JOURNAL OF SPORTS SCIENCES, 2016
http://dx.doi.org/10.1080/02640414.2016.1218040

Analysis of the predictive qualities of betting odds and FIFA World Ranking: evidence from the 2006, 2010 and 2014 Football World Cups

Fabian Wunderlich and Daniel Memmert

Institute of Cognitive and Team/Racket Sport Research, German Sport University Cologne, Cologne, Germany

ABSTRACT
The present study aims to investigate the ability of a new framework enabling to derive more detailed model-based predictions from ranking systems. These were compared to predictions from the bet market including data from the World Cups 2006, 2010, and 2014. The results revealed that the FIFA World Ranking has essentially improved its predictive qualities compared to the bet market since the mode of calculation was changed in 2006. While both predictors were useful to obtain accurate predictions in general, the world ranking was able to outperform the bet market significantly for the World Cup 2014 and when the data from the World Cups 2010 and 2014 were pooled. Our new framework can be extended in future research to more detailed prediction tasks (i.e., predicting the final scores of a match or the tournament progress of a team).

ARTICLE HISTORY
Accepted 21 July 2016

KEYWORDS
Football forecasting; betting odds; FIFA world ranking; World Cup; poisson model

The prediction of complex future developments, such as stock or sport forecasting, is a huge field, in which enormous financial turnovers are transacted every day. For example, in the area of stock forecasting, experts seem to attain, at best, similar levels of accuracy as completely laypeople (e.g., Hershey & Walsh, 2000/2001; Önkal & Muradoğlu, 1994; Yates & Tschirhart, 2006; Yates, McDaniel, & Brown, 1991; but see Önkal, Yates, Simga-Mugan, & Öztin, 2003; or Shanteau, 1992). This result is not surprising because the principle of financial economics assumes that stock prices should be more or less unpredictable (e.g., Fama, 1998; Törngren & Montgomery, 2003).

Football results also seem to possess a substantial degree of unpredictability and therefore discussing who will win the next match or the whole tournament is naturally an ambitious task for every person concerned with this sport. During the FIFA World Cup 2014 in Brazil, various groups like sport commentators, journalists, former football players, betting pools and communities, private and commercial betters, as well as fans tried to forecast the outcome of the games in a work-related or private surrounding over and over again. Making a prediction is indeed easy; however, making an accurate prediction is obviously more complicated and even evaluating the quality of a prediction is more complex than usually assumed. One reason is because of the complexity of the match, thus even so-called soccer experts (i.e., former soccer players) or professional betters cannot make any better predictions in general than naïve individuals (e.g., Andersson, Edman, & Ekman, 2005; Andersson, Memmert, & Popowic, 2009; Camerer & Johnson, 1991; but see Pachur & Biele, 2007).

For some time now one has tried to deliver more precise predictions by using various statistical prediction models. Betting odds as well as ranking systems are frequently examined to predict outcomes of sport events. Numerous findings approved the excellent predictive quality of betting odds (e. g., Spann & Skiera, 2009). Complex statistical models failed to outperform betting odds (Forrest, Goddard, & Simmons, 2005), moreover betting odds were found to be outperforming ranking systems in predictive tasks in football, tennis, and American football (Boulier & Stekler, 2003; Leitner, Zeileis, & Hornik, 2010; Scheibehenne & Bröder, 2007). However, there seems to be a general difficulty as rankings are usually not designed to enable predictions. Therefore, rankings are mostly used to merely classify teams (e.g., Leitner et al., 2010) or select the presumable winner of a match (e.g., Suzuki & Ohmori, 2008) whereas little models are known to deduce a percental prediction from a ranking (e.g., Dyte & Clarke, 2000; McHale & Davies, 2008). Moreover, to the best of our knowledge it has not been examined yet whether the FIFA World Ranking has improved its predictive qualities since the mode of calculation was changed in 2006.

Literature review and hypothesis development

Predicting the outcome of sport events is a quotidian activity, in which many people are willing to bet money on their predictions. Despite this fact, only scarce research on the effectiveness of complex statistical models to forecast the outcome of sport events seems to exist (cf. Dobson & Goddard, 2001; Goddard, 2005).

Past attempts to develop statistical models for predicting football matches have only mildly benefitted the predictive power and have thus proven that the forecast of match results is highly complex (cf. Dobson & Goddard, 2001; Goddard, 2005). However, it is noteworthy that objective variables like rankings (seedings) and betting odds are quite valuable

CONTACT Fabian Wunderlich Fabian.Wunderlich@gmx.de

predictors of the outcome of basketball, tennis, and football matches (Andersson et al., 2005; Boulier & Stekler, 1999; Forrest et al., 2005).

If the bet market can be considered to be effective and betting odds allow accurate predictions is not clear from the outset. This is due to the fact that the emergence of betting odds is a complex process and can be influenced by the opinion of professionals (i.e., bookmakers) as well as laypeople (i.e., punters) with various degrees of knowledge. It can hardly be quantified which group has which influence on the betting odds. Referring to the above mentioned results showing that experts do not necessarily predict better than laypeople, it is not even clear which group would be more successful in correctly judging the betting odds. Moreover the partially irrational behaviour of sport betters (e.g.; Andersson et al., 2005; Nilsson & Andersson, 2010) can indicate that betting odds do not necessarily need to be ideal.

Nevertheless, there is current evidence on the generally good prediction quality of betting odds (Scheibehenne & Bröder, 2007; Spann & Skiera, 2009) and even indications on continuous improvements of prediction quality of betting odds (e.g., Forrest et al., 2005; Štrumbelj & Šikonja, 2010).

Analyses exist which examined both betting odds and rating systems such as the FIFA World Ranking (Leitner et al., 2010). It was shown that classifications derived from betting odds outperformed those derived from ranking systems when predicting the results of the European Championship 2008. However, no percental predictions derived from the FIFA World Ranking were examined.

Motivated by the aforementioned paucity of research, the lack of studies with more complex sport forecasting models based on ratings and inconsistent findings of prior studies, the present paper aims to investigate the ability of a new framework enabling to derive more detailed predictions from ranking systems comparing these results to predictions from the bet market, also including the recent new data from the World Cup 2014. In particular, we took advantage of the last 3 World Cups in football to evaluate this topic and extended previous research in different ways. For our approach, we used easily accessible information to derive predictions for football matches from 3 different World Cup tournaments (2006, 2010, and 2014). Two different ways of predicting football results by using the following predictors were analysed: (a) the betting odds obtained from the world's largest bet exchange Betfair (referred to as bet market or BET) representing expectations for the future and (b) the official FIFA World Ranking (referred to as world ranking or RANK) reflecting results in the past.

Betting odds and rankings such as the FIFA World Ranking are regularly used to derive predictions and to evaluate their quality (Leitner et al., 2010; Scheibehenne & Bröder, 2007). As an absolute minimum criterion for a useful predictor, we expected to assure predictions that are able to outperform predictions made without any information. Based on prior research, we assumed that both predictors clearly fulfilled this task for all 3 World Cups, the first time for 2014, and expressed this expectation in Hypothesis 1.

The FIFA World Ranking until 2006 was criticised frequently and for various reasons including a poor differentiation of points between the teams, the inclusion of results from the previous 8 years and hence a deficient reflection of the teams' strengths. As a reaction to this reasonable criticism and attempt to improve the world ranking the mode of calculation was changed after the World Cup 2006. The ranking since 2006 is based on results from the previous 4 years and ensures a clearer differentiation of points between the teams. On the one hand, this implies a limited comparability between the tournaments in 2010, 2014, and 2006. On the other hand, the change of calculation mode might have led to an improvement or deterioration in predictive qualities. As one important purpose of the change was to improve the reflection of teams' strengths we expected an improvement, which we formulated in Hypothesis 2. To our knowledge, this question is investigated for the first time.

The FIFA World Ranking has a retrospective character structurally. Previous matches of the national teams are considered and used to calculate points. This procedure is naturally seen as an important purpose of the ranking to reward previous achievement of the teams. However, the mode of calculation does not include any further information that might be helpful to optimally predict future performances (e.g., home advantage, injured players). In contrast, the betting odds have a pure predictive character. The previous performance of teams might be used as a predictor for the upcoming performance, but the idea of rewarding teams for their previous achievements should not be included. Nevertheless, any information that bookmakers or punters consider to be relevant for predicting future performances should have an influence on the betting odds.

There are several obvious factors that influence betting odds but are not reflected in the ranking. The home advantage (for reviews, see Carron, Loughhead, & Bray, 2005) is the most relevant factor to be mentioned in this regard. Moreover, in group stages constellations can occur when a specific result helps both teams and this information will have an impact on the betting odds as well. Information about injured players, performance of players in domestic leagues or in rare cases even inside information and knowledge about match fixing might furthermore have an influence on the betting odds. This theoretical consideration shows that the predictor BET has a serious advantage over the predictor RANK when it comes to exploiting relevant information. If we assume the bet market to be effective, this information advantage will lead to a better quality of predictions. Hypothesis 3 is based on this idea.

The forecasting quality of both described predictors is evaluated. In light of the reviewed literature and our theoretical considerations, the following hypotheses were formulated:

Hypothesis 1: Both predictors are able to provide accurate predictions in a sense of being more accurate than a prediction made without any information.

Hypothesis 2: The predictor RANK has improved its predictive qualities since the mode of calculation was changed in order to improve team rankings after the World Cup 2006.

Hypothesis 3: The predictor BET is able to provide better predictions than the predictor RANK.

Two different kinds of predictions are addressed concerning the single matches of the World Cup. The first question is "Which team is more likely to win?". This means that one of the participating teams is appointed to be the "favourite" and 1 team is appointed to be the "outsider" during the match. We denote this classification as a model-free prediction of the match outcome for Hypothesis 1 and 2 as Hypothesis 1a and 2a. The second question is "How likely is each of the possible match outcomes?". To answer this question, a probability is assigned to each of the possible results (team 1 wins/draw/team 2 wins) of a match. We refer to those 3 probabilities as a percental prediction of the match for Hypothesis 1 and 2 as Hypothesis 1b and 2b.

Method

Data set

We studied data of the 3 World Cup Tournaments in 2006, 2010, and 2014 to evaluate and compare the quality of predictive techniques. Two sources of information were used to predict the outcome of in total 183 matches during the mentioned tournaments. To compile these predictions, betting odds from Betfair as well as the respective FIFA World Ranking before each World Cup were used.

Betting odds

The betting odds used in this study were obtained from Betfair Data, the official data supplier of the world's largest bet exchange Betfair (http://data.betfair.com/). Throughout this study the predictions derived from betting odds are denoted as predictions from the *bet market*. This wording might imply that the betting odds reflect the prediction of a whole market instead of 1 bookmaker or bet exchange. From a theoretical point of view, this seems reasonable as online betting leads to a high transparency of prices (in this case betting odds) and it can be assumed that differences between betting odds from various bookmakers and bet exchanges are highly limited. See Appendix 3 for additional empiric evidence that the results of this study do not depend on the choice of Betfair as representative for the bet market. In a bet exchange, customers are able to adopt the role of a bookmaker and offer betting odds or act as a punter and accept betting odds. Thus, the betting odds are traded and the mechanisms of a bet exchange can be compared to those of financial markets. As the betting odds are traded, there can be substantial changes prior to a match resulting in various different betting odds. To obtain unique betting odds for each outcome, we used those betting odds for each outcome having the highest betting volume.

Betting odds from a bet exchange were chosen because they are predestined to be used in predictive analysis. Betting odds from bookmakers include a margin and might furthermore be influenced by decisions of risk management or profit maximisation. In contrast, betting odds from a bet exchange do not directly include a margin and therefore seem more adequate to be used for predictions. Franck, Verbeek, and Nüesch (2010) provide empirical evidence that the prediction accuracy of Betfair is superior to the prediction accuracy of bookmakers in matches from major European leagues.

A disadvantage of using betting odds from a bet exchange is that they might not be useful to obtain predictions when liquidity is low (for more information on liquidity at Betfair, see Flepp, Nuesch, & Franck, 2014). In this study World Cup matches are considered which have an exceptionally high liquidity (i.e., volume of matched bets). Except for 2 matches all matches considered in this study possess a volume of more than 2.000.000€. 142 out of those 183 matches even exceed a volume of 5.000.000€.

FIFA World Ranking

The FIFA World Ranking is a ranking system for national football teams regularly published by FIFA and is used for instance to seed draws for tournaments. For each tournament the points for all 32 participants were obtained from FIFA World Ranking Website (2014). In all 3 cases we used the previous ranking published before the start of the tournament, which was released on the 17 May 2006, the 26 May 2010 and the 5 June 2014.

Prediction

Model-free prediction

A simple way of predicting World Cup matches is to calculate a model-free prediction. Therefore, the 2 predictors are used to determine which team has a higher probability to win a match (denoted as favourite) and which team has a lower probability to win the match (denoted as outsider). Concerning the predictor BET, a team is selected to be the favourite if it has the smaller betting odds compared to the opponent. For the predictor RANK, the team having the higher number of points and thus the lower position in the ranking is selected to be the favourite. By doing so, we obtained a model-free prediction for 183 of all 192 matches played during the World Cups 2006, 2010, and 2014. Betting odds were only available at Betfair Data for 55 of the 64 matches in the World Cup 2006. Thus, 9 matches from this tournament could not be included in this study.

Percental prediction

The model-free prediction is having a small level of detail as only a simple classification could be made. One team is picked to be the favourite and assumed to have a higher probability to win the match. Any information on the clarity of this classification is not included and no numeric value for the probability of different match outcomes is given. Hence, we also analysed percental predictions consisting of 3 probabilities for each match: the probability for each team to win and the probability for a draw. Deriving such a prediction from the bet market is fairly straightforward. The inverse betting odds can be interpreted as a probability for the corresponding result. Naturally, the 3 probabilities calculated that way will add to a value near 1 but not necessarily exactly equal to 1. Therefore, the probabilities are normalised to obtain a sum of exactly 1.

This calculation is shown on the example of the World Cup final 2014 Germany vs. Argentina. The betting odds having the highest betting volume were 2.36 (Germany), 3.45 (Argentina), and 3.35 (Draw). Note that in knockout matches the results after regular time are evaluated so that the case of a draw is included. The inverse betting odds are 0.4237 (Germany), 0.2899 (Argentina), and 0.2985 (Draw) summing up to 1.0121. Note that this deviation does not reflect a margin, but occurs from the choice of betting odds with the highest volume. After normalising to 100% (i.e., dividing by 1.0121) the percental prediction is 41.87% (Germany), 28.64% (Argentina), and 29.49% (Draw).

Deriving a similar prediction from the world ranking is more complicated, and a model-based approach is needed. This comes from the fact that the world ranking is not originally designed to obtain predictions. Our model can be divided into 2 steps: First, the ranking points need to be transferred into expected goals. The idea of expected goals is to quantify the (theoretical) average number of goals a team would score in a match if it was repeated a large number of times and thus random effects would be eliminated. Second, these expected goals need to be transferred into probabilities for each outcome of the match regarding the fact that in reality a match is obviously not repeated but only played a single time.

We started with the assumption that in a match between 2 teams we would expect the higher ranked team to score more goals on average than its opponent. This assumption was quite straightforward, but we still needed to design a concrete model that translates the difference of points in the FIFA World Ranking into the difference of expected goals. Denoting the ranking points of a team participating in a match as pts_{MAX} (team with higher number of points) and pts_{MIN} (team with lower number of points) and the expected goals as exp_{MAX} and exp_{MIN} the following equation shall hold:

$$\frac{exp_{MAX}}{exp_{MIN} + exp_{MAX}} = \frac{\alpha \cdot pts_{MAX}}{pts_{MIN} + \alpha \cdot pts_{MAX}}.$$

To ensure a reasonable differentiation between the teams, the points are transformed by using the factor α. Note that for the choice of $\alpha = 1$ the percentage difference of points equals the percentage difference of expected goals. To ensure an optimal comparability the factor α is chosen so that the average probability gap between favourite and outsider in the group stage (referred to as Δ_{win}) is equal for BET and RANK (for exact calculation see Appendix 1).

The above equation contains information on the gap of expected goals between favourite and outsider, but no information on the overall number of goals expected of both teams. As no information about special offensive or defensive qualities of teams can be derived from the ranking, all teams need to be treated equally. Therefore, we assumed the expected goals scored by both teams to be equal in each match and estimated this value by using the average number of goals scored in the previous World Cup. Denoting the average number of goals scored in the previous World Cup as $\hat{g}$ we claimed the following equation to hold:

$$exp_{MAX} + exp_{MIN} = \hat{g}.$$

By combining these 2 equations we were able to translate the points in the World Ranking into expected goals scored by both teams as in the following:

$$exp_{MAX} = \hat{g} \cdot \frac{\alpha \cdot pts_{MAX}}{pts_{MIN} + \alpha \cdot pts_{MAX}},$$

$$exp_{MIN} = \hat{g} - exp_{MAX}.$$

To transfer expected goals into probabilities for each outcome, slightly modified Poisson distributions are widely used in studies and in practice. We used a bivariate Poisson distribution suggested by Karlis and Ntzoufras (2003, 2005) (for exact calculation see Appendix 2).

The specified technique above makes it possible to translate the number of points in a ranking into a percental prediction. This calculation is expressed by the example of the 2010 World Cup Final Spain vs. Netherlands: As it was the mean number of goals scored in the World Cup 2006 we expected a total number of 2.25 goals in this match. Using the factor $\alpha = 1.382$ and the aforementioned equations, the points (1565 vs. 1231) were transferred into the expected goals which were 1.434 (Spain) and 0.816 (Netherlands). By using a bivariate Poisson distribution, we obtained a prediction of 51.48% (Spain), 28.64% (Draw), and 19.88% (Netherlands).

Estimation of prediction accuracy

Number of goals (model-free prediction)

As a quality criterion for the model-free prediction we used the number of goals scored by the favourite teams and those scored by the outsider teams. If the predictions were accurate we would have expected the teams selected as favourite to score more goals and win more matches than those teams selected as outsider. To evaluate the quality of predictions we chose the criterion goals instead of matches won for various reasons: (a) there is evidence that goal difference is a better measure for team strengths than matches won (Heuer & Rubner, 2009); (b) as the number of goals is higher than the number of matches, we expect a higher chance to obtain significant differences between both predictors; and (c) goals can easily be divided into 2 groups (goals scored by the favourite/goals scored by the outsider) whereas the match results also include the case of a draw.

Likelihood (percental prediction)

Obviously the use of a more detailed prediction requires the use of a more detailed quality criterion. As the analysis has a retrospective character, the actual results of the predicted matches are known. Thus, we were able to determine which probability was predicted for the actual result and multiply those probabilities for all evaluated matches. The value received by this technique is known as likelihood and used below to evaluate the quality of the percental predictions (for a conceptual explanation of basic maximum likelihood principles, see Myung, 2003).

Results

Model-free prediction

Figure 1 and Table 1 show the number of goals scored by the favourite and the outsider teams in all 3 World Cups. If we had to determine the favourite and the outsider without having any information at all on the teams participating, we would have expected both favourite and outsider to score 50% of the goals. In all cases, the favourite scored more goals than the outsider confirming that the model-free predictions from both predictors were more accurate than predictions made without any information. A one-sided binomial test indicated that for both predictors and all years the number of goals for the favourite teams significantly differed from 50% ($P < 0.05$ for RANK in 2006, $P < 0.01$ in all other cases), which strongly supports Hypothesis 1a.

Moreover, the results showed that in 2006 the predictor BET seemed to clearly outperform the predictor RANK whereas in 2010 and 2014 the predictor RANK was providing slightly better results. Therefore, we wanted to address the question which of the 2 predictors had the higher predictive quality. By using a McNemar's test, we found that in 2006 BET was performing significantly better than RANK ($P < 0.001$). In 2010 and 2014 a McNemar's test revealed that the difference between both predictors was not significant. When we combined the data of 2010 and 2014, no significant difference existed as well. We were therefore able to conclude that the predictive quality of RANK had improved since 2006 compared to the quality of BET. Recalling the change of calculation after 2006, this result appears to be consistent and is in line with our expectation (see Hypothesis 2a). It could as well be argued that instead of an improving ranking the bet market might have deteriorated its predictive quality over time. However, we do not have any indication that the predictive quality of the bet market might structurally decrease over time. Additionally, studies with indications on continuous improvements of prediction quality of betting odds exist (e.g., Forrest et al., 2005; Štrumbelj & Šikonja, 2010).

As we focused on the comparison between BET and RANK, we did not study the reasons for the slightly increasing number of correct predictions from BET in detail. Nevertheless, this could be a result of more balanced team skills and is not necessarily an indication for a structural deterioration of predictions. The improvement of RANK (compared to BET) after 2006 raises the question, whether it is in fact capable of providing equal or even better predictions than BET. We cannot clearly answer this question yet and are therefore in need of using the more detailed percental predictions.

Percental prediction

Figure 2 and Table 2 show the Log-likelihood values for BET and RANK for various years as well as a benchmark. The Log-likelihood benchmark is calculated as in the following: We estimate the probability of draws by using the empirical frequency of draws in the previous World Cup and the probability of a win equally for both teams. In 2014, for example, this results in a prediction of 35.9% (team 1), 28.1% (draw), and 35.9% (team 2) for each match. Thus we obtain a percental prediction we would have chosen without having any information about the teams. The log-likelihood value of this prediction is given in the table and referred to as benchmark.

Except for RANK in 2006, all predictions are able to provide significantly better predictions than the benchmark

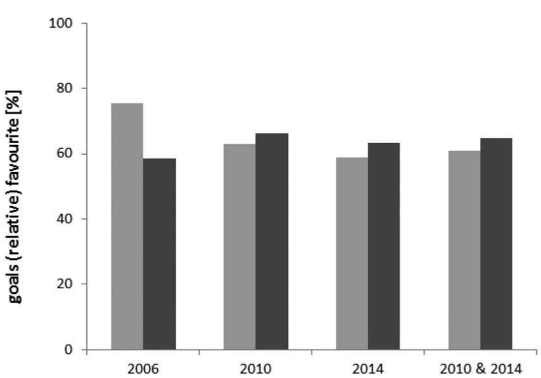

Figure 1. Percentage of goals scored by favourite teams according to the model-free prediction of both predictors.

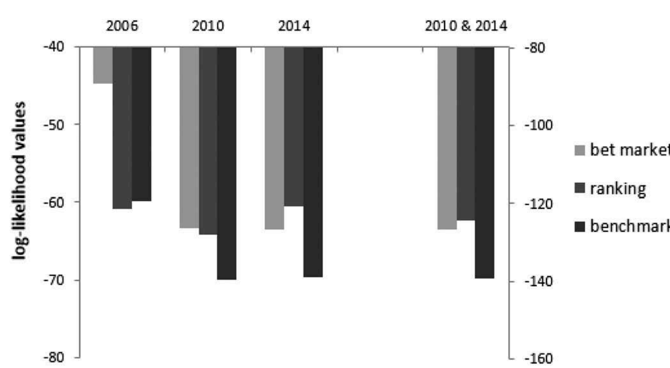

Figure 2. Log-likelihood values of the percental prediction by both predictors.

Table 1. Number of goals scored by favourite and outsider teams according to the model-free prediction of both predictors.

	Bet market				Ranking			
	Goals (Absolute)		Goals (Relative)		Goals (Absolute)		Goals (Relative)	
	Favourite	Outsider	Favourite	Outsider	Favourite	Outsider	Favourite	Outsider
World Cup 2006 (55 matches)	98	32	75.4%**	24.6%	76	54	58.5%*	41.5%
World Cup 2010	90	53	62.9%*	37.1%	95	48	66.4%*	33.6%
World Cup 2014	96	67	58.9%*	41.1%	103	60	63.2%*	36.8%
World Cup 2010 and 2014	186	120	60.8%*	39.2%	198	108	64.7%*	35.3%

* Significantly more than 50% ($P < 0.05$).
** Moreover significantly better than other predictor ($P < 0.001$).

Table 2. Log-likelihood values of the percental prediction by both predictors.

	Log-likelihood (bet market)	Log-likelihood (ranking)	Log-likelihood (benchmark)
World Cup 2006 (55 matches)	−44.71**	−60.80	−59.86
World Cup 2010	−63.36*	−64.20*	−69.95
World Cup 2014	−63.48*	−60.50**	−69.66
World Cup 2010 and 2014	−126.83*	−124.70**	−139.61

* Significantly better than benchmark ($P < 0.001$)
** Moreover significantly better than other predictor ($P < 0.05$)

(Likelihood ratio test, $P < 0.001$), proving the general usefulness of both predictors (Hypothesis 1b). There is no significant difference between the predictions derived from RANK in 2006 and the benchmark. This failure of outperforming the most simple benchmark model is clear evidence that the ranking system in 2006 is not comparable to the ranking system since 2006 and clear evidence for the improvement of RANK (Hypothesis 2b).

As it is not even able to outperform the benchmark, RANK was likewise performing significantly worse than BET in 2006 (Likelihood ratio test, $P < 0.0001$) supporting the general idea that betting odds should be able to provide better predictions than rankings. Surprisingly, this idea cannot be approved after 2006. In 2010, BET showed slightly better results than RANK, but a likelihood ratio test shows that no significant difference can be attested. In 2014, the predictive quality of RANK was clearly higher than the predictive quality of BET. Not only the bet market is not able to outperform the ranking, but in turn it seems to be outperformed by the ranking. A likelihood ratio test shows that the result is significant ($P < 0.02$). Moreover, if we pool the data from 2010 and 2014, the ranking is still able to outperform the bet market ($P < 0.05$). This result is completely contradicting our initial expectation (Hypothesis 3) and the findings of several studies that attest the excellent predictive quality of betting odds compared to ranking systems (Boulier & Stekler, 2003; Leitner et al., 2010; Scheibehenne & Bröder, 2007). Our results are a strong indication that rankings can possibly have a higher predictive quality than assumed usually.

It can be argued that matches involving the host of a World Cup are not fully comparable to other matches. Also matches in the group stage and matches in the knockout stage might not be fully comparable either. Tables 3 and 4 illustrate the results with distinction between these different categories of matches.

Discussion

Forecasting the outcome of worldwide sporting events is an area that does not only engage numerous individuals with varying knowledge, but also includes various types of prediction tasks. As indicated by the present study, our model is able to derive detailed percental predictions from ranking systems. This approach can even be extended to more detailed prediction tasks (i.e., predicting the final scores of a match or the tournament progress of a team).

As the results in this study indicate, the predictive quality of the world ranking has increased essentially since the mode of calculation was changed. Therefore, the comparability of the World Cups until 2006 and after 2006 is highly limited. Considering the prediction of World Cup matches, it leaves us with a quite small sample of data including the *new* ranking system (2 World Cups, 128 matches). Surely, a bigger sample of data would allow an even more precise examination of predictive qualities. Unfortunately the data set will not be extended before the following World Cup in 2018. The evaluation of additional friendly or qualification matches is principally possible, but these matches are structurally not comparable to World Cup matches for various reasons such as motivational aspects or the absence of matches on neutral venues. Likewise including matches from continental

Table 3. Number of goals scored by favourite and outsider teams according to the model-free prediction of both predictors divided by match categories.

	Bet market				Ranking			
	Goals (Absolute)		Goals (Relative)		Goals (Absolute)		Goals (Relative)	
	Favourite	Outsider	Favourite	Outsider	Favourite	Outsider	Favourite	Outsider
World Cup 2006								
Group stage (42 matches)	82	25	76.6%**	23.4%	65	42	60.7%*	39.3%
Knockout stage (13 matches)	16	7	69.6%*	30.4%	11	12	47.8%	52.2%
Host matches (7 matches)	14	4	77.8%*	22.2%	10	8	55.6%	44.4%
Without host (48 matches)	84	28	75.0%**	25.0%	66	46	58.9%*	41.1%
World Cup 2010								
Group stage (48 matches)	67	34	66.3%*	33.7%	65	36	64.4%*	35.6%
Knockout stage (16 matches)	23	19	54.8%	45.2%	30	12	71.4%**	28.6%
Host matches (3 matches)	5	3	62.5%	37.5%	5	3	62.5%	37.5%
Without host (61 matches)	85	50	63.0%*	37.0%	90	45	66.7%*	33.3%
World Cup 2014								
Group stage (48 matches)	84	52	61.8%*	38.2%	86	50	63.2%*	36.8%
Knockout stage (16 matches)	12	15	44.4%	55.6%	17	10	63.0%	37.0%
Host matches (7 matches)	11	14	44.0%	56.0%	17	8	68.0%	32.0%
Without host (57 matches)	85	53	61.6%*	38.4%	86	52	62.3%*	37.7%
World Cup 2010 and 2014								
Group stage (96 matches)	151	86	63.7%*	36.3%	151	86	63.7%*	36.3%
Knockout stage (32 matches)	35	34	50.7%	49.3%	47	22	68.1%**	31.9%
Host matches (10 matches)	16	17	48.5%	51.5%	22	11	66.7%*	33.3%
Without host (118 matches)	170	103	62.3%*	37.7%	176	97	64.5%*	35.5%

* Significantly more than 50% ($P < 0.05$).
** Moreover significantly better than other predictor ($P < 0.05$).

Table 4. Log-likelihood values of the percental prediction by both predictors divided by match categories.

	Log-likelihood (bet market)	Log-likelihood (ranking)	Log-likelihood (benchmark)
World Cup 2006			
Group stage (42 matches)	−32.43**	−43.91	−45.43
Knockout stage (13 matches)	−12.28**	−16.89	−14.44
Host matches (7 matches)	−4.46**	−7.90	−7.66
Without host (48 matches)	−40.25**	−52.90	−52.21
World Cup 2010			
Group stage (48 matches)	−49.35*	−50.59*	−52.62
Knockout stage (16 matches)	−14.01*	−13.61*	−17.33
Host matches (3 matches)	−3.40	−4.40	−3.33
Without host (61 matches)	−59.95*	−59.81*	−66.62
World Cup 2014			
Group stage (48 matches)	−45.63*	−43.33**	−51.33
Knockout stage (16 matches)	−17.84	−17.16	−18.34
Host matches (7 matches)	−6.69	−6.69	−7.65
Without host (57 matches)	−56.78*	−53.80**	−62.01
World Cup 2010 and 2014			
Group stage (96 matches)	−94.98*	−93.92*	−103.95
Knockout stage (32 matches)	−31.85*	−30.78*	−35.66
Host matches (10 matches)	−10.10	−11.09	−10.98
Without host (118 matches)	−116.74*	−113.61**	−128.63

* Significantly better than benchmark ($P < 0.05$).
** Moreover significantly better than other predictor ($P < 0.05$).

championships (like European Championship or Africa Cup of Nations) might enable further insights, but is lacking matches between teams from different continental confederations.

The choice of α can be criticised for 2 reasons. First, the prediction of RANK is (although to a small extend) dependent on data from BET. Second, Δ_{win} is calculated by using all group stage matches thus transferring information from these matches to the prediction of other matches. Nevertheless, we chose this way of calculating α to ensure an optimal comparability between RANK and BET. Moreover, we examined that the choice of α is not highly sensitive to Δ_{win} and the overall results are not highly sensitive to the choice of α.

Due to the structural idea of a bet exchange, various different betting odds can exist for the same match outcome. As explained before, we only considered those betting odds for each outcome, which had the highest betting volume, thus obtaining unique betting odds. It could be reasoned that the predictive quality of BET would increase if some kind of average betting odds was calculated for each outcome. For this reason we also studied volume-weighted average betting odds. Interestingly, virtually no differences in the results were provoked by this different approach. By reason of the less complex calculation, we therefore maintained the more simple approach of maximum volume betting odds.

In sum, the present study shows that the forecasting performance of FIFA World Ranking is better than usually assumed. It was shown that the change of calculation mode in 2006 led to an improved predictive quality compared to the bet market. While both predictors are clearly useful to obtain accurate predictions in general, the world ranking is able to outperform the bet market significantly for the World Cup 2014 and when the data from the World Cups 2010 and 2014 are pooled. We are reluctant to conclude that the world ranking in general is superior to betting odds because it is contradicting theoretical considerations and findings of other studies. Nevertheless we cannot ignore the results and therefore, in contrast to other studies, cannot confirm the superiority of betting odds over other predictors, yielding that its predictive qualities can not necessarily be considered optimal. Furthermore with the approach presented in this study, a useful framework is given which enables deduction of detailed predictions from ranking systems such as the FIFA World Ranking.

The world ranking outperforming the bet market is a surprising result, bearing in mind that betting odds have serious advantages over rankings when it comes to exploiting relevant information. It can be argued that rankings as well have a serious advantage over betting odds as they have a pure objective, quantitative character and thus cannot directly be subject to human misjudgement. Further studies investigating the emergence process of betting odds in detail, might help to gain practical insights on this theoretical consideration. Moreover it will be valuable to repeat the comparison of rankings and betting odds in the future including more data and thus approving or challenging the current results.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Andersson, P., Edman, J., & Ekman, M. (2005). Predicting the World Cup 2002 in Soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting*, *21*, 565–576. doi:10.1016/j.ijforecast.2005.03.004

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

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

Boulier, B. L., & Stekler, H. O. (2003). Predicting the outcomes of National Football League games. *International Journal of Forecasting*, *19*, 257–270. doi:10.1016/S0169-2070(01)00144-3

Camerer, C. F., & Johnson, E. J. (1991). The process-performance paradox in expert judgment: How can experts know so much and predict so badly?. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 195–217). New York, NY: Cambridge Press.
Carron, A. V., Loughhead, T. M., & Bray, S. (2005). The home advantage in sport competitions: Courneya and Carron's (1992) conceptual framework a decade later. *Journal of Sports Sciences*, *23*, 395–407. doi:10.1080/02640410400021542
Dobson, S., & Goddard, J. (2001). *The economics of football*. Cambridge, UK: Cambridge University Press.
Dyte, D., & Clarke, S. R. (2000). A ratings based Poisson model for World Cup soccer simulation. *The Journal of the Operational Research Society*, *51*(8), 993–998. doi:10.1057/palgrave.jors.2600997
Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, *49*, 283–306.
FIFA World Ranking Website. (2014). Retrieved September 26, 2014, from http://www.fifa.com/fifa-world-ranking/ranking-table/men/
Flepp, R., Nuesch, S., & Franck, E. (2014). *The liquidity advantage of the quote-driven market: Evidence from the betting industry* (UZH Business Working Paper No. 342). University of Zurich, Department of Business Administration. Retrieved March 6, 2015, from http://ssrn.com/abstract=2371185
Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, *21*, 551–564. doi:10.1016/j.ijforecast.2005.03.003
Franck, E., Verbeek, E., & Nüesch, S. (2010). Prediction accuracy of different market structures – bookmakers versus a betting exchange. *International Journal of Forecasting*, *26*, 448–459. doi:10.1016/j.ijforecast.2010.01.004
Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, *21*, 331–340. doi:10.1016/j.ijforecast.2004.08.002
Hershey, D. A., & Walsh, D. A. (2000/2001). Knowledge versus experience in financial problem solving performance. *Current Psychology*, *19*, 261–291. doi:10.1007/s12144-000-1020-x
Heuer, A., & Rubner, O. (2009). Fitness, chance, and myths: An objective view on soccer results. *The European Physical Journal B*, *67*(3), 445–458. doi:10.1140/epjb/e2009-00024-8
Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data using bivariate poisson models. *Journal of the Royal Statistical Association*, *52*, 381–393.
Karlis, D., & Ntzoufras, I. (2005). Bivariate poisson and diagonal inflated bivariate poisson regression models in R. *Journal of Statistical Software*, *14*(10), 1–36. doi:10.18637/jss.v014.i10
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*, 471–481. doi:10.1016/j.ijforecast.2009.10.001
McHale, I., & Davies, S. (2008). Statistical analysis of the effectiveness of the FIFA World Rankings. In J. Albert & R. H. Koning (Eds.), *Statistical thinking in sports* (pp. 77–90). Boca Raton, FL: Chapman & Hall/CRC.
Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90–100. doi:10.1016/S0022-2496(02)00028-7
Nilsson, H., & Andersson, P. (2010). Making the seemingly impossible appear possible: Effects of conjunction fallacies in evaluations of bets on football games. *Journal of Economic Psychology*, *31*, 172–180. doi:10.1016/j.joep.2009.07.003
OddsPortal Website. (2016). Retrieved April 15, 2016, from http://www.oddsportal.com
Önkal, D., & Muradoğlu, G. (1994). Evaluating probabilistic forecasts of stock prices in a developing stock market. *European Journal of Operational Research*, *74*, 350–358. doi:10.1016/0377-2217(94)90102-3
Önkal, D., Yates, J. F., Simga-Mugan, C., & Öztin, Ş.. (2003). Professional vs. amateur judgment accuracy: The case of foreign exchange rates. *Organizational Behavior and Human Decision Processes*, *91*, 169–185. doi:10.1016/S0749-5978(03)00058-X
Pachur, T., & Biele, G. (2007). Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events. *Acta Psychologica*, *125*, 99–116. doi:10.1016/j.actpsy.2006.07.002
Scheibehenne, B., & Bröder, A. (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting*, *23*, 415–426. doi:10.1016/j.ijforecast.2007.05.006
Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, *53*, 252–266. doi:10.1016/0749-5978(92)90064-E
Spann, M., & Skiera, B. (2009). Sports forecasting: A comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, *28*, 55–72. doi:10.1002/for.v28:1
Štrumbelj, E., & Šikonja, M. R. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting*, *26*, 482–488. doi:10.1016/j.ijforecast.2009.10.005
Suzuki, K., & Ohmori, K. (2008). Effectiveness of FIFA/Coca-Cola World Ranking in predicting the results of FIFA World Cup Finals. *Football Science*, *5*, 18–25.
Törngren, G., & Montgomery, H. (2003). Worse than chance? Performance and confidence among professionals and lay-people in the stock market. *The Journal of Behavioral Finance*, *5*, 146–153.
Yates, J. F., McDaniel, L. S., & Brown, E. S. (1991). Probabilistic forecasts of stock prices and earnings: The hazards of nascent expertise. *Organizational Behavior and Human Decision Processes*, *49*, 60–79. doi:10.1016/0749-5978(91)90042-R
Yates, J. F., & Tschirhart, M. D. (2006). Decision-making expertise. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 421–438). New York, NY: Cambridge University Press.

Appendices

Appendix 1

Calculation of α

Let $pFAV_i$, $pDRAW_i$, and $pOUT_i$ be the probabilities of the corresponding match outcome of match i. Denoting all matches of the group stage as matches 1 to n we calculate

$$\Delta_{\mathrm{win}} = \frac{1}{n}\sum_{i=1}^{n}(pFAV_i - pOUT_i),$$

as a measure for the overall degree of differences of team qualities. Note that the participants of knockout matches are not known in advance of the tournament and can therefore not be considered. We choose α such that Δ_{win} for BET and RANK are equal. The values of Δ_{win} and α are given below:

Table A1. Model parameters used for the percental prediction by predictor ranking.

World Cup	2006	2010	2014
Δ_{win}	42.6%	37.8%	36.7%
α	1.878	1.382	1.398

Note that the high value of α in 2006 is not mainly due to the slightly higher value of Δ_{win} but to the small differentiation of points in the ranking system until 2006.

Appendix 2

Bivariate Poisson model

Let $\exp_{\mathrm{MAX}}$ and $\exp_{\mathrm{MIN}}$ be the expected goals of both teams. The probability of the match ending with a result of X:Y is

$$P[X : Y|\lambda_1,\lambda_2\lambda_3] = e^{-(\lambda_{1,}+\lambda_2+\lambda_{3,})}\frac{\lambda_{1,}{}^{x}}{x!}\frac{\lambda_{2,}{}^{y}}{y!}\sum_{i=0}^{\min(x,y)}\binom{x}{i}\binom{y}{i}i!\left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^{i}, \text{ where}$$

$$\lambda_3 = 0.05(\exp_{\mathrm{MAX}} + \exp_{\mathrm{MIN}}),$$
$$\lambda_1 = \exp_{\mathrm{MAX}} - \lambda_3,$$
$$\lambda_2 = \exp_{\mathrm{MIN}} - \lambda_3.$$

Appendix 3

Data from further bookmakers

Throughout this study the predictions from the bet market are based solely on 1 representative (Betfair). We stated before, that this is reasonable from a theoretical point of view. However, it seems necessary to verify this consideration by examining empiric data.

The Odds Portal website (2016) offers information on average betting odds from various bookmakers on current and previous events. We used data from this website to derive predictions for the World Cups 2010 and 2014 with the same procedure as used for Betfair throughout the study. In 2010, for each match 7 different bookmakers were considered while in 2014 for each match at least 11 different bookmakers were considered. Note that extensive betting odds were not available for the World Cup 2006. This website is not an official source of the bookmakers but nevertheless, we found no indication for incorrect or inaccurate data.

We found that the predictions from Betfair and further bookmakers are almost perfectly correlated (correlation 0.999474 in 2010 and 0.999648 in 2014). In addition, Tables A2 and A3 show that the prediction quality of Betfair and further bookmakers are virtually equal and thus the predictions used throughout the study do actually reflect the predictions of the bet market.

Appendix 4

Betting returns

We examined football predictions from a scientific point of view and not from an economic point of view. Nevertheless, it is worth investigating whether positive betting returns could have been achieved by betting following our model-based predictions deducted from the world ranking.

A reasonable betting strategy would be to bet 100$ on a match outcome if our model yields a higher probability than the bet market. Consequently, no bet will be made if the probability determined by the model is lower than the probability of the bet market.

Table A4 shows the stakes, the net winnings and the winnings excluding a 5% commission on winnings as usual at Betfair. The highest winnings would have been achieved in 2014, followed by 2010. In 2006, a loss would have been made. These results are in line with the predictive quality of the model. Note that in general a higher predictive quality does not necessarily guarantee a positive betting return in every case and vice versa. Note also that no statement about the degree of significance is made at this point.

Table A2. Number of goals scored by favourite and outsider teams according to the model-free prediction of Betfair and Odds Portal.

	Betfair				Odds Portal			
	Goals (Absolute)		Goals (Relative)		Goals (Absolute)		Goals (Relative)	
	Favourite	Outsider	Favourite	Outsider	Favourite	Outsider	Favourite	Outsider
World Cup 2010	90	53	62.9%*	37.1%	90	53	62.9%*	37.1%
World Cup 2014	96	67	58.9%*	41.1%	96	67	58.9%*	41.1%
World Cup 2010 and 2014	186	120	60.8%*	39.2%	186	120	60.8%*	39.2%

* Significantly more than 50% ($P < 0.05$).
** Moreover significantly better than other predictor ($P < 0.05$).

Table A3. Log-likelihood values of the percental prediction by Betfair and Odds Portal.

	Log-likelihood (Betfair)	Log-likelihood (Odds Portal)	Log-likelihood (benchmark)
World Cup 2010	−63.36*	−63.48*	−69.95
World Cup 2014	−63.48*	−63.44*	−69.66
World Cup 2010 and 2014	−126.83*	−126.91*	−139.61

* Significantly better than benchmark ($P < 0.05$).
** Moreover significantly better than other predictor ($P < 0.05$).

Table A4. Betting returns using a betting strategy according to the predictions of predictor ranking.

	Stakes	Net winnings	Winnings excluding commission
World Cup 2006	8,000.00$	−3,413.00$	−3572.35$
World Cup 2010	9,900.00$	683.00$	278.85$
World Cup 2014	10,900.00$	2454.00$	1966.30$