Sports-related Twitter communication as a source of Big Data

It is not only data directly related to the performance of athletes at sporting events that has become increasingly available through technological progress. Digitisation has also led to increased availability of data reflecting communication about sporting events. Social media has made it possible for everyone to share opinions and reactions online in real-time. Twitter (Twitter, 2021) is a microblogging-platform that enables users to write and publish short textual messages, so-called tweets. Tweets can be tagged with special terms, so-called hashtags, that specify the topic and thus make it possible to easily find a large number of tweets concerned with a certain topic. Tweets are not only publicly available on the website of Twitter, but can also be obtained for analysis from a specific application programming interface (Twitter API, 2021).

Thus, Twitter communication has become a well-studied and highly discussed subject in various domains (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011; Bruns & Stieglitz, 2013; Gayo-Avello, 2013; Huberty, 2015; Zhang, Fuehres, & Gloor, 2011), including sports-related topics (Sanderson, 2014; Schumaker, Jarmoszko, & Labedz, 2016; Sheffer & Schultz, 2010; Witkemper, Lim, & Waldburger, 2012).

Twitter data undoubtedly qualifies as a source of Big Data, which is already evident by the sheer volume and velocity of data. On a regular day more than 500 million tweets are reported to be written on Twitter (Twitter Blog, 2013), while the World Cup 2014 as a monthlong competition individually triggered more than 672 million tweets (Twitter Blog, 2014) and even a single event, namely the loss of England to Iceland in the European Cup 2016, can trigger as many as 128,000 tweets per minute (Twitter Blog, 2016). On the basis of own data collection, it can be stated that textual and metadata on tweets and retweets related to one single hashtag with regard to one important match (#UELfinal, UEFA Europa League final 2019) can sum up to more than a gigabyte of raw data. Moreover, Twitter fulfils the criterion of variety, as it includes structured metadata (e.g. time or language of a tweet), but mainly textual data that can be considered highly unstructured and require specific algorithmic solutions to extract information (see Article VII). Being user-generated content including the inherent inaccuracies of textual human communication, Twitter data can also be considered to be subject to the issue of veracity.

Being a source of collaborative interaction, Twitter data – in the context of this dissertation – also raises theoretical questions with regard to crowd wisdom, that will be discussed with regard to Article VIII.

4.4. Event and positional data as a source of Big Data

Event and positional data represent a highly detailed representation of the events and movements taking place on the pitch over the course of a football match. As such, it can be argued that the full information complexity of a football match is already reduced and the data structured. This process itself is highly complex and while attempts for automatic extraction of information from video material exist (Ekin, Tekalp, & Mehrotra, 2003), manual annotation or at least manual quality control remains widely used (Pappalardo et al., 2019).

Event data represents an ordered list of all the events taking place in a football match. Events are enriched with information on the actors or players involved in the action as well as the location in the pitch. The number of events occurring in a football match can vary from 1,500 to 3,000 thus including the number of features per event may generate extensive list of information per match.

Positional data represents the raw trajectories of all the actors in a football match. It includes the dynamic x- and y-coordinates of 22 players and the ball, which (supposing a common sampling rate of 25 Hz) results in more than 6 million datapoints for a single match of 90 minutes. Rein and Memmert (2016) state that “storing position, event and video data from a single complete Bundesliga season results in 400 gigabytes of tracking data” (p. 7), which represents a remarkably high volume of data, while not being fully comparable to the order of magnitude of petabytes, exabytes or zettabytes connected to Big Data (Alwan & Ku-Mahamud, 2020).

Velocity, i.e. the fast renewal of data gains importance if focusing on in-match application such as in Article VIII & IX. The renewal of data is obviously limited to those time periods, where relevant matches take place. However, if aiming at practical in-play usage of event and positional data, data insights might be needed at low latency or the data captured a few minutes ago might already be useless if the game situation changed.

In terms of variety, event and positional data already represent two different sources of information, which, moreover, may not be linked. The systematic use of data is complicated by different data sources including commercial data provider regularly gathering data from professional football matches (e.g. VISTRACK, see Lorenzo-Martínez, Rein, Garnica-Caparrós, Memmert, & Rey, 2020) or tracking systems specifically used to collect experimental data for scientific purposes (e.g. KINEXON, see Memmert, Raabe, Schwab, & Rein, 2019), which implies different data coverage, specifications and accuracy. Although the trajectory data are stored in a highly structured way, the sport context gives the data source a highly unstructured character. Same actions in theory can happen in different time spans or players with the same role might behave completely different in the field. In particular, a lot of context is only implicitly derivable from the data, if at all. This commonly includes the information, whether the current situation reflects a corner, freekick or open play; whether an offensive action reflects a controlled build-up or counterattack; or what is the reason for a current interruption of a match.

In terms of veracity, several issues need to be considered, which includes missing information, irrelevant information, inefficient data specification or inconsistent data specification across different providers with regard to event data (cf. Decroos, Bransen, van Haaren, & Davis, 2019). In positional data, a missing z-coordinate of the ball can limit the possibility for automatic detection of events such as passes, crosses or shots. In general, the accuracy of positional data is imperfect and depends on the method, where GPS-based and video-based systems seem to be less accurate in determining the x and y coordinates than radar-based systems (Linke, Link, & Lames, 2018; Siegle, Stevens, & Lames, 2013).

In summary, event and in particular positional data are accompanied by several aspects of complexity and can thus be associated with the different Vs of Big Data. For this reason, it is not surprising that sports data in general and event and positional data in particular are increasingly assessed as Big Data (Morgulev et al., 2018; Rein & Memmert, 2016).

4.5. Sentiment analysis in football-specific textual messages

4.5.1. Previous research

Digitisation and the growth of social media has led to the emergence of massive datasets reflecting human online communication, which has stimulated the interest of various stakeholders from business and research to benefit from this development. Textual data being at the heart of online communication, however, has a highly unstructured character, which makes it difficult to be analysed automatically. The high potential to gain insights from these data, in combination with the complexity of analysing it algorithmically, has driven the relevance of the field of sentiment analysis (also called opinion mining), which refers to the automatic extraction of sentiments from textual data. Sentiment analysis is considered “one of the fastest growing research areas in computer science, making it challenging to keep track of all the activities in the area” (Mäntylä, Graziotin, & Kuutila, 2018, p. 16). Attempts to review the current state of sentiment analysis literature can therefore easily include a few hundred (Piryani, Madhavi, & Singh, 2017) or even thousands (Mäntylä et al., 2018) of articles.

Within the analysis of social media data as well as further online or offline communication, the use of tweets from Twitter (Agarwal et al., 2011) is one potential and well-studied aspect of sentiment analysis. At the same time, the non-algorithmic analysis of sports-related aspects of Twitter communication (Sheffer & Schultz, 2010; Witkemper et al., 2012) has become popular to such an extent, that Sanderson (2014) suspected "that there is a good number of reviewers who are probably tired of being asked to review Twitter and other sport and social media studies" (p. 128) as early as in 2014.

Sentiment analysis techniques are generally applicable to texts from any domain, which includes rather scarce applications in sports-related communication (Brown, Rambaccussing, Reade, & Rossi, 2017; Godin, Zuallaert, Vandersmissen, Neve, & van de Walle, 2014; Schumaker et al., 2016). Article VIII will also pursue this approach, however, some caution is warranted as sports can be considered a side-aspect of sentiment analysis, if at all. For example, Mäntylä et al. (2018) identified society, security, travel, finance and corporate, medical, entertainment and other as the predominant application areas.

Consequently, it can be concluded that sentiment analysis techniques including sentiment analysis lexica (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011), i.e. predefined lists of words associated with certain sentiments or sentiment scores, have not been optimized for or even validated in the domain of sports. To the best of the authors knowledge, no prior article has focused on the validation of sentiment analysis techniques applied particularly to sports-related Twitter data. Based on the importance of sentiment analysis and the opportunities to analyse sports-related communication, in combination with the lack of validation in this specific domain, the question whether algorithmic methods of sentiment analysis are accurate enough to correctly classify the sentiment in sports-specific textual messages should be posed. Article VII contributes to answering this question by validating sentiment analysis techniques on a dataset from top-class football matches.

In the context of this dissertation, it can be considered a preliminary work being a necessary precondition for the use of football-related sentiment analysis in Article VIII.

4.5.2. Answer to Research Question 3a)

With regard to research question 3a), Article VII is evidence that lexicon-based sentiment analysis methods are able to accurately classify the sentiment of sets of tweets above a certain sample size, that is commonly available in real-world applications. The accuracy of classifying single tweets, however, is highly limited.

10,000 English language tweets related to ten top-class football matches (1,000 randomly chosen tweets per match) were evaluated manually by human annotators and categorized as either ‘positive’, ‘negative’, ‘neutral’ or ‘nonsense’. After pre-processing the data, tweets were algorithmically evaluated by three publicly available lexicon-based sentiment analysis techniques.

The accuracy for binary classification of individual tweets (positive or negative) was found to range between 61.0% and 63.6% for the three individual methods and 67.4% when using a combined classification of all methods. These results can be explained by the inherently complex task of extracting the sentiment from a single short textual message. Other studies analysing non-sports-related tweets have also reported limited accuracies of around 75% (Agarwal et al., 2011; Kharde & Sonawane, 2016). Moreover, the data suggest that football-specific tweets have their own characteristics and in particular are shorter than tweets from other domains. Tweets from the football domain on average contained 13.3 words, while tweets from comparative sets of tweets contained an average of 18.0 words (finance-related tweets) or 20.8 words (politics-related tweets), implying a higher complexity for correct classification.

Practical applications of sentiment analysis, however, do not depend on the specific sentiment of an individual tweet, but aim at correctly classifying the general sentiment in large sets of tweets. For this reason, test sets with a larger number of tweets and a certain percentage of one polarity were constructed. Then, the accuracy of classifying the predominant sentiment in these test sets were investigated.

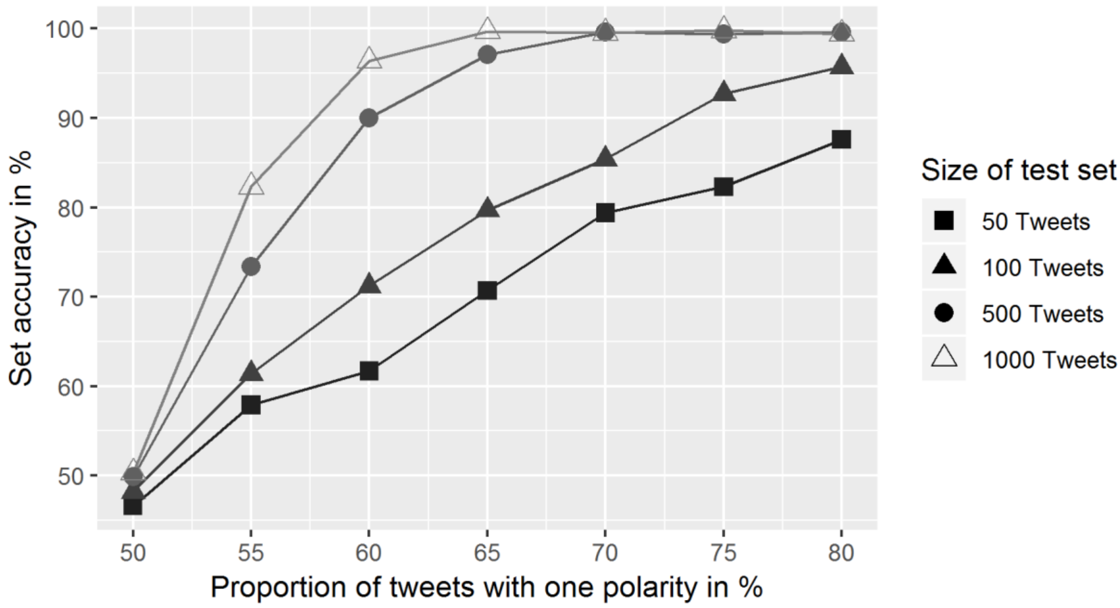

Figure 9: Accuracy (in %) of correctly classified sets of tweets with a given size and a given percentage of tweets with the same polarity.

This more detailed analysis shows that realistic sets of tweets (for example 1,000 tweets with 60% of the tweets containing a certain polarity), can be classified correctly in more than 95% of all cases.

4.6. Using Twitter data for in-play forecasting

4.6.1. Previous research

Article VIII combines two strands of research, namely football forecasting and forecasting from Twitter data. Football forecasting has already been extensively discussed in this dissertation and Article VIII tackles two neglected aspects. First, forecasting over-under goals in football (see Boshnakov et al., 2017; Wheatcroft, 2020 for rare exceptions). Second, the idea of improving forecasts in-play that, to the best of the authors knowledge, has only been focused by Zou, Song, and Shi (2020) and only regarding goals and no further data.

The use of Twitter in forecasting has been mainly discussed in domains, where the outcome of events is highly dependent on public opinion, such as political elections or stock prices. Huberty (2015) investigated election forecasting from social media in general and Twitter in particular by discussing existing studies and an own approach. Although Twitter forecasts seem to have some value, he argues that even simple benchmarks based on incumbency were not outperformed. Overall, he concludes that known methods in forecasting elections from social media have failed. An example from the German federal election in 2009 shows how inconsistent the evaluation of Twitter in forecasting can be. Tumasjan, Sprenger, Sandner, and Welpe (2010) reported that by analysing the number of mentions of German political parties in tweets from Twitter, a forecasting accuracy comparable to election polls can be achieved. Jungherr, Jürgens, and Schoen (2012) later repeated this analysis and found evidence that the results and optimistic implications were attributable to arbitrary choices of the authors. In particular, the simple inclusion of an additional political party to the analysis evoked opposite results and conclusions. In a meta-analysis, Gayo-Avello (2013) likewise identifies serious weaknesses of current approaches, but also gives several suggestions for future research in the domain of election forecasting with Twitter.

Some promising results were found in relation to Twitter forecasts on stock prices. In an investigation of daily changes in the Dow Jones Industrial Average (Bollen, Mao, & Zeng, 2011), the general usefulness of Twitter data in order to analyse public mood was confirmed, as some mood dimensions were shown to be predictive of stock prices. Further mood dimensions and the pure classification of positive or negative mood, however, showed no predictive value. Emotions in tweets as a reflection of public mood served as valuable predictors of stock market indicators such as Dow Jones, NASDAQ or S&P 500 (Zhang et al., 2011). The percentage of emotional tweets on a specific day was shown to be negatively correlated to stock market indicators of the next day and positively correlated to an index of market volatility.

A few results on forecast football matches from Twitter have been published so far and have reported mixed results. Forecasting accuracy appears to be poor, as forecasts based on volume and sentiment of tweets performed similar to the naïve prediction rule of selecting the home team to win and were outperformed by forecasts based on statistical analysis, experts or bookmakers (Godin et al., 2014). Likewise, forecasts based on tone and polarity of tweets were outperformed by betting odds in terms of predictive accuracy (Schumaker et al., 2016). At the same time, the two aforementioned publications claim promising results in terms of profitability, however, using a non-odds-dependent betting strategy and without addressing the issue of high randomness in betting returns (see Article II). The arguably most detailed analysis of Twitter data in football forecasting has been conducted by Brown et al. (2017), who investigated whether tweets include information that is not reflected in in-play betting odds. The authors claim that this is the case, in particular when considering the tone of tweets and after significant match events. Despite a very detailed and sound statistical analysis, the results need to be assessed with some caution, as the authors do not address the issue of matching betting data and Twitter data, where the analysis of data in terms of seconds could be distorted even by short timely inconsistencies.

The analysis of Twitter data also has impact on theoretical questions. As argued in the previous publications (Brown et al., 2017; Godin et al., 2014; Schumaker et al., 2016) and driven by the involvement of a large group of different users, Twitter qualifies as a source of crowd wisdom. However, in sharp contrast to the betting market there is neither a market situation with financial incentives for accurate assessments of the situation, nor is Twitter focused on forecasts. Furthermore, there is no a priori mechanism of aggregating opinions. In this sense, despite the argument of crowd wisdom, the value of Twitter data in forecasting is to be seen very critically from a theoretical point of view.

4.6.2. Answer to Research Question 3b)

With regard to research question 3b), no evidence was found that information extracted from Twitter data can help to improve in-play forecasting models in football. This conclusion is drawn from an analysis of almost two million tweets related to more than 400 Premier League matches and an additional analysis of goals in more than 30,000 professional matches in major European leagues.

The sentiment of tweets was analysed using the method of Article VII, moreover, further tweet specific features, namely the intensity as well as the number of words, hashtags and emoticons were investigated. When including Twitter information into forecasting models, no improvement of forecasting accuracy was found in comparison to models based on pre-game betting odds. This result is consistent both when using information from the first half to forecast the number of goals in the second half and when using information from time intervals of five minutes to forecast whether a goal is scored in the next five minutes.

Besides the forecasting results, further findings were revealed when analysing tweet information over the full course of 90 minutes as well as in the time periods shortly before and after goals. First, the overall sentiment reflected in tweets decreases continuously over the course of football matches. Second, tweets clearly react to goals in a way that more and shorter tweets are posted. Surprisingly the overall level of sentiment is rather unaffected by goals. Except for very slight increases in sentiment, the tweets do not show any relevant anomalies in the minutes before goals, which is in line with the notion of a missing predictive value.

While Twitter data did not add predictive value to pre-game information, this result does not necessarily need to be attributed to missing information in the tweets or an insufficient extraction of such information. It is also possible that the process of a football match is virtually stable across the 90 minutes of play and thus information becoming available in-play do not inherently contain additional value. This idea, although probably in contrast to the intuition of football experts, is backed up by a goal-based analysis. It was demonstrated that goals in the first half hardly outperform forecasts based on naïve benchmarks and do not further improve on forecasts once controlling for pre-game expectation. Given a large database of more than 30,000 matches, this can be considered a very robust result.

4.7. Using event and positional data for in-play forecasting

4.7.1. Previous research

Article IX combines two strands of research for the first time. First, in-play forecasting in football, that has already been discussed with regard to Article VIII before. Second, performance analysis in football based on event and positional data.

Performance analysts face the problem that the collaborative movement of players on the pitch as well as the sequences of events imply a high degree of complexity. Theoretically, researchers have used dynamical systems theory as a framework to analyse the movements of players, which is inspired by movement patterns of animals (Passos, Araújo, & Davids, 2013). One basic characteristic of this theoretical concept is the existence of organismic, environmental and task constraints (Glazier & Robins, 2013). In football, such constraints can be operationalised by means of situational variables including match location, current scoreline or team strength (Fernandez-Navarro et al., 2018; Lago, 2009; Lago-Peñas, 2012, see also Article X & Article XI). From a data perspective, the action on the pitch is summarized by means of event and positional data as has been elaborated on before.

Attempts to effectively extract relevant information from positional data include analysing variables like team centroids, space control, playing spaces including spread (i.e. dispersion) of teams, passing networks and several applications of machine learning to detect tactical patterns including team formations (for reviews see Low et al., 2020; Rein & Memmert, 2016). However, describing and understanding the collective behaviour on the pitch is not sufficient to establish practical relevance for coaches and professional football teams. To draw conclusions on performance-enhancing strategies, links between measures deduced from positional data and actual performance need to be established (Rein et al., 2017).

Event data, being a much more structured data source, represents a lower complexity, as the events (e.g. number of shots) can directly be assessed as a potential measure of performance. Consequently, there is a solid body of literature investigating the link between event based measures and success in football (Castellano et al., 2012; Lago-Peñas, Lago-Ballesteros, Dellal, & Gómez, 2010; Lepschy et al., 2020).

However, such attempts fail to explain whether success actually can be explained by a performance measure, or whether both are just a reflection of general team strength. Moreover, the direction of influence has not been sufficiently examined, which refers to the problem that performance measures can influence success, but (intermediate) success, in a sense of the current scoreline, can also influence performance measures. Both problems are circumvented by a framework to analyse the ability of performance indicators for in-play forecasting of football matches in Article IX.

4.7.2. Answer to Research Question 3c)

Article IX did not provide evidence that information extracted from event and positional data can help to improve in-play forecasting models in football. This answer is based on an analysis of 18 performance indicators (denoted as PIs subsequently), based on event and positional data from 50 matches including more than 300 million datapoints. The PIs were tested using a novel framework to investigate the predictive in-play value and distinguish between explanatory and predictive power.

The framework considers differences between the two competing teams in both halves of the match in terms of betting odds as a measure of team strength, goals as a measure of success and PIs as a measure of performance. Based on this, the strength dependence (relation between team strength and PIs), explanatory power (relation between PIs and goals in the first half), predictive power (relation between PIs in the first half and goals in the second half) and predictive overperformance (analogous to predictive power, but controlling for pre-game expectation on PIs and goals) are calculated.

The results of strength dependence, explanatory power and predictive power are highly inconsistent for the majority of PIs. This sheds light on the complex relation between team strength, performance and success and is clear evidence that links between PIs and success in the same match as common in performance analysis in football (Castellano et al., 2012; Lago-Peñas et al., 2010; Lepschy et al., 2020) should be viewed very critically. As one example, the number of clearances (strength dependence: $r(48) = -.15$, $p = .29$; explanatory power: $r(48) = .34$, $p < .05$; predictive power $r(48) = -.09$, $p = .53$) shows a non-significant negative relation to both team strength and future success, while it shows a significant positive relation to success in the same half of the match. This is clear evidence that clearances are not associated with strong or successful football teams, but mainly a consequence of the scoreline, as leading teams are more prone to use clearances to defend the lead.

While a variety of PIs (namely shots, passes, short passes, ball distance and several measures of space control) show a significant ($p < .05$) relationship to team strength, only passes ($r(48) = .24$, $p < .1$) and short passes ($r(48) = .24$, $p < .1$) reveal significant predictive value and only at a ten percent significance level. In terms of predictive overperformance, i.e. if controlling for pre-game expectation in terms of betting odds, no PI shows significant relation anymore, which is evidence for the limited predictive value in-play. The highest non-significant correlations were found for passes, short passes, ball distance and ball possession, which are the most promising variables and might reveal some weak value if increasing the sample size of matches.

In full analogy to Article VIII a dataset of 30,000 additional matches was analysed and revealed that the goal difference from the first half does not significantly improve forecasts made for the goal difference in the second half of a match when controlling for the pre-game expectation. This is clear evidence that in-play information in terms of goals does not reveal any information that was not known before the start of the match. This puts the results of PIs into context and underlines that in-play forecasting in football seems to be a highly difficult task, independent of the data used. Scoring in football matches thus can be seen as a relatively stable process with predefined goal expectations.

5. Discussion, Conclusions and Implications

5.1. Discussion

The results with regard to the various research questions have already been discussed in detail in the respective articles and have been embedded in the context of the literature in chapters 2 to 4 of this dissertation. At this point, the dissertation as a whole will be discussed, with a special focus on the interdisciplinarity and differences to experimental sports science research.

The variety of applications for sports forecasting in different research areas as outlined in the Introduction implies the possibility to build bridges by making use of strongly interdisciplinary research approaches. The integration of different disciplines, however, involves the risk of creating methodical fields of tension. Although sports is at the centre of this dissertation and it was written at a university with a strong sports science focus, some fundamental differences to experimental sports science research exist and shall be discussed. In particular, this concerns differences between theory-based and application-based research, the contrast between experimental designs and the analysis of existing data as well as questions of statistical inference referring to the ongoing discussion on the usage of p-values and effect sizes (Wasserstein, Schirm, & Lazar, 2019). The current dissertation has inevitably been influenced by these aspects and thus should be understood and discussed in the light of these three fields of tension, that will be further explained below.

While in theory-based research, empirical data is used to find evidence in favour or against a scientific theory, the focus of application-based research is rather on the practical usability of the insights. Related to sports forecasting, this means, that providing an improved predictive power by either presenting a novel method, incorporating additional data, or using a modified model specification has value in itself. The reasons and the exact process underlying the superiority and whether this is evidence in favour or against an existing scientific theory is often of secondary importance, if at all. Within this dissertation, links to theory-based research have particularly been discussed with regard to the usage of betting odds (see section 3.1), that is associated with theories on market efficiency and crowd wisdom. Further theoretical aspects are touched with regard to mathematical theory and proofs (in Article II), dynamic systems theory with regard to performance analysis in football (in Article VI), or several theories with a socio-psychological or biological background in relation to explaining the home advantage (in Article V). However, the predominant part of the research questions in this dissertation can be considered to have a strongly application-based focus.

A second notable difference is the method of data collection. Experimental designs in sports science usually start with the design of the experiment and the related data collection, which means that data are typically collected for the sole purpose of analysis. Forecasting, on the other hand, can rather be characterized as data mining (Raval, 2012), which means that the starting point of analysis are existing data, that have not been collected or at least not primarily for the purpose of scientific investigation. This has implications with regard to so-called p-hacking, referring to various methods of unethically increasing the possibility of finding significant p-values which includes that “researchers try out several statistical analyses and/or data eligibility specifications and then selectively report those that produce significant results” (Head, Holman, Lanfear, Kahn, & Jennions, 2015, p. 1). Data mining refers to the process of extracting relevant knowledge from large databases (Raval, 2012) and thus could be erroneously disregarded as p-hacking. None of the articles being part of the synopsis presents self-conducted experiments with test persons and only for Article VI and VII, existing data was extended by manual human annotation. All further analyses are based on either literature review, theoretical mathematical considerations or the usage of existing data. In this sense, the analysis methods naturally need to be adapted to the data and its availability, instead of customising the data collection to fit the intended analysis methods. The author would like to acknowledge that this needs to be carefully considered when discussing the results.

The very current debate on statistical inference (Wasserstein et al., 2019) particularly affects sports science for two reasons: First, sample sizes for experimental designs can be strongly limited by difficulties in finding suitable test persons or a costly and time-consuming data collection. As a consequence, statistical analysis of such experiments might fail to find significant results, although relevant effects exist. Second, in intervention studies the cost of an intervention (in terms of financial resources, time or physical effort) can be considerable. In this respect, even with a significant result, the future implementation of such an intervention is only reasonable if the effect achieved justifies the effort. For these reasons, it is essential, especially in sports science, to discuss the different implications of p-values and effect sizes. Although forecasting might be regarded as an even more statistical field of research, the interpretation of p-values and the usage of effect sizes can be considered less problematic, as the two above-mentioned problems are not equally present. In the majority of cases, forecasting models can be tested on a large database (i.e. large sample size), which accordingly more often leads to p-values being small enough to present unambiguous results. The lower importance of effect sizes in forecasting is a consequence of the fact that different methods or model specifications usually do not differ considerably in terms of effort, which means that a better model will be preferred, even if the positive effect of using the superior model is minimal. In future, however, the discussion about effect sizes could also become increasingly relevant in the field of sports forecasting. If using more data-intensive and complex methods from the domain of computer science, computational time and financial effort for the required computational power need to be put into relation to a possibly superior forecasting accuracy of these methods.

The articles in this dissertation are not only heterogenous in terms of scientific disciplines, but also strongly differ in terms of sample sizes (e.g. 50 matches in Article IX or a few ten thousand matches in Article V). Despite the justified criticism, p-values and the terminology significant or non-significant are still standard in the scientific world and have accordingly been used throughout the dissertation. The author would like to clarify that readers should not understand this as a strict dichotomy and be aware that breaking results down to significant or non-significant is driven by the need to simplify complex relationships into clear answers. Readers should carefully consider the statistical results as a whole, instead of merely relying on a black or white result in terms of significance. In line with the above arguments, effect sizes have only been partially reported and in particular not been discussed with regard to assessing the quality of forecasting models.

5.2. Conclusions

In conclusion, the present dissertation strongly supports the notion that betting odds are an excellent predictor of success in football. While the benefit of betting odds in football forecasting is already a well-established result (Forrest et al., 2005; Hvattum & Arntzen, 2010; Štrumbelj & Šikonja, 2010), the exceptional value of betting odds with regard to several additional aspects, has been evidenced for the first time within this dissertation: Betting odds known before a match allow for more accurate information about team strength than the actual match result (Article III). Betting odds are a valuable predictor of success in penalty shootouts (Article IV). Models based on the first half goals are clearly outperformed by models based on betting odds when forecasting the second half result in football matches with regard to both the number (Article VIII) and difference (Article IX) of goals. Moreover, it has been found to be highly challenging to improve forecasts based on betting odds by using further in-play information such as Twitter data (Article VIII) or event and positional data (Article IX).

The results of this work have theoretical implications with regard to crowd wisdom and information efficiency in markets (see section 3.1). As betting odds can be considered an aggregated estimation influenced by various stakeholders, the accuracy of betting odds is in line with the idea that collaborative estimation leads to high forecasting success (crowd wisdom). While no formal tests of market efficiency are employed, the results are moreover in line with the notion that relevant information is correctly processed in the sports betting market and reflected by the betting odds. Furthermore, the well-documented value of betting odds has methodological implications in sports science that will be discussed in section 5.3.

The present dissertation has conceptionally, statistically and empirically assessed the substantial role of randomness in sports in general and in football in particular. Article I suggests that unsystematic, i.e. random contributions should be conceptionally regarded and modelled as an individual fundamental aspect contributing to the outcomes of sports events. Statistically, the inherent randomness in sports outcomes is large enough, that even seemingly substantial betting returns can be achieved by an arbitrary selection of bets (Article II). Empirically, it has been shown that almost every second goal scored in football was subject to at least one random influence and that random influence remained substantial in virtually all match situations, except for penalties (ArticleVI). The highly limited informative value of goals in football, that has been demonstrated in Article III (the result of a match contains less information on team strength than betting odds before a match) as well as Article VIII and IX (goals in the first half do not improve forecasts for the second half based on betting odds), can be considered a direct consequence of this. Practical and theoretical implications of these results for performance analysis in football will be discussed in section 5.3.

The failure to improve pre-game forecasts based on the number of goals in the first half, does not only provide evidence for the high randomness of goals, but might also more generally call into question the in-play predictability of football matches. This idea is backed up by the highly limited success of Twitter data (Article VIII) and positional data (Article IX) for in-play forecasting. This implies a need to make further efforts in order to understand to what extent information becoming available over the course of football matches does have predictive value at all. While a conclusive answer to this question has not been given yet, a limited predictive value as indicated by this dissertation, would have serious implications for bookmakers and match analysts. Bookmakers should not attribute too much importance on using complex in-play forecasting models and match analysts should not be overconfident in drawing conclusions from the observation of an ongoing match to the further course of this match.

5.3. Implications

5.3.1. Implications from a mathematical and economic perspective

Article I is a contribution to a more general view on sports forecasting that has been missing in the literature so far. However, as an implication of this dissertation, additional effort should be made to investigate general aspects of sports forecasting that concern several use cases (e.g.

sports or competitions). At the same time, it seems necessary to investigate differences between these use cases in a more systematic approach. One possibility to tackle such questions is the use of artificial data that can be beneficial to particularly identify strengths and weaknesses of rating procedures and forecasting models by intentionally varying single aspects in the data creation.

Researchers should stop assessing accuracy and profitability as just two concurrent measures of predictive quality in sports forecasting or even carelessly mixing both aspects when interpreting the results. Instead, it appears necessary to actively investigate and discuss differences, as has already been done for a long time in other fields of economics (Boothe & Glassman, 1987; Ertimur et al., 2007; Fuertes et al., 2015; Leitch & Tanner, 1991) and very recently in sports betting by Hubáček and Šourek (2020) with reference to Article II. If models are intended to draw conclusions about the underlying processes, for example when using them in sports science, models should be developed and assessed with the sole purpose of achieving the highest possible accuracy. If the purpose of a model, however, is to generate positive betting returns, the goal of profitability should already be reflected when fitting the model. One idea that points in this direction is to optimize model profitability by intentionally decreasing the correlation between the forecasts of model and bookmaker (Hubáček et al., 2019). Moreover, positive betting returns should be assessed more critically by using measures similar to p-values to reflect the inherent randomness in betting returns.

5.3.2. Implications from a sports science perspective

The most straightforward implication of this dissertation is to start using the information enclosed in betting odds for the purpose of sports science. Researchers in sports science unfortunately do not seem to be aware or convinced of the usefulness of betting odds so far, although the high accuracy of betting odds is accepted in the forecasting literature (Baker & McHale, 2013; Forrest et al., 2005; Hvattum & Arntzen, 2010; Kovalchik, 2016; Štrumbelj & Šikonja, 2010; Štrumbelj & Vračar, 2012). The present dissertation has given several application examples of using betting odds to answer sports scientific questions. In this context, betting odds can be used to obtain a measure of relative (i.e. match-specific) team strength (Article IV & VI), a measure of absolute team strength (Article III), an indicator for balanced or unbalanced matches (Article VI) and a measure of home advantage (Article V).

Moreover, it is recommendable to stop overestimating the informative value of goals or assuming a clear connection between skill and success in football by using goals in a match as a measure of team strength. This dissertation has presented further evidence for the inherent randomness in sports in general and in football in particular. In contrast to the betting odds, it is evident that many researchers are already well aware of the limited informative value of goals, as they have started to tackle this problem by using score box possessions (Sgrò et al., 2017; Tenga, Holme, et al., 2010; Tenga, Ronglan, & Bahr, 2010) instead of goals or by investigating and developing potential key performance indicators (Brechot & Flepp, 2020; Castellano et al., 2012; Rein et al., 2017) that minimize random noise when assessing performance.

Further sports scientific implications, that are not directly connected to forecasting and betting odds, are the following: Stronger football teams tend to be more successful in penalty shootouts than weaker teams (Article IV); home advantage in football is not predominantly driven by the presence of spectators (Article V) and coaches should discuss the potential value of deliberately creating uncontrolled actions to score goals (Article VI).

5.3.3. Implications from a data analysis and computer science perspective

The analysis of sources of Big Data related to sports forecasting has yielded mixed results. One implication is that the complexity of data should not be overstated by seeing it as an end in itself. Analysis of Big Data with regard to Twitter (Article VIII) as well as event and positional data (IX) did not help to improve in-play forecasts based on betting odds, representing data as strikingly simple as obtaining a few numbers from the website of a bookmaker. This is obviously bad news if adopting the opinion of Mauro et al. (2016) “that information, not data, is the fundamental fuel of the current Big Data phenomenon.” (p. 124). It supports critics of

Big Data (Ekbia et al., 2015) and is in line with the notion that in forecasting simple models can be superior to complex ones (cf. Gilliland, 2020).

However, this does not necessarily negate the usefulness of complex data in sports forecasting. First, if a process inherently doesn't allow for in-play forecasting, it will not be possible with any amount of data or sophisticated method. It appears that in-play forecasting in football is such a highly difficult task independent from the data used, which can be considered a further important implication. The fact that goals from the first half likewise did not improve forecasts from betting odds for the number (Article VIII) and difference (Article IX) of goals clearly points in this direction. Second, the assessment of betting odds as 'simple' is driven by the fact that they are already the result of a process, which, in itself, can be highly complex or involve complex data. The direct comparison is therefore flawed as the final result of a sophisticated machine learning model can also be represented by a single number, but involves a high degree of complexity. Further research is needed to give conclusive answers on whether in-play forecasting in football is feasible at all and if Big Data can contribute to improve in-play or pre-game forecasting.

Besides the difficulty of in-play forecasting, the present dissertation has further implications with regard to Big Data analysis in sports, such as the fact that sentiment analysis of football-related tweets was able to accurately determine the sentiment in samples of hundreds or thousands of tweets common in real-world applications. The analysis of Twitter data also revealed insights into the sentiment of users over the course of matches and after goals. Moreover, some performance indicators based on event and positional data – although failing to reach significance – showed promising results with regard to predictive value. Although the present dissertation cannot provide proof that sports forecasting can benefit from sources of Big Data, further research in this direction is certainly needed and useful as a variety of research questions in this domain have not been sufficiently investigated.

6. Summary and Outlook

In summary, this dissertation has pursued a holistic approach and has given a diverse view on sports forecasting from three different directions. From a methodological standpoint, Articles III – VI have shown how betting odds can be used as a tool to improve analysis in sports science and the results are evidence for the usefulness of this approach. Moreover, the results of this dissertation encourage awareness for important issues that need to be considered in future research on sports forecasting and performance analysis in sports. This includes the differences of accuracy and profitability in evaluating sports forecasting methods, the prominent contribution of randomness on the outcomes of football matches, the difficulty of in-play forecasting in football as well as the contribution of Big Data, that needs to be driven forward critically. The contributions made by this dissertation imply a variety of domains and research topics that should be considered in the near future. These primarily aim at tackling general weaknesses of sports forecasting processes, an improved connection between predictive modelling and sports science, as well as the critical assessment of Big Data in sports forecasting. Table 4 gives a summary of possible future domains.

Table 4: Future domains and research topics in sports forecasting

	Future domains and research topics	Based on
1	***Economic and Mathematical Aspects*** General aspects of sports forecasting	
	- Using artificial (i.e. simulated) data to understand the strengths and weaknesses of rating procedures and forecasting methods.	Article I
	- Fitting forecasting models based on profitability instead of accuracy when aiming at profitable models.	Article II
2	***Aspects of Sports Science*** The role of randomness and betting odds in the domain of sports science	

	- Implementing a generalized concept to distinguish more clearly between skill, performance and success in football (and other sports games).	Article III, VI & IX
	- Implementing betting odds as a standard tool to estimate team strength as a situational variable in performance analysis.	Article I, III – VI, VIII, IX
	- Understanding and embracing the limits of predictability in football (and further sports games) by particularly investigating random contributions.	Article I & VI
	- Using in-play forecasting to understand the process of football matches (and further sports games) as well as the value of information becoming available over the course of a match.	Article VIII & IX
3	***Aspects of Data Analysis and Computer Science*** Using Big Data in sports forecasting	
	- Properly validating and using automatization processes including, but not limited to sentiment analysis in football (and further sports)	Article VII
	- Expanding Big Data analysis in sports forecasting to larger sets of matches and pre-game forecasting	Article VIII & IX
	- Critically investigating and discussing the usefulness of Big Data in sports forecasting	Article VII - IX

Besides the above-mentioned research areas, that are directly linked to this work, the topic of sports forecasting can be extended in several directions. One direction is related to aspects of sports betting, that could be argued to be more socially relevant. Attempts to use knowledge from sports forecasting in order to investigate match fixing exist in science (Forrest & McHale, 2019) and practice (Sports Integrity Initiative, 2017) and appear worth expanding. A novel and promising aspect is to analyse the betting behaviour of sports bettors in detail, with regard to economic and psychological aspects. While this would be highly valuable to better understand gambling addiction related to sports betting, it would require availability of internal bookmaker data as in contrast to the betting odds, the bets actually taken by the bettors are not publicly observable. Another direction is to link sports forecasting research stronger to other domains of forecasting. Such connections already exist, for example, as the value of betting odds has been demonstrated in the domain of political elections (Erikson & Wlezien, 2012; Wolfers & Leigh, 2002). More generally, the idea of benefitting from market structures and collaborative knowledge is not exclusive to sports betting, but the basis for the research domain of prediction markets (Wolfers & Zitzewitz, 2004), that can be applied to various aspects of daily life.

7. References

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment Analysis of Twitter Data. *Proceedings of the Workshop on Languages in Social Media*, 30–38.

Alberti, G., Iaia, F. M., Arcelli, E., Cavaggioni, L., & Rampinini, E. (2013). Goal scoring patterns in major European soccer leagues. *Sport Sciences for Health*, *9*(3), 151–153. https://doi.org/10.1007/s11332-013-0154-9

Alwan, H. B., & Ku-Mahamud, K. R. (2020). Big data: definition, characteristics, life cycle, applications, and challenges. *IOP Conference Series: Materials Science and Engineering*, *769*, 12007. https://doi.org/10.1088/1757-899X/769/1/012007

Anderson, M., Wolfson, S., Neave, N., & Moss, M. (2012). Perspectives on the home advantage: A comparison of football players, fans and referees. *Psychology of Sport and Exercise*, *13*(3), 311–316. https://doi.org/10.1016/j.psychsport.2011.11.012

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. https://doi.org/10.1016/j.psychsport.2008.07.008

Angelini, G., & Angelis, L. de (2019). Efficiency of online football betting markets. *International Journal of Forecasting*, *35*(2), 712–721. https://doi.org/10.1016/j.ijforecast.2018.07.008

Armatas, V., & Yiannakos, A. (2010). Analysis and evaluation of goals scored in 2006 World Cup. *Journal of Sport and Health Research*, *2*, 119–128.

Asif, M., & McHale, I. G. (2016). In-play forecasting of win probability in One-Day International cricket: A dynamic logistic regression model. *International Journal of Forecasting*, *32*(1), 34–43. https://doi.org/10.1016/j.ijforecast.2015.02.005

Assunção, R., & Pelechrinis, K. (2018). Sports Analytics in the Era of Big Data: Moving Toward the Next Frontier. *Big Data*, *6*(4), 237–238. https://doi.org/10.1089/big.2018.29028.edi

Baboota, R., & Kaur, H. (2019). Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting*, *35*(2), 741–755. https://doi.org/10.1016/j.ijforecast.2018.01.003

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

Baumeister, R. F., & Steinhilber, A. (1984). Paradoxical effects of supportive audiences on performance under pressure: The home field disadvantage in sports championships. *Journal of Personality and Social Psychology*, *47*(1), 85–93. https://doi.org/10.1037/0022-3514.47.1.85

Beal, R., Norman, T. J., & Ramchurn, S. D. (2019). Artificial intelligence for team sports: a survey. *The Knowledge Engineering Review*, *34*. https://doi.org/10.1017/S0269888919000225

Bernardo, G., Ruberti, M., & Verona, R. (2019). Semi-strong inefficiency in the fixed odds betting market: Underestimating the positive impact of head coach replacement in the main European soccer leagues. *The Quarterly Review of Economics and Finance*, *71*, 239–246. https://doi.org/10.1016/j.qref.2018.08.007

Berrar, D., Lopes, P., Davis, J., & Dubitzky, W. (2019). Guest editorial: special issue on machine learning for soccer. *Machine Learning*, *108*(1), 1–7. https://doi.org/10.1007/s10994-018-5763-8

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, *22*(3), 547–581. https://doi.org/10.1016/j.ijforecast.2006.04.001

Boothe, P., & Glassman, D. (1987). Comparing exchange rate forecasting models. *International Journal of Forecasting*, *3*(1), 65–79. https://doi.org/10.1016/0169-2070(87)90079-3

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. https://doi.org/10.1016/j.ijforecast.2016.11.006

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

Boulier, B. L., Stekler, H. O., Coburn, J., & Rankins, T. (2010). Evaluating National Football League draft choices: The passing game. *International Journal of Forecasting*, *26*(3), 589–605. https://doi.org/10.1016/j.ijforecast.2009.10.009

Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007). Referee bias contributes to home advantage in English Premiership football. *Journal of Sports Sciences*, *25*(11), 1185–1194. https://doi.org/10.1080/02640410601038576

Brandt, P. T., Freeman, J. R., & Schrodt, P. A. (2014). Evaluating forecasts of political conflict dynamics. *International Journal of Forecasting*, *30*(4), 944–962. https://doi.org/10.1016/j.ijforecast.2014.03.014

Braun, S., & Kvasnicka, M. (2013). National Sentiment and Economic Behavior. *Journal of Sports Economics*, *14*(1), 45–64. https://doi.org/10.1177/1527002511414718

Brechot, M., & Flepp, R. (2020). Dealing With Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals. *Journal of Sports Economics*, *21*(4), 335–362. https://doi.org/10.1177/1527002519897962

Brefeld, U., & Zimmermann, A. (2017). Guest editorial: Special issue on sports analytics. *Data Mining and Knowledge Discovery*, *31*(6), 1577–1579. https://doi.org/10.1007/s10618-017-0530-1

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. https://doi.org/10.1111/ecin.12506

Bruns, A., & Stieglitz, S. (2013). Towards more systematic Twitter analysis: metrics for tweeting activities. *International Journal of Social Research Methodology*, *16*(2), 91–108. https://doi.org/10.1080/13645579.2012.756095

Buraimo, B., Forrest, D., & Simmons, R. (2010). The 12th man? refereeing bias in English and German soccer. *Journal of the Royal Statistical Society: Series a (Statistics in Society)*, *173*(2), 431–449. https://doi.org/10.1111/j.1467-985X.2009.00604.x

Castellano, J., Casamichana, D., & Lago, C. (2012). The Use of Match Statistics that Discriminate Between Successful and Unsuccessful Soccer Teams. *Journal of Human Kinetics*, *31*, 139–147. https://doi.org/10.2478/v10078-012-0015-7

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

Cervone, D., D’Amour, A., Bornn, L., & Goldsberry, K. (2014). Pointwise: Predicting points and valuing decisions in real time with nba optical tracking data. In *Proceedings of the 8th MIT Sloan Sports Analytics Conference, Boston, MA, USA*.

Constantinou, A. C., & Fenton, N. E. (2012). Solving the Problem of Inadequate Scoring Rules for Assessing Probabilistic Football Forecast Models. *Journal of Quantitative Analysis in Sports*, *8*(1). https://doi.org/10.1515/1559-0410.1418

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. https://doi.org/10.1016/j.knosys.2012.07.008

Courneya, K. S., & Carron, A. V. (1992). The home advantage in sport competitions: A literature review. *Journal of Sport and Exercise Psychology*, *14*(1), 28–39.

Dalton, K., Guillon, M., & Naroo, S. A. (2015). An Analysis of Penalty Kicks in Elite Football Post 1997. *International Journal of Sports Science & Coaching*, *10*(5), 815–827. https://doi.org/10.1260/1747-9541.10.5.815

Davis-Stober, C. P., Budescu, D. V., Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision*, *1*(2), 79–101. https://doi.org/10.1037/dec0000004

Decroos, T., Bransen, L., van Haaren, J., & Davis, J. (2019). Actions Speak Louder than Goals. In A. Teredesai (Ed.), *ACM Digital Library, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1851–1861). New York,NY,United States: Association for Computing Machinery. https://doi.org/10.1145/3292500.3330758

Dey, A. (2016). Machine learning algorithms: a review. *International Journal of Computer Science and Information Technologies*, *7*(3), 1174–1179.

Dick, U., & Brefeld, U. (2019). Learning to Rate Player Positioning in Soccer. *Big Data*, *7*(1), 71–82. https://doi.org/10.1089/big.2018.0054

Dicks, M., Uehara, L., & Lima, C. (2011). Deception, Individual Differences and Penalty Kicks: Implications for Goalkeeping in Association Football. *International Journal of Sports Science & Coaching*, *6*(4), 515–521. https://doi.org/10.1260/1747-9541.6.4.515

Direr, A. (2011). Are betting markets efficient? Evidence from European Football Championships. *Applied Economics*, *45*(3), 343–356. https://doi.org/10.1080/00036846.2011.602010

Dohmen, T. J. (2008). Do professionals choke under pressure? *Journal of Economic Behavior & Organization*, *65*(3-4), 636–653. https://doi.org/10.1016/j.jebo.2005.12.004

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., . . . Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, *66*(8), 1523–1545. https://doi.org/10.1002/asi.23294

Ekin, A., Tekalp, A. M., & Mehrotra, R. (2003). Automatic soccer video analysis and summarization. *IEEE Transactions on Image Processing : A Publication of the IEEE Signal Processing Society*, *12*(7), 796–807. https://doi.org/10.1109/TIP.2003.812758

Erikson, R. S., & Wlezien, C. (2012). Markets vs. polls as election predictors: An historical assessment. *Electoral Studies*, *31*(3), 532–539. https://doi.org/10.1016/j.electstud.2012.04.008

Ertimur, Y., Sunder, J., & Sunder, S. V. (2007). Measure for Measure: The Relation between Forecast Accuracy and Recommendation Profitability of Analysts. *Journal of Accounting Research*, *45*(3), 567–606. https://doi.org/10.1111/j.1475-679X.2007.00244.x

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486

Fernandez-Navarro, J., Fradua, L., Zubillaga, A., & McRobert, A. P. (2018). Influence of contextual variables on styles of play in soccer. *International Journal of Performance Analysis in Sport*, *18*(3), 423–436. https://doi.org/10.1080/24748668.2018.1479925

Fister, I. (2021). Computational Intelligence and Data Mining in Sports. *Applied Sciences*, *11*(6), 2637. https://doi.org/10.3390/app11062637

Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, *21*(3), 551–564. https://doi.org/10.1016/j.ijforecast.2005.03.003

Forrest, D., & McHale, I. G. (2019). Using statistics to detect match fixing in sport. *IMA Journal of Management Mathematics*, *30*(4), 431–449. https://doi.org/10.1093/imaman/dpz008

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. https://doi.org/10.1016/j.ijforecast.2010.01.004

Fuertes, A.-M., Kalotychou, E., & Todorovic, N. (2015). Daily volume, intraday and overnight returns for volatility prediction: profitability or accuracy? *Review of Quantitative Finance and Accounting*, *45*(2), 251–278. https://doi.org/10.1007/s11156-014-0436-6

Furley, P., Noël, B., & Memmert, D. (2017). Attention towards the goalkeeper and distraction during penalty shootouts in association football: A retrospective analysis of penalty shootouts from 1984 to 2012. *Journal of Sports Sciences*, *35*(9), 873–879. https://doi.org/10.1080/02640414.2016.1195912

Galton, F. (1907). Vox populi. *Nature*, 450–451.

Garicano, L., Palacios-Huerta, I., & Prendergast, C. (2005). Favoritism Under Social Pressure. *Review of Economics and Statistics*, *87*(2), 208–216. https://doi.org/10.1162/0034653053970267

Gauriot, R., & Page, L. (2019). Fooled by Performance Randomness: Overrewarding Luck. *Review of Economics and Statistics*, *101*(4), 658–666. https://doi.org/10.1162/rest_a_00783

Gayo-Avello, D. (2013). A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data. *Social Science Computer Review*, *31*(6), 649–679. https://doi.org/10.1177/0894439313493979

Gilliland, M. (2020). The value added by machine learning approaches in forecasting. *International Journal of Forecasting*, *36*(1), 161–166. https://doi.org/10.1016/j.ijforecast.2019.04.016

Glazier, P. S., & Robins, M. T. (2013). Constraints in Sports Performance. *Routledge Handbook of Sports Performance Analysis*, 42–51.

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

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, *21*(2), 331–340. https://doi.org/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. https://doi.org/10.1002/for.877

Godin, F., Zuallaert, J., Vandersmissen, B., Neve, W. de, & van de Walle, R. (2014). Beating the bookmakers: leveraging statistics and Twitter microposts for predicting soccer results. *KDD Workshop on Large-Scale Sports Analytics*.

Gonzalez-Rodenas, J., Lopez-Bondia, I., Calabuig, F., Pérez-Turpin, J. A., & Aranda, R. (2017). Creation of goal scoring opportunities by means of different types of offensive actions in US major league soccer. *Human Movement*, *2017*(5), 106–116. https://doi.org/10.5114/hm.2017.73616

Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime. *International Journal of Forecasting*, *19*(4), 579–594. https://doi.org/10.1016/S0169-2070(03)00092-X

Goumas, C. (2014a). Home advantage and referee bias in European football. *European Journal of Sport Science*, *14 sup1*, 243-249. https://doi.org/10.1080/17461391.2012.686062

Goumas, C. (2014b). Home advantage in Australian soccer. *Journal of Science and Medicine in Sport*, *17*(1), 119–123. https://doi.org/10.1016/j.jsams.2013.02.014

Green, K. C., Armstrong, J. S., & Soon, W. (2009). Validity of climate change forecasting for public policy decision making. *International Journal of Forecasting*, *25*(4), 826–832. https://doi.org/10.1016/j.ijforecast.2009.05.011

Groll, A., Heiner, J., Schauberger, G., & Uhrmeister, J. (2020). Prediction of the 2019 IHF World Men's Handball Championship – A sparse Gaussian approximation model. *Journal of Sports Analytics*, *6*(3), 187–197. https://doi.org/10.3233/JSA-200384

Grunz, A., Memmert, D., & Perl, J. (2012). Tactical pattern recognition in soccer games by means of special self-organizing maps. *Human Movement Science*, *31*(2), 334–343. https://doi.org/10.1016/j.humov.2011.02.008

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, *13*(3), e1002106. https://doi.org/10.1371/journal.pbio.1002106

Herzog, S. M., & Hertwig, R. (2011). The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. *Judgment and Decision Making*, *6*(1), 58–72.

Heuer, A., Müller, C., & Rubner, O. (2010). Soccer: Is scoring goals a predictable Poissonian process? *EPL (Europhysics Letters)*, *89*(3), 38007. https://doi.org/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. https://doi.org/10.1140/epjb/e2009-00024-8

Heuer, A., & Rubner, O. (2012). How does the past of a soccer match influence its future? Concepts and statistical analysis. *PloS One*, *7*(11), e47678. https://doi.org/10.1371/journal.pone.0047678

Hill, I. D. (1974). Association Football and Statistical Inference. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *23*(2), 203–208. https://doi.org/10.2307/2347001

Ho, T. K. (1995). Random decision forests. In *Proceedings of the third International Conference on Document Analysis and Recognition: August 14-16, 1995, Montréal, Canada* (pp. 278–282). Los Alamitos, Calif: IEEE Computer Society Press. https://doi.org/10.1109/ICDAR.1995.598994

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting*, *32*(3), 896–913. https://doi.org/10.1016/j.ijforecast.2016.02.001

Horvat, T., & Job, J. (2020). The use of machine learning in sport outcome prediction: A review. *WIREs Data Mining and Knowledge Discovery*, *10*(5). https://doi.org/10.1002/widm.1380

Hosseini, M., Moore, J., Almaliki, M., Shahri, A., Phalp, K., & Ali, R. (2015). Wisdom of the Crowd within enterprises: Practices and challenges. *Computer Networks, 90*, 121–132. https://doi.org/10.1016/j.comnet.2015.07.004

Hubáček, O., & Šourek, G. (2020). Beating the market with a bad predictive model. *ArXiv Preprint ArXiv:2010.12508*.

Hubáček, O., Šourek, G., & Železný, F. (2019). Exploiting sports-betting market using machine learning. *International Journal of Forecasting*, *35*(2), 783–796. https://doi.org/10.1016/j.ijforecast.2019.01.001

Huberty, M. (2015). Can we vote with our tweet? On the perennial difficulty of election forecasting with social media. *International Journal of Forecasting*, *31*(3), 992–1007. https://doi.org/10.1016/j.ijforecast.2014.08.005

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

Jamieson, J. P. (2010). The Home Field Advantage in Athletics: A Meta-Analysis. *Journal of Applied Social Psychology*, *40*(7), 1819–1848. https://doi.org/10.1111/j.1559-1816.2010.00641.x

Jones, M. B. (2013). The home advantage in individual sports: An augmented review. *Psychology of Sport and Exercise*, *14*(3), 397–404. https://doi.org/10.1016/j.psychsport.2013.01.002

Jordet, G., Hartman, E., & Sigmundstad, E. (2009). Temporal links to performing under pressure in international soccer penalty shootouts. *Psychology of Sport and Exercise*, *10*(6), 621–627. https://doi.org/10.1016/j.psychsport.2009.03.004

Jordet, G., Hartman, E., Visscher, C., & Lemmink, K. A. P. M. (2007). Kicks from the penalty mark in soccer: The roles of stress, skill, and fatigue for kick outcomes. *Journal of Sports Sciences*, *25*(2), 121–129. https://doi.org/10.1080/02640410600624020

Jungherr, A., Jürgens, P., & Schoen, H. (2012). Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. “Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment”. *Social Science Computer Review*, *30*(2), 229–234. https://doi.org/10.1177/0894439311404119

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.

Kharde, V., & Sonawane, S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, *139*(11), 5–15. https://doi.org/10.5120/ijca2016908625

Kocher, M., Lenz, M., & Sutter, M. (2008). Performance under pressure-The case of penalty shootouts in football. In *Myths and facts about football: The economics and psychology of the world's greatest sport* (pp. 61–72). Cambridge Scholars Publishing.

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. https://doi.org/10.1111/rssa.12042

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. https://doi.org/10.1016/j.ijforecast.2018.10.011

Kovalchik, S. A. (2016). Searching for the GOAT of tennis win prediction. *Journal of Quantitative Analysis in Sports*, *12*(3), 127–138. https://doi.org/10.1515/jqas-2015-0059

Lago, C. (2009). The influence of match location, quality of opposition, and match status on possession strategies in professional association football. *Journal of Sports Sciences*, *27*(13), 1463–1469. https://doi.org/10.1080/02640410903131681

Lago-Ballesteros, J. [Joaquin], & Lago-Peñas, C. (2010). Performance in Team Sports: Identifying the Keys to Success in Soccer. *Journal of Human Kinetics*, *25*(2010), 85–91. https://doi.org/10.2478/v10078-010-0035-0

Lago-Peñas, C. (2012). The role of situational variables in analysing physical performance in soccer. *Journal of Human Kinetics*, *35*, 89–95. https://doi.org/10.2478/v10078-012-0082-9

Lago-Peñas, C., Lago-Ballesteros, J. [Joaquín], Dellal, A., & Gómez, M. (2010). Game-Related Statistics that Discriminated Winning, Drawing and Losing Teams from the Spanish Soccer League. *Journal of Sports Science & Medicine*, *9*(2), 288–293.

Lames, M. (2018). Chance involvement in goal scoring in football – an empirical approach. *German Journal of Exercise and Sport Research*, *48*(2), 278–286. https://doi.org/10.1007/s12662-018-0518-z

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. https://doi.org/10.1504/IJAPR.2013.052339

Le, H., Carr, P., Yue, Y., & Lucey, P. (2017). Data-Driven Ghosting using Deep Imitation Learning. In *Proceedings of the 11th Annual MIT Sloan Sports Analytics Conference 2017*. Boston, MA.

Leitch, G., & Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventual error measures. *The American Economic Review*, 580–590.

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. https://doi.org/10.1016/j.ijforecast.2009.10.001

Lepschy, H., Wäsche, H., & Woll, A. (2020). Success factors in football: an analysis of the German Bundesliga. *International Journal of Performance Analysis in Sport*, *20*(2), 150–164. https://doi.org/10.1080/24748668.2020.1726157

Lessmann, S., Sung, M.-C., & Johnson, J. E. (2010). Alternative methods of predicting competitive events: An application in horserace betting markets. *International Journal of Forecasting*, *26*(3), 518–536. https://doi.org/10.1016/j.ijforecast.2009.12.013

Linke, D., Link, D., & Lames, M. (2018). Validation of electronic performance and tracking systems EPTS under field conditions. *PloS One*, *13*(7), e0199519. https://doi.org/10.1371/journal.pone.0199519

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9020–9025. https://doi.org/10.1073/pnas.1008636108

Lorenzo-Martínez, M., Rein, R., Garnica-Caparrós, M., Memmert, D., & Rey, E. (2020). The Effect of Substitutions on Team Tactical Behavior in Professional Soccer. *Research Quarterly for Exercise and Sport*, 1–9. https://doi.org/10.1080/02701367.2020.1828563

Low, B., Coutinho, D., Gonçalves, B., Rein, R., Memmert, D., & Sampaio, J. (2020). A Systematic Review of Collective Tactical Behaviours in Football Using Positional Data. *Sports Medicine (Auckland, N.Z.)*, *50*(2), 343–385. https://doi.org/10.1007/s40279-019-01194-7

Luckner, S., Schröder, J., & Slamka, C. (2008). On the Forecast Accuracy of Sports Prediction Markets. In H. Gimpel (Ed.), *Lecture Notes in Business Information Processing: Vol. 2. Negotiation, auctions, and market engineering: International seminar, Dagstuhl Castle, Germany, November 12-17, 2006, rev. selected papers* (1st ed., Vol. 2, pp. 227–234). Berlin: Springer. https://doi.org/10.1007/978-3-540-77554-6_17

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, *36*(3), 109–118.

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

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, *27*, 16–32. https://doi.org/10.1016/j.cosrev.2017.10.002

Matos, R., Monteiro, D., Antunes, R., Mendes, D., Botas, J., Clemente, J., & Amaro, N. (2021). Home-Advantage during COVID-19: An Analysis in Portuguese Football League. *International Journal of Environmental Research and Public Health*, *18*(7). https://doi.org/10.3390/ijerph18073761

Mauro, A. de, Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, *65*(3), 122–135. https://doi.org/10.1108/LR-06-2015-0061

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

McHale, I., & Swartz, T. (2019). Editorial: Forecasting in sports. *International Journal of Forecasting*, *35*(2), 710–711. https://doi.org/10.1016/j.ijforecast.2019.01.002

Memmert, D., Hüttermann, S., Hagemann, N., Loffing, F., & Strauss, B. (2013). Dueling in the penalty box: evidence-based recommendations on how shooters and goalkeepers can win penalty shootouts in soccer. *International Review of Sport and Exercise Psychology*, *6*(1), 209–229. https://doi.org/10.1080/1750984X.2013.811533

Memmert, D., Lemmink, K. A. P. M., & Sampaio, J. (2017). Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine (Auckland, N.Z.)*, *47*(1), 1–10. https://doi.org/10.1007/s40279-016-0562-5

Memmert, D., Raabe, D., Schwab, S., & Rein, R. (2019). A tactical comparison of the 4-2-3-1 and 3-5-2 formation in soccer: A theory-oriented, experimental approach based on positional data in an 11 vs. 11 game set-up. *PloS One*, *14*(1), e0210191. https://doi.org/10.1371/journal.pone.0210191

Michailidis, Y., Michailidis, C., & Primpa, E. (2013). Analysis of goals scored in European Championship 2012. *Journal of Human Sport and Exercise*, *8*(2), 367–375. https://doi.org/10.4100/jhse.2012.82.05

Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, *5*(4), 213–222. https://doi.org/10.1007/s41060-017-0093-7

Mueller, S. Q. (2020). Pre- and within-season attendance forecasting in Major League Baseball: a random forest approach. *Applied Economics*, *52*(41), 4512–4528. https://doi.org/10.1080/00036846.2020.1736502

Neave, N., & Wolfson, S. (2003). Testosterone, territoriality, and the ‘home advantage’. *Physiology & Behavior*, *78*(2), 269–275. https://doi.org/10.1016/S0031-9384(02)00969-1

Nederlandse Online Gambling Associatie (2015). Sports betting: commercial and integrity issues. Retrieved from https://no-ga.nl/wp-content/uploads/2020/08/Sports-Betting-Commercial-and-Integrity-Issues.pdf

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

Nevill, A. M., Newell, S. M., & Gale, S. (1996). Factors associated with home advantage in English and Scottish soccer matches. *Journal of Sports Sciences*, *14*(2), 181–186. https://doi.org/10.1080/02640419608727700

Noël, B., Furley, P., van der Kamp, J., Dicks, M., & Memmert, D. (2015). The development of a method for identifying penalty kick strategies in association football. *Journal of Sports Sciences*, *33*(1), 1–10. https://doi.org/10.1080/02640414.2014.926383

O’donoghue, P., & Robinson, G. (2016). Score-line effect on work-rate in English FA Premier League soccer. *International Journal of Performance Analysis in Sport*, *16*(3), 910–923. https://doi.org/10.1080/24748668.2016.11868938

Ofoghi, B., Zeleznikow, J., MacMahon, C., & Raab, M. (2013). Data Mining in Elite Sports: A Review and a Framework. *Measurement in Physical Education and Exercise Science*, *17*(3), 171–186. https://doi.org/10.1080/1091367X.2013.805137

Ottaviani, M., & Sorensen, P. N. (2008). The favorite-longshot bias: An overview on the main explanations. In *Handbook of Sports and Lottery markets* (pp. 83–101). Elsevier.

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. https://doi.org/10.1016/j.actpsy.2006.07.002

Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., Ferragina, P., Pedreschi, D., & Giannotti, F. (2019). A public data set of spatio-temporal match events in soccer competitions. *Scientific Data*, *6*(1), 236. https://doi.org/10.1038/s41597-019-0247-7

Passos, P., Araújo, D., & Davids, K. (2013). Self-organization processes in field-invasion team sports : Implications for leadership. *Sports Medicine (Auckland, N.Z.)*, *43*(1), 1–7. https://doi.org/10.1007/s40279-012-0001-1

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. https://doi.org/10.1016/j.ijforecast.2017.08.002

Pettersson-Lidbom, P., & Priks, M. (2010). Behavior under social pressure: Empty Italian stadiums and referee bias. *Economics Letters*, *108*(2), 212–214. https://doi.org/10.1016/j.econlet.2010.04.023

Piryani, R., Madhavi, D., & Singh, V. K. (2017). Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Information Processing & Management*, *53*(1), 122–150. https://doi.org/10.1016/j.ipm.2016.07.001

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences*, *4*(3), 237–248. https://doi.org/10.1080/02640418608732122

Pollard, R. (2008). Home Advantage in Football: A Current Review of an Unsolved Puzzle. *The Open Sports Sciences Journal*, *1*(1), 12–14. https://doi.org/10.2174/1875399X00801010012

Pollard, R., & Gomez, M. A. (2014). Components of home advantage in 157 national football leagues worldwide. *International Journal of Sport and Exercise Psychology*, *12*(3), 218–233.

Pollard, R., & Pollard, G. (2005). Long-term trends in home advantage in professional team sports in North America and England (1876-2003). *Journal of Sports Sciences*, *23*(4), 337–350. https://doi.org/10.1080/02640410400021559

Pratas, J. M., Volossovitch, A., & Carita, A. I. (2018). Goal scoring in elite male football: A systematic review. *Journal of Human Sport and Exercise*, *13*(1), 218–230. https://doi.org/10.14198/jhse.2018.131.19

Raval, K. M. (2012). Data mining techniques. *International Journal of Advanced Reserach in Computer Science and Software Engineering*, *2*(10), 439–442.

Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *SpringerPlus*, *5*(1), 1410. https://doi.org/10.1186/s40064-016-3108-2

Rein, R., Raabe, D., & Memmert, D. (2017). "Which pass is better?" Novel approaches to assess passing effectiveness in elite soccer. *Human Movement Science*, *55*, 172–181. https://doi.org/10.1016/j.humov.2017.07.010

Riedl, D., Strauss, B., Heuer, A., & Rubner, O. (2015). Finale furioso: Referee-biased injury times and their effects on home advantage in football. *Journal of Sports Sciences*, *33*(4), 327–336. https://doi.org/10.1080/02640414.2014.944558

Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernàndez, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PloS One*, *13*(7), e0201264. https://doi.org/10.1371/journal.pone.0201264

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. In W. W. Smari (Ed.), *2013 International Conference on Collaboration Technologies and Systems (CTS 2013): San Diego, California, USA, 20 - 24 May 2013 ; [including symposia and workshops* (pp. 42–47). Piscataway, NJ: IEEE. https://doi.org/10.1109/CTS.2013.6567202

Sanderson, J. (2014). What Do We Do With Twitter? *Communication & Sport*, *2*(2), 127–131. https://doi.org/10.1177/2167479514527429

Savelsbergh, G. J. P., van der Kamp, J., Williams, A. M., & Ward, P. (2005). Anticipation and visual search behaviour in expert soccer goalkeepers. *Ergonomics*, *48*(11-14), 1686–1697. https://doi.org/10.1080/00140130500101346

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: the real-world use of big data: How innovative enterprises extract value from uncertain data, Executive Report. *IBM Institute for Business Value and Said Business School at the University of Oxford*.

Schumaker, R. P., Jarmoszko, A. T., & Labedz, C. S. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, *88*, 76–84. https://doi.org/10.1016/j.dss.2016.05.010

Scoppa, V. (2008). Are subjective evaluations biased by social factors or Are subjective evaluations biased by social factors or connections. An economic analysis of soccer referee decisions. *Empirical Economics*, *35*(1), 123–140.

Scoppa, V. (2021). Social pressure in the stadiums: Do agents change behavior without crowd support? *Journal of Economic Psychology*, *82*, 102344. https://doi.org/10.1016/j.joep.2020.102344

Seidl, T., Cherukumudi, A., Hartnett, A., Carr, P., & Lucey, P. (2018). Bhostgusters: Realtime interactive play sketching with synthesized nba defenses. In *12 th Annual MIT Sloan Sports Analytics Conference*.

Sgrò, F., Aiello, F., Casella, A., & Lipoma, M. (2017). The effects of match-playing aspects and situational variables on achieving score-box possessions in Euro 2012 Football Championship. *Journal of Human Sport and Exercise*, *12*(1), 58–72. https://doi.org/10.14198/jhse.2017.121.05

Sheffer, M. L., & Schultz, B. (2010). Paradigm Shift or Passing Fad? Twitter and Sports Journalism. *International Journal of Sport Communication*, *3*(4), 472–484. https://doi.org/10.1123/ijsc.3.4.472

Siegle, M., Stevens, T., & Lames, M. (2013). Design of an accuracy study for position detection in football. *Journal of Sports Sciences*, *31*(2), 166–172. https://doi.org/10.1080/02640414.2012.723131

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. https://doi.org/10.1016/j.ijforecast.2007.05.003

Sors, F., Grassi, M., Agostini, T., & Murgia, M. (2020). The sound of silence in association football: Home advantage and referee bias decrease in matches played without spectators. *European Journal of Sport Science*, 1–9. https://doi.org/10.1080/17461391.2020.1845814

Soto Valero, C. (2016). Predicting Win-Loss outcomes in MLB regular season games – A comparative study using data mining methods. *International Journal of Computer Science in Sport*, *15*(2), 91–112. https://doi.org/10.1515/ijcss-2016-0007

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. https://doi.org/10.1002/for.1091

Sports Integrity Initiative (2017). Retrieved from https://www.sportsintegrityinitiative.com/uefas-betting-fraud-detection-system-cas-regard-monitoring-tool/

Stekler, H. O., Sendor, D., & Verlander, R. (2010). Issues in sports forecasting. *International Journal of Forecasting*, *26*(3), 606–621. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/j.ijforecast.2011.01.004

Surowiecki, J. (2005). *The wisdom of crowds*: Anchor.

Sutter, M., & Kocher, M. G. (2004). Favoritism of agents – The case of referees' home bias. *Journal of Economic Psychology*, *25*(4), 461–469. https://doi.org/10.1016/S0167-4870(03)00013-8

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267–307. https://doi.org/10.1162/COLI_a_00049

Taylor, J. W., & Buizza, R. (2004). A comparison of temperature density forecasts from GARCH and atmospheric models. *Journal of Forecasting*, *23*(5), 337–355. https://doi.org/10.1002/for.917

Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010). Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. *Journal of Sports Sciences*, *28*(3), 245–255. https://doi.org/10.1080/02640410903502766

Tenga, A., Ronglan, L. T., & Bahr, R. (2010). Measuring the effectiveness of offensive match-play in professional soccer. *European Journal of Sport Science*, *10*(4), 269–277. https://doi.org/10.1080/17461390903515170

Tilp, M., & Thaller, S. (2020). Covid-19 Has Turned Home Advantage Into Home Disadvantage in the German Soccer Bundesliga. *Frontiers in Sports and Active Living*, *2*, 593499. https://doi.org/10.3389/fspor.2020.593499

Timmermann, A. (2000). Density forecasting in economics and finance. *Journal of Forecasting*, *19*(4), 231–234.

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of the International AAAI Conference on Web and Social Media*, *4*(1). Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14009

Twitter (2021). Retrieved from https://twitter.com/

Twitter API (2021). Retrieved from https://developer.twitter.com/en/docs/twitter-api

Twitter Blog (2013). New Tweets per second record, and how! Retrieved from https://blog.twitter.com/engineering/en_us/a/2013/new-tweets-per-second-record-and-how.html

Twitter Blog (2014). The 2014 #YearOnTwitter. Retrieved from https://blog.twitter.com/official/en_us/a/2014/the-2014-yearontwitter.html

Twitter Blog (2016). Events #ThisHappened2016. Retrieved from https://blog.twitter.com/en_gb/topics/events/2016/thishappened2016.html

Unkelbach, C., & Memmert, D. (2010). Crowd noise as a cue in referee decisions contributes to the home advantage. *Journal of Sport & Exercise Psychology, 32*(4), 483–498. https://doi.org/10.1123/jsep.32.4.483

Van Reeth, D. (2019). Forecasting Tour de France TV audiences: A multi-country analysis. *International Journal of Forecasting, 35*(2), 810–821. https://doi.org/10.1016/j.ijforecast.2018.06.003

Vaughan Williams, L., & Stekler, H. O. (2010). Sports forecasting. *International Journal of Forecasting*, *26*(3), 445–447. https://doi.org/10.1016/j.ijforecast.2009.12.005

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond " p < 0.05". *The American Statistician*, *73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

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

Wheatcroft, E. (2019, August 23). *Evaluating probabilistic forecasts of football matches: The case against the Ranked Probability Score*. Retrieved from http://arxiv.org/pdf/1908.08980v1

Wheatcroft, E. (2020). A profitable model for predicting the over/under market in football. *International Journal of Forecasting*, *36*(3), 916–932. https://doi.org/10.1016/j.ijforecast.2019.11.001

Williams, L. V. (1999). Information Efficiency in Betting Markets: a Survey. *Bulletin of Economic Research*, *51*(1), 1–39. https://doi.org/10.1111/1467-8586.00069

Witkemper, C., Lim, C. H., & Waldburger, A. (2012). Social Media and Sports Marketing: Examining the Motivations and Constraints of Twitter Users. *Sport Marketing Quarterly*. (21), 170–183.

Wolfers, J., & Leigh, A. (2002). Three Tools for Forecasting Federal Elections: Lessons from 2001. *Australian Journal of Political Science*, *37*(2), 223–240. https://doi.org/10.1080/10361140220148115

Wolfers, J., & Zitzewitz, E. (2004). Prediction Markets. *Journal of Economic Perspectives*, *18*(2), 107–126. https://doi.org/10.1257/0895330041371321

Wolfson, S., Wakelin, D., & Lewis, M. (2005). Football supporters' perceptions of their role in the home advantage. *Journal of Sports Sciences*, *23*(4), 365–374. https://doi.org/10.1080/02640410400021567

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear". *Procedia - Social and Behavioral Sciences*, *26*, 55–62. https://doi.org/10.1016/j.sbspro.2011.10.562

Zou, Q., Song, K., & Shi, J. (2020). A Bayesian In-Play Prediction Model for Association Football Outcomes. *Applied Sciences*, *10*(8), 2904. https://doi.org/10.3390/app10082904

8. Appendix

For copyright reasons, the Appendix of the accepted and published version of this dissertation does not include any full article. Subsequently, only title, reference and abstract of each article are included. All manuscripts were published in English and thus the abstracts are presented in the original language of creation.

Appendix I: Article I

Title:

Forecasting the outcomes of sports events: A review

Reference:

Wunderlich, F., & Memmert, D. (2021). Forecasting the outcomes of sports events: A review. *European Journal of Sport Science*, 21(7), 944-957. https://doi.org/10.1080/17461391.2020.1793002.

[IF 2019: 2.6, 24/85 in Sports Sciences, 72th percentile, Q2]

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.

Appendix II: Article II

Title:

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

Reference:

Wunderlich, F., & Memmert, D. (2020). Are betting returns a useful measure of accuracy in (sports) forecasting?. *International Journal of Forecasting*, 36(2), 713-722. https://doi.org/10.1016/j.ijforecast.2019.08.009.

[IF 2019: 2.8, 66/373 in Economics, 83th percentile, Q1]

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.

Appendix III: Article III

<u>Title:</u>

The Betting Odds Rating System: Using soccer forecasts to forecast soccer.

<u>Reference:</u>

Wunderlich, F., & Memmert, D. (2018). The Betting Odds Rating System: Using soccer forecasts to forecast soccer. *PloS one*, 13(6). https://doi.org/10.1371/journal.pone.0198668.

[IF 2018: 2.8, 24/69 in Multidisciplinary Sciences, 66th percentile, Q2]

Abstract

Betting odds are frequently found to outperform mathematical models in sports related forecasting tasks, however the factors contributing to betting odds are not fully traceable and in contrast to rating-based forecasts no straightforward measure of team-specific quality is deducible from the betting odds. The present study investigates the approach of combining the methods of mathematical models and the information included in betting odds. A soccer forecasting model based on the well-known ELO rating system and taking advantage of betting odds as a source of information is presented. Data from almost 15.000 soccer matches (seasons 2007/2008 until 2016/2017) are used, including both domestic matches (English Premier League, German Bundesliga, Spanish Primera Division and Italian Serie A) and international matches (UEFA Champions League, UEFA Europe League). The novel betting odds based ELO model is shown to outperform classic ELO models, thus demonstrating that betting odds prior to a match contain more relevant information than the result of the match itself. It is shown how the novel model can help to gain valuable insights into the quality of soccer teams and its development over time, thus having a practical benefit in performance analysis. Moreover, it is argued that network based approaches might help in further improving rating and forecasting methods.

Appendix IV: Article IV

Title:

Almost a lottery: the influence of team strength on success in penalty shootouts.

Reference:

Wunderlich, F., Berge, F., Memmert, D., & Rein, R. (2020). Almost a lottery: the influence of team strength on success in penalty shootouts. *International Journal of Performance Analysis in Sport*, *20*(5), 857-869. https://doi.org/10.1080/24748668.2020.1799171.

[IF 2019: 1.5, 60/85 in Sports Sciences, 30th percentile, Q3]

Abstract

Three aspects of penalty shootouts that have not been examined in the literature so far are 1) the influence of overall team strength on penalty success, 2) the viability of a forecasting model for penalty shootouts, and 3) the existence of a penalty-specific home advantage. To this end, a sample consisting of 1067 penalty shootouts from 14 cup competitions was investigated. Team strength was estimated based on betting odds and results show that stronger teams win significantly more shootouts compared to weaker teams. A forecasting model, based on an out-of-sample approach, suggests that the effect of team strength on success is rather small as the winning probability remains around 40 % even for very weak teams against very strong teams. Thus, for weaker teams it seems advantageous to focus on drawing a game against stronger teams as their probability of success is much greater during a penalty shootout compared to normal game play. In contrast to the robust evidence of a home advantage during normal game play the results further indicate an absence for a home (or away) advantage during penalty shootouts. The results presented are therefore highly valuable for coaches in supporting clear tactical recommendations.

Appendix V: Article V

Title:

How does spectator presence affect football? Home advantage remains in European top-class football matches played without spectators during the COVID-19 pandemic.

Reference:

Wunderlich, F., Weigelt, M., Rein, R., & Memmert, D. (2021). How does spectator presence affect football? Home advantage remains in European top-class football matches played without spectators during the COVID-19 pandemic. *Plos one*, 16(3). https://doi.org/10.1371/journal.pone.0248590

[IF 2019: 2.7, 27/71 in Multidisciplinary Sciences, 62th percentile, Q2]

Abstract

The present paper investigates factors contributing to the home advantage, by using the exceptional opportunity to study professional football matches played in the absence of spectators due to the COVID-19 pandemic in 2020. More than 40,000 matches before and during the pandemic, including more than 1,000 professional matches without spectators across the main European football leagues, have been analyzed. Results support the notion of a crowd-induced referee bias as the increased sanctioning of away teams disappears in the absence of spectators with regard to fouls ($p < .001$), yellow cards ($p < .001$), and red cards ($p < .05$). Moreover, the match dominance of home teams decreases significantly as indicated by shots ($p < .001$) and shots on target ($p < .01$). In terms of the home advantage itself, surprisingly, only a non-significant decrease is found. While the present paper supports prior research with regard to a crowd-induced referee bias, spectators thus do not seem to be the main driving factor of the home advantage. Results from amateur football, being naturally played in absence of a crowd, provide further evidence that the home advantage is predominantly caused by factors not directly or indirectly attributable to a noteworthy number of spectators.

Appendix VI: Article VI

Title:

The influence of randomness on goals in football decreases over time. An empirical analysis of randomness involved in goal scoring in the English Premier League.

Reference:

Wunderlich, F., Seck, A., & Memmert, D. (2021): The influence of randomness on goals in football decreases over time. An empirical analysis of randomness involved in goal scoring in the English Premier League. *Journal of Sports Sciences*. *39*(20), 2322-2337. https://doi.org/10.1080/02640414.2021.1930685.

[IF 2019: 2.6, 27/85 in Sports Sciences, 69th percentile, Q2]

Abstract

Performance analysis in football predominantly focuses on systematic contributions to success, thus neglecting the role of randomness. The present paper pursues a direct approach to quantify and analyse randomness in football by identifying random influences in the goal scoring process. The dataset includes all matches from the seasons 12/13 to 18/19 of the English Premier League, adding up to a total of 7,263 goals, that were checked for the occurrence of six variables of random influence. Additionally, the influence of nine situational variables was investigated. Results show that randomness was present for almost 50% of all goals. Moreover, it was demonstrated that the proportion of random goals decreased over the seven seasons ($p < .001$), is more pronounced for weaker teams ($p < .05$) as well as if the current scoreline is a draw ($p < .05$) and depends on the match situation (open play, freekick, corner, penalty). An improved understanding of randomness in football has important implications for both researchers and practitioners. Performance analysts should acknowledge randomness as a crucial factor to distinguish clearly between performance and success. Coaches could even consider the conscious creation of uncontrollable situations as a possible tactic to provoke random influences on goal scoring.

Appendix VII: Article VII

<u>Title:</u>

Innovative Approaches in Sports Science - Lexicon-Based Sentiment Analysis as a Tool to Analyze Sports-Related Twitter Communication

<u>Reference:</u>

Wunderlich, F., & Memmert, D. (2020). Innovative Approaches in Sports Science - Lexicon-Based Sentiment Analysis as a Tool to Analyze Sports-Related Twitter Communication. *Applied Sciences*, 10(2), 431. https://doi.org/ 10.3390/app10020431.

[IF 2019: 2.5, 32/91 in Engineering, Multidisciplinary, 65th percentile, Q2]

Abstract

Sentiment analysis refers to the algorithmic extraction of subjective information from textual data and – driven by the increasing amount of online communication – has become one of the fastest growing research areas in computer science with applications in several domains. Although sports events such as football matches are accompanied by a huge public interest and large amount of related online communication, social media analysis in general and sentiment analysis in particular are almost unused tools in sports science so far. The present study tests the feasibility of lexicon-based tools of sentiment analysis with regard to football-related textual data on the microblogging platform Twitter. The sentiment of a total of 10,000 tweets with reference to ten top-level football matches was analyzed both manually by human annotators and algorithmically by means of publicly available sentiment analysis tools. Results show that the general sentiment of realistic sets (1000 tweets with a proportion of 60% having the same polarity) can be classified correctly with more than 95% accuracy. The present paper demonstrates that sentiment analysis can be an effective and useful tool for sports-related content and is intended to stimulate the increased use of and discussion on sentiment analysis in sports science.

Appendix VIII: Article VIII

<u>Title:</u>

A big data analysis of Twitter data during premier league matches: do tweets contain information valuable for in-play forecasting of goals in football?

<u>Reference:</u>

Wunderlich, F., & Memmert, D. (2022). A big data analysis of Twitter data during premier league matches: do tweets contain information valuable for in-play forecasting of goals in football?. *Social Network Analysis and Mining*, 12(1), 1-15.

[Impact 2020: 3.9, Q1 in Media Technology according to Scimago]

Abstract

Data-related analysis in football increasingly benefits from Big Data approaches and machine learning methods. One relevant application of data analysis in football is forecasting, which relies on understanding and accurately modelling the process of a match. The present paper tackles two neglected facets of forecasting in football: Forecasts on the total number of goals and in-play forecasting (forecasts based on within-match information). Sentiment analysis techniques were used to extract the information reflected in almost two million tweets from more than 400 Premier League matches. By means of wordclouds and timely analysis of several tweet-based features, the Twitter communication over the full course of matches and shortly before and after goals was visualized and systematically analysed. Moreover, several forecasting models including a random forest model have been used to obtain in-play forecasts. Results suggest that in-play forecasting of goals is highly challenging, and in-play information does not improve forecasting accuracy. An additional analysis of goals from more than 30,000 matches from the main European football leagues supports the notion that the predictive value of in-play information is highly limited compared to pre-game information. This is a relevant result for coaches, match analysts and broadcasters who should not overestimate the value of in-play information. The present study also sheds light on how the perception and behaviour of Twitter users change over the course of a football match. A main result is that the sentiment of Twitter users decreases when the match progresses, which might be caused by an unjustified high expectation of football fans before the match.

Appendix IX: Article IX

Title:

In-Play forecasting in football using event and positional data

Reference:

Klemp, M.*, **Wunderlich, F.***, & Memmert D. (2021). In-Play forecasting in football using event and positional data. *Scientific reports*, 11(1), 1-10.

*both authors contributed equally to this manuscript

[IF 2020: 4.4, 17/72 in Multidisciplinary Sciences, 77th percentile, Q1]

Abstract

Two highly relevant aspects of football, namely forecasting of results and performance analysis by means of performance indicators, are combined in the present study by analysing the value of in-play information in terms of event and positional data in forecasting the further course of football matches. Event and positional data from 50 matches, including more than 300 million datapoints were used to extract a total of 18 performance indicators. Moreover, goals from more than 30,000 additional matches have been analysed. Results suggest that surprisingly goals do not possess any relevant informative value on the further course of a match, if controlling for pre-game market expectation by means of betting odds. Performance indicators based on event and positional data have been shown to possess more informative value than goals, but still are not sufficient to reveal significant predictive value in-play. The present results are relevant to match analysts and bookmakers who should not overestimate the value of in-play information when explaining match performance or compiling in-play betting odds. Moreover, the framework presented in the present study has methodological implications for performance analysis in football, as it suggests that researchers should increasingly segment matches by scoreline and control carefully for general team strength.

Appendix X: Article X

<u>Title:</u>

Evaluation of the Technical Performance of Football Players in the UEFA Champions League

<u>Reference:</u>

Yi, Q., Gómez-Ruano, M. Á., Liu, H., Zhang, S., Gao, B., **Wunderlich, F.**, & Memmert, D. (2020). Evaluation of the Technical Performance of Football Players in the UEFA Champions League. *International Journal of Environmental Research and Public Health*, 17(2), 604, https://doi.org/10.3390/ijerph17020604.

[IF 2019: 2.8, 58/193 in Public, Environmental & Occupational Health, 70th percentile, Q2]

Abstract

The aim of this study was to assess the technical match performance of top-class football players from a long-term perspective. Technical performance profiles of players according to five playing positions (*central defender, full back, wide midfielder, central midfielder*, *forward*) and five situational variables (*competition stage, match location, quality of team, quality of opponent, match outcome*) were established. Technical match data of players in the UEFA Champions League from season 2009-2010 to 2016-2017 were analysed. The true effects of positional and situational variables on players' technical performance were evaluated by the non-clinical magnitude-based inference. Results showed that the effect of *competition stage* on player's performance was negligible. Situational variables related to team strength (*quality of team, quality of opponent* and *match outcome*) revealed the strongest effects on player's performance while the effect of *match location* was relatively lower. The technical performance of wide midfielders and forwards were more susceptible to the competing contexts when compared with central defenders, full backs and central midfielders. Differences of players' match performance could mainly be identified in variables related to goal scoring, passing and organising, while there were less differences in most of attacking and defending related variables.

Appendix XI: Article XI

<u>Title:</u>

Situational and Positional Effects on the Technical Variation of Players in the UEFA Champions League

<u>Reference:</u>

Yi, Q., Gómez-Ruano, M. Á., Liu, H., Gao, B., **Wunderlich, F.**, & Memmert, D. (2020). Situational and Positional Effects on the Technical Variation of Players in the UEFA Champions League. *Frontiers in Psychology*, 11, 1201, https://doi.org/10.3389/fpsyg.2020.01201.

[IF 2019: 2.1, 45/138 in Psychology, Multidisciplinary, 68th percentile, Q2]]

Abstract

This study aimed to identify the situational and positional effects on the variation of players' technical performance in the UEFA Champions League from a long-term perspective. The technical performance of full match observations from outfield players in the UEFA Champions League from season 2009/10 to 2016/17 was analysed. The coefficient of variation of each variable of each player in each season was calculated to evaluate the match-to-match variation of technical performance. The variation of technical performance between players was compared across five playing positions and five situational variables using the non-clinical magnitude-based inference. Results showed that variables related to goal scoring, passing and organising from five playing positions showed a relatively higher variation among five competing contexts (ES: -0.72±0.38 -0.82±0.61). Quality of team, quality of opponent and match outcome showed relatively greater influences than competition stage and match location on the variation of player's technical performance (ES: -0.72±0.38 -0.57±0.56). The technical performances of wide players (full backs and wide midfielders) were more variable between the group and knockout stage (ES: -0.37±0.32 -0.28±0.19). This study provides an important understanding of the associations among the variation of technical indicators, playing positions and situational variables. These profiles of technical variation could be used by coaches and analysts for talent identification, player recruitment, pre-match preparation, and post-match evaluation.