International Journal of Forecasting 34 (2018) 17–29

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results

Thomas Peeters *

Erasmus School of Economics, Tinbergen Institute, Erasmus Research Institute of Management (ERIM), Netherlands

ARTICLE INFO

Keywords:
Wisdom of Crowds
Forecasting
Football
Transfermarkt

ABSTRACT

This paper investigates the value of collective judgments which stem from settings that have not been designed explicitly to elicit the 'Wisdom of Crowds'. In particular, I investigate information obtained from transfermarkt.de, an online platform where a crowd of registered users assess the value of professional soccer players. I show that forecasts of international soccer results based on the crowd's valuations are more accurate than those based on standard predictors, such as the FIFA ranking and the ELO rating. When this improvement in forecasting performance is applied to betting strategies, it leads to sizable monetary gains. I further exploit information on the preferences of individual crowd members in order to investigate whether wishful thinking hampers the accuracy of crowd valuations, but fail to find evidence that such is the case.

1. Introduction

A growing strand of the literature highlights the value of collective judgments, often referred to as the 'Wisdom of Crowds', for assessing the probability of future events (e.g., Surowiecki, 2004). Researchers usually elicit such wisdom by relying on controlled experiments (e.g., Herzog & Hertwig, 2011; Lorenz, Rauhut, Schweitzer, & Helbing, 2011; Simmons, Nelson, Galak, & Frederick, 2011), prediction markets (e.g., Cowgill & Zitzewitz, 2015, Forsythe, Rietz, & Ross, 1999, Wolfers & Zitzewitz, 2004) or even prediction polls (e.g., Atanasov, Rescober, Stone, Swift, Servan-Schreiber, & Tetlock, 2016). These settings allow researchers retain varying degrees of control over the mechanism through which collective judgments arise. As a consequence, recent research has looked into questions regarding the design of these mechanisms, such as how to select crowd members optimally (Budescu & Chen, 2015; Goldstein, McAfee, & Suri, 2014; Lambertson & Page, 2012), how to safeguard iversity in the crowd (Economo, Hong, & Page, 2016) or whether experimental subjects should be allowed to communicate (Li & Liu, 2015). In this context, Simmons et al. propose four conditions for crowd wisdom: crowd members should be "(1) knowledgeable, (2) motivated to be accurate, (3) independent, and (4) diverse" (Simmons et al., 2011 p. 2). When these conditions are met, crowd predictions are thought to provide useful information for managerial decision making and public policy.

Recently, though, a new source of potential crowd wisdom has surfaced, in the form of various highly popular internet forums where large groups of users express their opinions on a variety of issues. Some of the most prominent examples include Twitter, which has been studied as a predictor of soccer games (see Brown, Rambaccussing, Reade, & Rossi, 2016), and IMDb and rottentomatoes.com, which aggregate user scores on movies and TV series (see Camara & Dupuis, 2014). One could even consider crowd funding websites such as kickstarter.com as a source of crowd wisdom (see Mollick & Nanda, 2015). However, as these websites have not been set up with the aim of eliciting crowd wisdom, they usually generate collective judgments in ways that do not satisfy the conditions set out by Simmons et al. (2011). First, they typically do not

* Correspondence to: Erasmus School of Economics, Department of Applied Economics, Postbus 1738, 3000DR Rotterdam, Netherlands.
E-mail address: peeters@ese.eur.nl.

http://dx.doi.org/10.1016/j.ijforecast.2017.08.002
0169-2070/

provide users with any explicit incentives to induce accurate reporting. Second, they usually allow (and indeed stimulate) communication between users, which may limit the independence of user opinions. Finally, they often make little attempt to reach a diverse user population. In itself, the absence of monetary rewards may not harm the prediction accuracy (e.g., Servan-Schreiber, Wolfers, Pennock, & Galebach, 2004), but the combination of low independence, low diversity and low powered incentives obviously raises concerns about the usefulness of these crowd judgments for decision makers (e.g., Lorenz et al., 2011).

Nonetheless, three recent papers provide supportive evidence of the value of this type of crowd wisdom. First, Chen, De, Hu, and Hwang (2014) show that a textual analysis of users' posts on seekingalpha.com, a popular opinion forum for stock market investors, has predictive power for future stock returns. Second, Mollick and Nanda (2015) argue that support on the crowd funding website kickstarter.com is a better predictor of the success of theatre productions than evaluations by a designated expert panel. Finally, Brown et al. (2016) show that sentiment analyses from selected Twitter messages may improve the prediction accuracy of a betting exchange.

This paper aims to shed further light on the value of this wisdom of crowds 'from the field'. To this end, I analyze player valuation data from a popular soccer website called Transfermarkt. I examine three closely related issues. First, I compare the forecasting performance of a model predicting international soccer results (i.e., games played between countries) based on the website's valuations with those of several benchmark forecasting models from the sports forecasting literature, namely betting odds, ELO ratings and FIFA rankings (e.g., Forrest, Goddard, & Simmons, 2005; Hvattum & Arntzen, 2010; Stekler, Sendor, & Verlander, 2010).[1] Second, I calculate the monetary gains that can be obtained when using the predictions from the Transfermarkt valuations model to bet on soccer matches. Third, I exploit data on the individual preferences of website users in order to check for wishful thinking bias in the Transfermarkt valuations. As such, I directly test a potential bias arising from a lack of crowd diversity in this setting.

My analysis uncovers three interconnected results. First, a simple model based on Transfermarkt values predicts game results more accurately than models based on either ELO ratings or FIFA rankings. In contrast, averaged betting odds provide more accurate predictions than the Transfermarkt model. However, the Transfermarkt valuations still contain information that allows the prediction performance of the betting odds to be improved. Second, the Transfermarkt predictions can be employed to devise profitable betting strategies, which again outperform the rival predictors, ELO and FIFA. Finally, I fail to find evidence of wishful thinking in the Transfermarkt valuations model.

Taken together, my findings indicate that crowd valuations can potentially be a powerful source of information for predicting international soccer results. As such, I find further support for the value of collective judgements from non-experimental settings, even when they clearly violate the conditions set out by Simmons et al. (2011). In addition to contributing to our understanding of collective judgement, these results are also of practical importance, as Transfermarkt valuations have recently become an important reference for practitioners in both the professional soccer and betting industries.

The next section describes the functioning of the Transfermarkt platform in more detail. Section 3 outlines the dataset and the methodology used for forecasting game outcomes. Section 4 then evaluates the forecasting performances of the valuations model and various rival forecasting models, in terms of both accuracy and potential returns from a set of betting strategies. Section 5 contains the empirical results pertaining to wishful thinking bias. Finally, Section 6 concludes and provides several avenues for future research.

2. Crowd valuations on Transfermarkt

This analysis uses data from Transfermarkt, a soccer statistics website which is known mainly for publishing monetary valuations, dubbed 'market values', for a very large sample of professional soccer players. These market values are referred to regularly both by researchers in the sports economics or management literature (e.g., Bryson, Frick, & Simmons, 2013; Herm, Callsen-Backer, & Kreis, 2014)[2] and in the popular press (e.g., Bloomberg, 2016). Moreover, several club officials have revealed privately that player agents tend to refer to Transfermarkt valuations during player contract negotiations, indicating their increasing importance for the soccer player transfer market itself.

Transfermarkt derives its market values from the assessments of the site's registered users, who reveal their opinions about the player valuations on specialized forums designated for individual players. Each user entry indicates whether the user believes a particular player valuation to be too high or too low, and suggests an updated valuation. Typically, users also provide a short explanation for their assessment, which keeps the discussion on the forum alive. The site then updates the market value of each player at regular intervals through the soccer season based on the user posts since the last update. The evolution of this valuation over the course of a player's career is then depicted on the player's profile page, which also shows the player's current and former clubs, playing position, personal characteristics, and performance in terms of titles and cups.

1 Other forecasts in the literature include expert opinions (Frick & Wicker, 2016) and betting exchanges (Franck, Verbeek, & Nuesch, 2010; Smith, Pateon, & Williams, 2009). However, data for comparisons with these predictors could not be obtained for the full sample used in this study.

2 This literature uses the valuations primarily to proxy for player wages or transfer fees. As such, Herm et al. (2014) look at the correlation between the Transfermarkt valuations and the actual transfer fees paid. While they find a strong correlation, this is a difficult exercise for at least two reasons. First, the secrecy of most deals prohibits the accumulation of a large database from a single, reliable and open source. Second, the peculiar labor market institutions in professional football make it difficult to argue there should be one-on-one matching between valuations and fees, as, for example, end-of-contract players can move for a zero transfer fee even when they clearly provide economic value to the respective teams.

Table 1
Summary statistics: game results and sample composition.

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Loss team i	1020	0.369	0.483	0	1
Draw	1020	0.221	0.415	0	1
Win team i	1020	0.411	0.492	0	1
Goal margin	1020	0.103	2.230	−10	11
Tournaments					
WC 10 qualifier	26.3%				
World Cup 10	2.3%				
WC 14 qualifier	42.2%				
EC 12 qualifier	24.3%				
Euro Cup 12	3.0%				
Copa America	2.0%				

Notes: The game results and final scores are collected from the football analytics websites http://uk.soccerway.com, http://www.national-football-teams.com and http://www.footballdatabase.eu.

Next to this, each registered user has a personal profile, which shows the user's chosen name and the number of entries he or she has made. Most user profiles also reveal which club the user supports. As was shown by Massey, Simmons, and Armor (2011), sports fans tend to overvalue the quality of their preferred team. Hence, an undiversified presence of fans for different teams in the user population could bias the valuations towards wishful thinking, in much the same way as election futures markets may be biased by an overrepresentation of democrats or republicans in the market (e.g., Forsythe et al., 1999). This analysis exploits this feature of the website in order to devise a test for the presence of wishful thinking in the Transfermarkt valuations.

3. Prediction models for international soccer games

3.1. Data

I investigate the forecasting performance of the Transfermarkt valuations based on a dataset containing non-friendly international games played between September 2008 and May 2014. I specifically select games among and between the 63 countries from UEFA and CONMEBOL,[3] because the players from these nations constitute the vast majority of those covered by Transfermarkt. At the game-country level, 81.4% of all observations relate to UEFA countries, versus 18.6% for the CONMEBOL countries. The choice of games between national teams rather than club teams stems from three considerations. First, the literature presents more readily available rival predictors for matches between countries than for club games (see below for more detail). Second, the competition format for games between countries is usually simpler than that for club teams. Club teams are often active in multiple competitions at the same time, leading to the strategic resting of key players or reduced efforts by fielded players in less important games. National teams, on the other hand, only play a limited number of tournaments with fairly straightforward formats, and are never normally active in several competitions at once. Third, national team managers are restricted to selecting players with the nationality of their respective country. This largely mitigates the main endogeneity concern that can be present in the analysis of club soccer, i.e., that winning games creates feedback, allowing better players to be obtained through an increase in club revenue (and therefore player wages).

Table 1 presents an overview of this dataset. The data contain 1020 games, each of which is contested by two countries, dubbed i and j. For each game, I randomly assign one country to take the role of i and the other to be j. More than two thirds of all observations are FIFA World Cup qualifying games (26.3% for the 2010 and 42.2% for the 2014 World Cup). Slightly under a quarter (24.3%) of all observations are qualifier games for Euro 2012, and the remaining 8% are taken from the 2012 Euro Cup (3%), the 2010 FIFA World Cup (2.3%) and the 2011 Copa America (2%). All games in the qualifying tournaments are played in the home stadium of one of the countries involved. The Euro Cup, World Cup and Copa America are played in a single host country, meaning that only one team in the tournament is at home and all others play on neutral turf. In terms of the game results, around 22% of all games end in a draw, while the remainder are won by either team i or team j. Because of the random allocation to team i or j, the average goal margin between teams stands close to zero, but the distribution is fairly wide, with a standard deviation of about 2.2 goals.

The logarithms of the average Transfermarkt valuations (v_{it} and v_{jt}) of the players *selected* by teams i and j form the main predictor in the crowd valuations model. Once a player has appeared in the selection of a particular country, he automatically ceases to be eligible to play for another country in the future. Thus, restricting the analysis to players who were selected for the game avoids issues with dual-nationality players. However, I refrain from using information on which players actually *played* during the game, as this is not known before kick-off. The average is calculated by matching all players to their most recent valuations on the Transfermarkt website prior to the game. The online values are usually updated twice per season, meaning that a player typically has the same value over a sequence of about five international games. If a player does not have a listed value on the website, I assign him a value of €1. Although a distinct body of literature addresses the importance of the intra-team quality distribution in soccer (e.g. Coates, Frick, & Jewell, 2016; Franck & Nüesch, 2010), I refrain from including any distributional measures in the model, because my aim is to present the simplest possible prediction model.[4] Table 2 shows summary statistics for the selection-wide average of the player Transfermarkt values. The mean at the game level is around €4.7 m, but this number varies widely both across countries and over time. Several small associations, such as San Marino and Liechtenstein, have occasionally fielded teams in which no player has a value above €1, whereas soccer powerhouse

[3] UEFA is the federation of European national soccer associations, whereas CONMEBOL groups the South American soccer nations. However, these geographical boundaries are not followed strictly.

[4] Experimenting with the inclusion of distributional measures did not yield any significant gains in prediction accuracy either.

Table 2
Summary statistics: forecasting and bias control variables.

Forecasting variables	Country i					Country j			
	Obs.	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.
Market value (€)	1020	4 693 453	5 407 573	1	28 600 000	4 821 070	5 839 096	1	31 800 000
Home advantage	1020	0.468	0.499	0	1	0.466	0.499	0	1
Number players	1020	19.63	2.210	16	23	19.62	2.221	15	23
Odd loss	1020	6.151	9.551	1.010	71.35	6.260	10.04	1.010	76.69
Odd draw	1020	5.180	3.409	1.690	22.56	5.180	3.409	1.690	22.56
Odd win	1020	6.260	10.04	1.010	76.69	6.151	9.551	1.010	71.35
ELO points	1020	1641	247.3	857	2140	1645	247.7	859	2136
FIFA points	1020	688.2	350.3	0	1883	699.2	366.6	0	1883
Bias controls									
Av. number fans	1020	231.8	451.5	1	3514.4	209.1	390.3	1	3514.4
Av. rel. number fans	1020	0.005	0.011	0	0.073	0.004	0.008	0	0.074
Players at supp. clubs	1020	2.308	3.723	0	23	2.133	3.160	0	22
Support weighted value	1020	71 528	203 088	0	2 035 881	57 751	165 420	0	1 815 932

Notes: The identities, characteristics and team affiliations of players are collected from the football analytics websites http://uk.soccerway.com, http://www.national-football-teams.com and http://www.footballdatabase.eu. Player valuations are published at http://www.Transfermarkt.de. Historical betting odds for all games are taken from http://www.betbase1.info/ and represent an average of decimal odds across different bookmakers. FIFA points can be found at http://www.fifa-world-box.com and ELO ratings at http://www.eloratings.net.

Spain regularly fields a team with an average value of around €30 m.

Two further covariates enter into the forecasting model too. First, I add the numbers of players selected (n_{it} and n_{jt}). In several games, the two teams select different numbers of players, which implies that the average value is then calculated over a different number of players for each team. Table 2 reveals that an average team's selection, including both starters and those on the reserve bench, consists of 19.6 players, with a minimum of 15 and a maximum of 23. Second, it is generally accepted that a home advantage persists in soccer games (see e.g. Garicano, Palacios-Huerta, & Prendergast, 2005). Hence, I include indicator variables (h_{it} and h_{jt}) in the regression model to account for this. As is apparent from the home advantage indicators, about 3.3% of all games are played on neutral ground, with no team enjoying the home advantage.

I benchmark the forecasting performance of the Transfermarkt model by comparing it to those of four alternative prediction models. The first alternative uses only the relative frequencies of a loss, home win and draw to predict game outcomes (see Table 1). Naturally, this "null" model should be interpreted as a lower bound for the performance of the Transfermarkt model. A second alternative is based on averaged decimal betting odds gathered from the online archive betbase1.info. Table 2 details the summary statistics for these raw odds. Although the interpretation is slightly tricky because these values still include the bookmaker's profit margin, it is clear that the odds indicate a wide divergence in the pre-game expectations about win probabilities. For the third alternative, I follow Hvattum and Arntzen (2010), who use ELO ratings to forecast soccer results. The ELO rating system, which originated in chess, allocates points to countries based on a weighting of their past performances, taking the strength of their opponents into account. The final predictor, which is often used as a benchmark in the literature on forecasting soccer results, is the FIFA country ranking (e.g., Herzog & Hertwig, 2011). This ranking, developed by the international soccer federation FIFA, assigns points to each FIFA member country based on a weighted average of their international results over the previous four years. Although the exact formula has been criticized (and consequently changed[5]), FIFA tries to control for the importance of the game and the strength of the association and of the opposing team. Table 2 also contains summary statistics for the ELO and FIFA ratings. While these values are hard to interpret in isolation, they too appear considerably dispersed.

A final point that is worth noting is that the rival predictors are updated more frequently than the Transfermarkt valuations: the ELO scores and betting odds are updated on a game-by-game basis, while the FIFA ranking is recalculated after each international break, which usually contains at most two games. This would be expected to give the rival predictors an advantage over Transfermarkt in terms of forecasting performance.

3.2. A simple forecasting model for Transfermarkt valuations

I present two regression models for linking Transfermarkt valuations to international soccer results. In the first model, the dependent variable is a categorical measure y_{ijt} that describes the result of the game played between teams$i$ and j at time t as a loss, draw or win for team i. In the baseline approach, I relate this discrete outcome to the explanatory variables set out above using an ordered probit model. I also test two alternative link functions, namely an ordered logit and a multinomial logit.

Using the notation introduced above, the linear part of the ordered probit model can be written formally as

$$y^*_{ijt} = \beta_h h_{it} - \beta_h h_{jt} + \beta_v v_{it} - \beta_v v_{jt} \\ + \beta_n n_{it} - \beta_n n_{jt} + \varepsilon_{ijt}, \quad (1)$$

where h_{it} and h_{jt} are the home advantage indicators, v_{it} and v_{jt} are the logged average Transfermarkt valuations, and n_{it} and n_{jt} represent the numbers of players in each country's selection. Each β_x is a parameter to be estimated, where the subscript x refers to the respective explanatory variable. I

5 The last overhaul took place in July 2006, before the start of the data sample used in this analysis.

expect to find positive estimates for all coefficients. Clearly, a country with a home advantage, a higher average Transfermarkt value and a wider selection of players should have higher probability of winning a given game. The continuous expression in Eq. (1) is scaled into the three categories of the dependent variable y_{ijt} through the ordered probit procedure estimating two cut values. When y^*_{ijt} falls below the first cut value, the model predicts a loss for team*i*. When y^*_{ijt} is estimated to be between the two cut values, the model predicts a draw. If y^*_{ijt} exceeds the second cut value, the model predicts a win for team *i*. Since I dub teams as either *i* or *j* randomly, the two cut values should be exactly the opposite, and therefore symmetric around zero. The intuition for this can be understood by thinking of the hypothetical case where two exactly equivalent teams face one another. This leads Eq. (1) to equal zero, and the game outcome has a high predicted probability of being a draw. Then suppose that one team deviates by increasing its average value. This should increase the chance of this team winning the game by the same amount, irrespective of its identity (being dubbed *i* or *j*) in the dataset. This can only be the case when the two cut values are symmetric around zero.

In a second regression model, I introduce two stages. In the first stage, an OLS regression relates the goal difference between teams *i* and *j* in game *t* to the explanatory variables detailed above. In the second stage, an ordered probit regression links the predicted goal difference from stage 1 to the discrete set of potential game outcomes. As such, I obtain comparable sets of predicted probabilities under both approaches, derived from the same set of explanatory variables.

The parameters in both regression models are estimated by differencing each variable in Eq. (1) between the two opposing teams. This allows the home advantage parameter to be identified under the conditions that (a) the sampling does not result in team *i* being either at home or away in all observations, and (b) the cut value estimates are restricted to be symmetric around zero, because otherwise they are collinear to the home advantage dummy. The random assignment of countries as either team *i* or team *j* for any given game ensures that these conditions are met.

3.3. Competing prediction models

Next, I also have to formulate predicted outcome probabilities based on the alternative forecasting variables. While this is straightforward for the 'null' model (the observed frequency equals the predicted probability), the other rival predictors require a transformation. I calculate implied probabilities from decimal odds by taking the inverse of the odd and subtracting the "over", i.e., the bookmaker's profit margin. For example, I obtain the probability of a win by applying the formula

$$Prob\left(win_{ijt}\right) = \frac{1}{odd\left(win_{ijt}\right)}\frac{1}{1+over_{ijt}},$$

where

$$over_{ijt} = \frac{1}{odd\left(win_{ijt}\right)} + \frac{1}{odd\left(draw_{ijt}\right)} + \frac{1}{odd\left(loss_{ijt}\right)} - 1.$$

Implied probabilities of a draw and a loss can be obtained similarly. In contrast to other prediction models in this analysis, betting odds may incorporate all of the information available leading up to the game. Furthermore, bookmakers have an increasingly strong financial incentive to set accurate odds (e.g., Forrest et al., 2005). Thus, the betting odds probably present an upper-bound for the prediction accuracy of the Transfermarkt model, even though individual odds may still be less efficient than, say, betting exchanges, for example (e.g., Franck et al., 2010).

For the baseline ELO and FIFA models, I employ a regression approach similar to that depicted in Eq. (1), where I replace the log average Transfermarkt valuation with the log of the last ELO rating/FIFA points published by eloratings.net/FIFA prior to the game. While this approach implies a divergence from the ordered logit approach used by Hvattum and Arntzen (2010), it creates a level playing field in terms of the information used across different models, when comparing the prediction accuracies.[6] I present the ELO and FIFA models both including and excluding the number-of-players variable. On the one hand, there is not such a clear rationale for including this variable as in the Transfermarkt model. On the other hand, excluding it could put the ELO and FIFA model at an unfair disadvantage if it has additional predictive power that is not related directly to the average Transfermarkt valuations. Again, I expect positive coefficient estimates for all variables, as both a higher ELO rating and more FIFA points indicate that a country has a better past performance.

3.4. Estimation results

It is clearly not permissible for information that is not revealed until after the start of an event to be used to form predictions of the outcome of the event. I therefore estimate the regression models on a hold-out sample using a rolling estimation algorithm. Specifically, I re-estimate the model for each game day t using only information from games played before t. I then use these estimates to forecast the results for all games played on day t. Since it is clearly not feasible to report the updated results for each match day separately, Table 3 reports estimation results based on the full dataset, for the models of both game results and goal margins. Note that no actual forecasts will be based on these estimates, as information on even the very last games is taken into account here. Still, Table 3 provides insights into the overall fits of the regression models and the directions in which different predictors drive the forecasted probabilities.

As expected, Table 3 shows that teams with higher average Transfermarkt valuations have higher probabilities of winning a game. When this average value is derived from a wider selection of players, the chance of winning increases even more. These parameters also have a positive effect on the expected margin of victory. The alternative predictors show similar patterns; i.e., higher ELO ratings and FIFA points both increase the probability of winning the game and drive up the expected goal margin. The

6 I have also explored alternative link functions (ordered and multinomial logit) for these rival models in additional analyses, but they did not lead to any significant differences in the prediction accuracies achieved.

Table 3
Estimation results: alternative prediction models, full dataset.

Outcome	Game results					Goal margin			Game results		
Av. TM value	0.450***					0.552***			0.757***	−1.054***	−0.477***
	(0.024)					(0.020)			(0.044)	(0.068)	(0.056)
ELO rating		4.498***	4.572***				6.338***				
		(0.252)	(0.247)				(0.220)				
FIFA points				0.940***	0.959***			0.841***			
				(0.058)	(0.057)			(0.034)			
Num. of players	1.775**	1.337		2.187**		2.323**	1.528*	3.470***	2.950*	−4.143*	−2.773
	(0.880)	(0.875)		(0.881)		(0.918)	(0.908)	(0.952)	(1.523)	(2.180)	(1.934)
Home adv.	0.281***	0.283***	0.289***	0.295***	0.305***	0.408***	0.429***	0.431***	0.471***	−0.658***	-0.211**
	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.052)	(0.051)	(0.055)	(0.072)	(0.102)	(0.094)
Cut value/ constant	0.409***	0.400***	0.400***	0.390***	0.388***	0.088*	0.116**	0.083	0.694***	−0.182*	-0.217**
	(0.024)	(0.024)	(0.024)	(0.023)	(0.023)	(0.050)	(0.049)	(0.053)	(0.043)	(0.101)	(0.096)
Est. method	Ordered probit					OLS regression			Ord. logit	Multinomial logit	
Obs.	1,020	1,020	1,020	1,020	1,020	1,020	1,020	1,020	1,020	1,020	
(Pseudo-)R^2	0.247	0.237	0.235	0.222	0.219	0.482	0.501	0.434	0.246	0.246	
AIC	1648.4	1671.3	1671.6	1703.9	1708.2	3866.5	3829.0	3957.4	1651.4	1656.9	
BIC	1668.1	1691.0	1686.4	1723.6	1723.0	3886.2	3848.7	3977.1	1671.1	1696.3	
Likelihood	−820.2	−831.6	−832.8	−847.9	−851.1	−1929.3	−1910.5	−1974.7	−821.7	−820.4	

Notes: The table presents estimation results for each model, estimated on the full game sample. The first five columns compare the Transfermarkt model to alternative predictor models in an ordered probit model, with game results as the dependent variable. The next three columns compare the same models in an OLS regression, with goal margin as the dependent variable. The final three columns show a comparison between the ordered probit model and both an ordered logit and a multinomial logit model. The reference category in the multinomial logit is a win for team *i*.

*** Significance is denoted by: $p < 0.01$.

** Significance is denoted by: $p < 0.05$.

* Significance is denoted by: $p < 0.1$.

coefficient on the number of players is significant for all FIFA models and for the ELO model on goal margins, but not for the ELO probit model. To keep the rival models as similar as possible, I opted to report results where the variable is included in both the FIFA and ELO models.[7] In terms of goodness-of-fit, the FIFA ranking model clearly falls behind, as it has the lowest (pseudo) R^2 values and highest Akaike (AIC) and Bayesian (BIC) information criteria values throughout. The Transfermarkt model has a better fit than the ELO ratings in the ordered probit, but this is reversed for the goal margin regressions. Comparing the ordered probit to the ordered and multinomial logit models reveals that the ordered probit obtains a better model fit (lower AIC and BIC). In terms of parameter estimates, we find similar results across all link functions (note that the interpretation of the signs is reversed in the multinomial logit). Given these findings, the remainder of this paper continues to report results for the ordered probit.

4. Forecasting performance

4.1. Head-to-head comparisons of prediction accuracy

I began my comparison of the forecasting performances of the Transfermarkt model and its rivals by calculating the Brier score. The Brier score for a game between teams i and j played at time t is a quadratic loss function of the form

$$BS_{ijt} = \frac{1}{3}\Big(\big(f_{ijt-1}(win) - win_{ijt}\big)^2 + \big(f_{ijt-1}(draw) - draw_{ijt}\big)^2 + \big(f_{ijt-1}(loss) - loss_{ijt}\big)^2\Big),$$

where f_{ijt-1} represents the forecasted probability for each outcome based on the information available before the game starts. The outcome variables $win_{ijt}/draw_{ijt}/loss_{ijt}$ take values of one if the game ends in a win/draw/loss for team i, and zero otherwise. As with any loss function, lower values of the Brier score indicate better forecasting performances. Since Brier score values can be compared across models at the level of individual observations, the significance of differences in prediction accuracy can be examined through pairwise t-tests (see Hvattum & Arntzen, 2010 for more detail on this). Moreover, Strumbelj (2014)[8] presents a decomposition of the Brier scores of sports predictions into three dimensions, namely reliability (REL), generalized resolution (GRES) and uncertainty (UNC). Prediction models with lower scores on the reliability dimension are generally preferred, because this dimension has an interpretation similar to that of a loss function. For the generalized resolution, a higher score indicates a better forecasting performance. Finally, the uncertainty component is an inherent feature of the data and does not depend on the prediction model used.

A second criterion for the forecasting performance (e.g., Goddard, 2005) is the pseudo-likelihood statistic. This measure first takes the log of the predicted probabilities for the true outcome of each observation, i.e.,

$$PL_{ijt} = \log\big(f_{ijt-1}(win) * win_{ijt} + f_{ijt-1}(draw) * draw_{ijt} + f_{ijt-1}(loss) * loss_{ijt}\big).$$

The geometric mean of this expression over the forecasted sample is then used for comparison. Unfortunately, there

[7] Experiments with models that excluded the variable led to similar conclusions in terms of forecasting performances.

[8] The decomposition formulas are provided by Strumbelj (2014). In this application, I perform the calculations based on deciles of the predicted probabilities.

Table 4
Comparison of the forecasting performances of the rival prediction models.

Sample	Games	Forecaster	Brier score		Brier decomposition			Pseudo likelihood	Success ratio
			Total	p-val. vs. TM values	REL	GRES	UNC		
Results model									
Last 60% of games	592	TM values	0.1578		0.0034	0.0581	0.2124	0.4485	0.6503
		Betting odds	0.1558	0.1920	0.0018	0.0584		0.4491	0.6334
		ELO model	0.1613	0.0642	0.0034	0.0547		0.4402	0.6318
		FIFA model	0.1645	0.0042	0.0041	0.0520		0.4341	0.6199
		Null model	0.2162	0.0000	0.0069	0.0031		0.3431	0.4628
All games after match day 20	961	TM values	0.1584		0.0020	0.0555	0.2119	0.4448	0.6420
		Betting odds	0.1556	0.0495	0.0020	0.0582		0.4485	0.6358
		ELO model	0.1609	0.0846	0.0021	0.0531		0.4386	0.6327
		FIFA model	0.1643	0.0018	0.0026	0.0501		0.4327	0.6160
		Null model	0.2157	0.0000	0.0050	0.0012		0.3439	0.4641
Last 60% of UEFA games	477	TM values	0.1499		0.0036	0.0666	0.2128	0.4659	0.6667
		Betting odds	0.1482	0.2422	0.0030	0.0677		0.4641	0.6520
		ELO model	0.1539	0.0722	0.0032	0.0621		0.4556	0.6520
		FIFA model	0.1585	0.0022	0.0030	0.0573		0.4466	0.6352
		Null model	0.2149	0.0000	0.0056	0.0036		0.3454	0.4486
Margin model									
Last 60% of games	592	TM values	0.1574		0.0030	0.0581	0.2124	0.4487	0.6436
		Betting odds	0.1558	0.2305	0.0018	0.0584		0.4491	0.6334
		ELO model	0.1606	0.0820	0.0023	0.0542		0.4417	0.6318
		FIFA model	0.1672	0.0001	0.0041	0.0494		0.4282	0.6030
		Null model	0.2162	0.0000	0.0069	0.0031		0.3431	0.4628

Notes: The table compares the forecasting performances of alternative forecasting models over the last 60% of all games, the last 60% of games between UEFA countries and all games after the 20th match day, which are the largest feasible datasets given the need to establish initial model estimates. The underlined values are the test value for the best performing model in each comparison. The Brier scores are taken as averages per game and per outcome. The reported p-values are derived from one-sided t-tests on the differences in Brier scores, where observations are paired at the game level. In each case, the reference value is the Transfermarkt model. The decomposition into reliability (REL), generalized resolution (GRES) and uncertainty (UNC) is based on decile partitions in the formulas given by Strumbelj (2014).

is no test for the significance of the differences in this measure across models. Unlike the Brier score, higher values of the pseudo likelihood indicate better forecasting performances.

Finally, I examine the success ratio of each prediction model. Unlike the measures above, the success ratio looks at discrete predictions of game outcomes (i.e., a loss, draw or win). First, I define the predicted outcome for each game as the outcome that attains the highest predicted probability in the model. Then, a game prediction is labeled a 'success' when the predicted outcome of a game equals the observed outcome. Finally, the total number of successes in the sample is divided by the total number of games. As such, I obtain a relative measure, bounded by zero and one, which is referred to commonly as the success ratio. Note that higher success ratios again indicate better forecasting performances.

Table 4 contains the results for selected subsamples of the dataset. To allow for a fair comparison, the estimated models require a 'learning' period in order to increase the precision of the estimated coefficients in the estimation algorithm. Thus, Table 4 shows the forecasting accuracy measures calculated on the final 60% of the games. Here, all games are first ordered by the date they took place. Then all games up until the 40th percentile according to this ordering are used to form initial estimates. These are then updated before each game day in the final 60% of the dataset. In this subsample, the Transfermarkt valuations comfortably outperform the null, FIFA and ELO models, yielding significantly lower Brier scores and higher pseudo likelihoods and success ratios throughout. The betting odds produce better probabilistic predictions, as is evidenced by both the Brier scores and pseudo likelihoods; however, the difference is not significant at conventional significance levels. In terms of forecasting discrete game outcomes, the success ratio indicates that the Transfermarkt model has a slight advantage over the betting odds. When looking at the split-up of the Brier scores, the betting odds outperform all rival models in each dimension. The Transfermarkt model comes second in both reliability and resolution, outdoing the ELO, FIFA and null models in each respect.

The bottom panel of Table 4 repeats the same analysis for the model based on goal margins. The results are mostly similar. The betting odds beat the forecasting accuracy of the Transfermarkt model in every measure except the success ratio. In turn, the Transfermarkt model leads the ELO, FIFA and null models in almost every accuracy measure. The exceptions here are the Brier score split-ups, where the Transfermarkt values achieve a higher resolution and better absolute performance than the ELO model, but a lower reliability.

The second panel of Table 4 checks the robustness of these results for varying choices of the learning period by depicting the forecasting accuracy for the largest possible subsample, i.e., with a minimal learning period. For this subsample, we start predicting after just 20 match days, in this case at the 60th game. Here, the betting odds have a clear advantage, because they do not require any coefficients to be estimated. This shows in the results, with the Brier score favoring the betting odds model significantly at the 5% level. However, the Transfermarkt values still outperform the alternative estimated models. As such, the results found for the last 60% of games do not seem to be sensitive to the exact choice of the learning period.

Since Transfermarkt users primarily follow European clubs (see Table 7), the crowd might be wiser when it is forecasting games between UEFA countries. To look into this, Table 4 also reports the forecasting results for a subsample of games played between UEFA countries. In this subsample, the odds still yield better Brier scores, but the Transfermarkt model obtains higher pseudo likelihoods and success ratios. The results relative to other rival models remain largely unchanged.

4.2. Analysis of rival model forecast errors

One alternative to the head-to-head comparisons above is to regress the forecast error from a given rival predictor on the difference in average Transfermarkt valuations between the countries in the game being forecast. Then, the regression equation for a given game outcome y_{ijt} (loss-win-draw) is

$$f_{ijt-1}\left(y_{ijt}\right) - y_{ijt} = \beta_0 + \beta_v\left(v_{it} - v_{jt}\right) + \varepsilon_{ijt}.$$

A significant estimate for β_v indicates that the Transfermarkt values contain information that improves the prediction accuracy beyond that achieved by the rival predictor. I run this regression for each outcome and each rival predictor separately, using similar subsamples of the dataset as in the head-to-head comparisons above.

Table 5 presents the results of this exercise. In each case, the Transfermarkt value could improve the prediction accuracy for wins and losses, but not draws. The most extensive gains are found relative to the null and FIFA models, followed by the betting odds. The ELO model has the least significant results. Hence, the Transfermarkt values contain less additional information with respect to the ELO ratings than with respect to any of the other models.

4.3. Scatterplots of predicted probabilities

Having established that the Transfermarkt valuations allow improvements upon established forecasters, I next examine how its predictions diverge from those based on rival predictors. Inspired by Strumbelj (2014, 2016), I generate plots of the predicted probability of each game outcome in the model against those obtained by alternative prediction models. In the resulting scatterplots (see Fig. 1), each x-axis shows the Transfermarkt predictions, while each y-axis shows one of the alternative predictions.

The largest diversions between the Transfermarkt and betting odds predictions are in the tails of the predicted probability distribution. At the lower end of the draw probability, the Transfermarkt model predicts lower probabilities, while its predictions at the upper ends of the loss and win probabilities are clearly higher. The differences between the Transfermarkt and the ELO and FIFA models are less straightforward to characterize. On the whole, the FIFA model seems to diverge more from the Transfermarkt predictions, because the dispersion around the 45° line is slightly wider. However, there is no clear point along the distribution at which this effect is more or less pronounced for either model.

4.4. Comparing betting returns

Finally, I calculate the differences between the returns on several betting strategies based on both the Transfermarkt and rival forecasting models. The betting returns are calculated using the raw betting odds, as summarized in Table 2. These include the 'over', i.e., the profit margin of the bookmaker. Thus, in expectation, the returns on 'naïve' betting strategies will be negative and a successful strategy may simply achieve a less negative return.

Following Hvattum and Arntzen (2010), I analyze three betting strategies. Under the first strategy, dubbed 'unit bet', the bettor puts a one-unit stake on each bet with a positive expected return. Using similar notation to before, the expected return $R_{y_{ijt}}$ for a bet on an outcome y_{ijt} in a game between i and j at time t is given as

$$E\left[R_{y_{ijt}}\right] = odd\left(y_{ijt}\right) f_{ijt-1}\left(y_{ijt}\right) - 1.$$

As multiple bets in a single game may yield positive expected returns, the number of bets typically does not equal the number of games. However, some bets on very probable outcomes yield very low returns under this strategy, because uniform stakes are bet regardless of the odds placed by the bookmaker. A second strategy, called 'unit win', corrects for this by adjusting the stake on each bet to yield a return of one unit if the bet is successful. To achieve this, a stake equal to $\frac{1}{odd(y_{ijt})-1}$ has to be put on each bet with a positive expected return. One drawback here is that some bets require very high stakes to yield a unit return, which puts more of the bettor's capital at risk. The third strategy that I explore is betting according to the Kelly Criterion. Here, the bettor again puts out a stake on each bet with a positive expected return, but the size of the stake is designed to maximize the long-run capital growth of the bettor. The stake can be calculated using the formula $\frac{f_{ijt-1}(y_{ijt})(odd(y_{ijt})-1)-(1-f_{ijt-1}(y_{ijt}))}{odd(y_{ijt})-1}$. Note that the probability of success/failure in this betting strategy determines not only when to bet, but also what to stake.

Table 6 summarizes the resulting betting returns using similar subsamples to before. The Transfermarkt model clearly attains higher returns than any of the rival prediction models. Unfortunately, though, it is not possible to assess the statistical significance of this difference in returns, because the betting strategies select different subsets of games to bet on for each prediction model. However, with the advantage over its closest competitor ranging between 1% and 17%, it seems fair to state that this result is both sizeable and robust for each choice of subsample and betting strategy. It is less clear, though, which of the alternative models should be preferred to the others. Neither the FIFA model nor the ELO model consistently yields the second-highest average return. Comparing across betting strategies, it clearly pays to adjust the stakes to the odds, because the unit win and Kelly strategies consistently outperform the unit bet strategy. However, the results between the unit win and Kelly strategies are less clear-cut. In most cases the positive unit win returns are higher than the positive

Table 5
Regression results of the forecast errors of the rival prediction models on the average Transfermarkt value.

Forecast error	Betting odds			ELO			FIFA			Null		
Outcome	Win	Draw	Loss	Win	Draw	Loss	Win	Draw	Loss	Win	Draw	Loss
Sample: Last 60% of games												
Av. TM value	0.013**	−0.000	−0.013**	0.008	0.003	−0.011*	0.014**	0.001	−0.015**	0.104***	−0.009	−0.096***
	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)
Constant	0.001	0.002	−0.003	−0.029*	0.007	0.022	−0.025	0.005	0.019	0.067***	0.005	−0.072***
	(0.017)	(0.017)	(0.015)	(0.017)	(0.017)	(0.016)	(0.017)	(0.017)	(0.016)	(0.017)	(0.017)	(0.016)
Observations	592	592	592	592	592	592	592	592	592	592	592	592
R^2	0.007	0.000	0.008	0.002	0.000	0.005	0.007	0.000	0.010	0.288	0.003	0.279
Sample: All games after match day 20												
Av. TM value	0.013***	−0.001	−0.012***	0.007	0.004	−0.011**	0.014***	0.001	−0.015***	0.099***	−0.008*	−0.091***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Constant	0.000	−0.005	0.005	−0.028**	−0.008	0.036***	−0.022	−0.012	0.034***	0.074***	−0.012	−0.061***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)
Observations	961	961	961	961	961	961	961	961	961	961	961	961
R^2	0.007	0.000	0.007	0.002	0.001	0.005	0.008	0.000	0.010	0.277	0.003	0.266
Sample: Last 60% of UEFA games												
Av. TM value	0.016**	−0.003	−0.014**	0.009	−0.000	−0.009	0.015**	−0.002	−0.013*	0.110***	−0.010	−0.099***
	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Constant	−0.002	−0.003	0.005	−0.031*	0.007	0.024	−0.024	0.005	0.019	0.060***	−0.007	−0.053***
	(0.018)	(0.018)	(0.017)	(0.018)	(0.018)	(0.017)	(0.019)	(0.018)	(0.018)	(0.019)	(0.019)	(0.018)
Observations	477	477	477	477	477	477	477	477	477	477	477	477
R^2	0.012	0.000	0.009	0.003	0.000	0.004	0.010	0.000	0.008	0.340	0.004	0.309

Notes: The table reports OLS regression results where the dependent variable is the forecasting error for a given outcome and forecasting model. The explanatory variables included are a constant term and the log difference in average Transfermarkt value between the countries. The subsamples are similar to those in Table 4.

*** Significance is denoted by: $p < 0.01$

** Significance is denoted by: $p < 0.05$

* Significance is denoted by: $p < 0.1$.

Table 6
Betting returns: alternative forecasting models.

Sample	Games	Model	Kelly bet		Unit win		Unit bet		Number of bets
			Mean return	Mean stake	Mean return	Mean stake	Mean return	Mean stake	
Last 60% of games	592	TM values	<u>0.38%</u>	0.13	<u>8.70%</u>	4.24	<u>1.91%</u>	1.00	574
		TM values margin	0.38%	0.12	4.82%	3.88	−0.40%	1.00	588
		ELO	−0.97%	0.13	2.44%	5.02	−15.36%	1.00	544
		ELO margin	−1.20%	0.10	−0.70%	4.24	−18.79%	1.00	523
		FIFA	−1.13%	0.15	3.99%	4.14	−19.36%	1.00	635
		FIFA margin	−1.67%	0.16	2.69%	4.15	−21.27%	1.00	644
All games after match day 20	961	TM values	<u>0.11%</u>	0.12	<u>5.06%</u>	3.93	<u>−3.56%</u>	1.00	944
		ELO	−0.89%	0.11	−3.16%	3.94	−13.98%	1.00	908
		FIFA	−1.06%	0.13	2.42%	4.64	−19.53%	1.00	1089
Last 60% of UEFA games	477	TM values	<u>3.93%</u>	0.16	<u>10.93%</u>	5.71	<u>1.10%</u>	1.00	426
		ELO	−1.01%	0.16	3.93%	4.90	−15.80%	1.00	466
		FIFA	−1.49%	0.15	1.76%	6.27	−28.88%	1.00	512

Notes: The table compares the betting returns and betted stakes under three alternative betting strategies, namely Kelly criterion betting, unit win and unit bet. The subsamples in the comparison are similar to those reported in Table 4. The highest returns for each strategy and subsample are underlined. Note that the rules often require betting on more than one outcome per game, which explains why the number of bets may exceed the number of game observations.

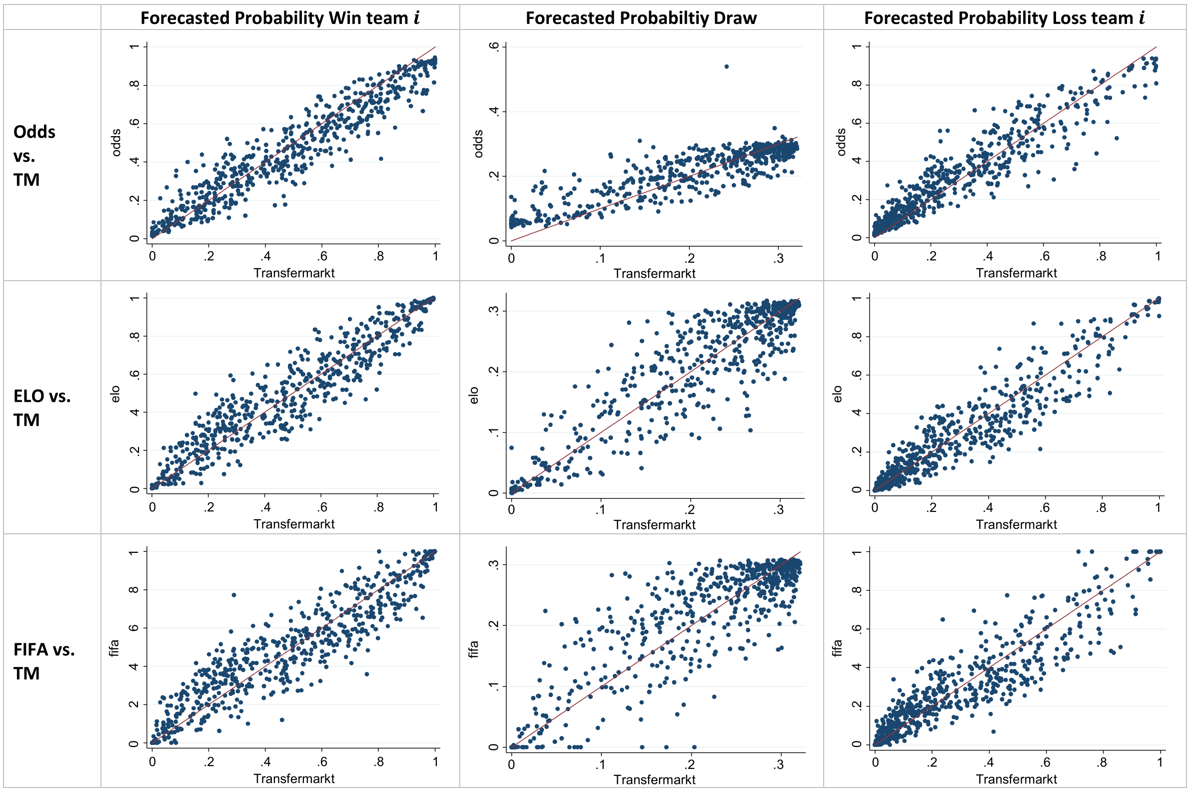

Fig. 1. Scatterplots of forecasted probabilities of a win, draw or loss.

Kelly returns, but this gain comes at the expense of a higher capital exposure, as is clear from the higher average stakes and larger potential losses under unit win.

5. Wishful thinking bias in Transfermarkt valuations

In the Transfermarkt setting, I interpret a wishful thinking bias to mean that crowd members overestimate the quality of certain players because they prefer those players to be of high value.[9] I infer information on the preferences of crowd members by exploiting the fact that Transfermarkt users almost always report their favorite club(s) on their profile. Clearly, users should prefer high quality players to be playing at their favorite clubs, rather than at

9 This interpretation is linked closely to that of Massey et al. (2011), who show that people persistently overestimate the quality (i.e., win probability) of their preferred team.

Table 7
Top 15 teams by user affiliations.

Rank	Team	League	Country	No. of fans	% total fans
1	FC Bayern München	1.Bundesliga	Germany	5913	12.47%
2	Borussia Dortmund	1.Bundesliga	Germany	5174	10.91%
3	Hamburger SV	1.Bundesliga	Germany	2215	4.67%
4	FC Schalke 04	1.Bundesliga	Germany	2131	4.49%
5	SV Werder Bremen	1.Bundesliga	Germany	2077	4.38%
6	Borussia Mönchengladbach	1.Bundesliga	Germany	1893	3.99%
7	Galatasaray Istanbul	Süper Lig	Turkey	1414	2.98%
8	VfB Stuttgart	1.Bundesliga	Germany	1348	2.84%
9	1.FC Kaiserslautern	2.Bundesliga	Germany	1082	2.28%
10	1.FC Köln	2.Bundesliga	Germany	1006	2.12%
11	Eintracht Frankfurt	1.Bundesliga	Germany	970	2.05%
12	Fenerbahce Istanbul	Süper Lig	Turkey	961	2.03%
13	1.FC Nürnberg	1.Bundesliga	Germany	822	1.73%
14	Hertha BSC	1.Bundesliga	Germany	784	1.65%
15	Hannover 96	1.Bundesliga	Germany	768	1.62%

Notes: The table lists the numbers of users on Transfermarkt that support particular clubs for the top 15 clubs. This refers to the situation on 31/10/2013. Note that several teams may have switched divisions since because of promotion or relegation.

competing clubs. Thus, if users suffer from wishful thinking, I expect them ceteris paribus to post higher valuations for the players of clubs they support. Even then, a sufficiently diverse crowd might alleviate individual biases (see Simmons et al., 2011) through countervailing preferences cancelling out. However, if such is not the case, the value of players at well supported clubs could be overstated relative to those playing on poorly supported teams.

Table 7 reports on the top 15 teams in terms of support among Transfermarkt users.[10] Overall, the distribution is skewed heavily towards the most popular German and Turkish teams. Taken together, Bayern Munich and Borussia Dortmund, the two top teams in Germany, claim more than 20% of all fan support. The top 15 clubs jointly command the support of around 60% of all users who reveal their preferences. This distribution of fans reflects the German origin of the website, as the support for Turkish teams can be tied back to the large minority of Germans who are of Turkish descent. Overall, Table 7 suggests that club preferences in the crowd are largely similar, making it less likely that the diversity of the crowd will alleviate wishful thinking.

I assess the presence of potential wishful thinking bias in the valuations by employing a methodology similar to the market test for racial discrimination that was introduced by Szymanski (2000). This test exploits the idea that controls for potential biases should not add significant explanatory power to a regression of game outcomes on unbiased predictors. However, if the predictor is biased, the explanatory power of the model can be improved by controlling for characteristics of the team, which should be irrelevant to game results in the absence of biases (e.g., Szymanski, 2000), uses the number of black players on the team roster). The signs on the estimated parameters can then be interpreted as an indication of the direction of the bias.

I propose four measures for gauging how well the clubs of the players in a national team are supported at the time of a game. First, I count the average numbers of registered users who support the current clubs of the players in each national team. For the second measure, I divide this number by the total number of users, to create a percentage measure of support. A third alternative is to count the numbers of players who are active at popular clubs, which I define as clubs that are supported by more than 1% of all users. As the final measure, I calculate the average team value weighted by the relative support of each player's club. This results in a higher value for teams in which the more valuable players play at relatively more popular clubs. Table 2 shows summary statistics for all four bias measures. However, before introducing these in the match outcome and goal margin models, I take the natural logarithm for the average number of fans, number of players at popular clubs and weighted team value.

Table 8 provides the estimation results for this analysis. First note that I find no significant negative estimates for the fan support measures in any of the models reported in Table 8. Hence, the estimation results do not point in the direction of a wishful thinking bias. On the contrary, there are some indications that the crowd undervalues players from teams with large support in the crowd. Furthermore, the estimates for the variables included previously, namely the average value and roster size, are never significantly different from those of the baseline model. In other words, the introduction of the bias controls does not make any fundamental change to the way in which valuations predict the game outcomes. Finally, the AIC and BIC reveal that the introduction of these measures enhances the explanatory power of the models slightly. Therefore, introducing these variables could still make a marginal improvement in the predictive performance of the model.

6. Conclusions

This paper has provided a test of the wisdom of crowds 'in the field' by assessing how well player valuations from non-expert users of an online platform (Transfermarkt) predict the results of international soccer games. A simple model that contains nothing but the average Transfermarkt valuation, the number of players and a home advantage

[10] The figures reflect the situation on 31/10/2013. Unfortunately, more recent versions of the website do not report overviews of the user statistics.

Table 8
Estimation results: regression model with wishful thinking controls.

Game result	Game result model					Goal difference model				
Average value	0.450***	0.447***	0.422***	0.433***	0.479***	0.552***	0.500***	0.518***	0.516***	0.467***
	(0.024)	(0.036)	(0.026)	(0.031)	(0.047)	(0.020)	(0.029)	(0.022)	(0.025)	(0.036)
Number of fans		0.003					0.061**			
		(0.023)					(0.025)			
Rel. number of fans			10.012***					16.444***		
			(3.765)					(4.051)		
Players support teams				0.044					0.139**	
				(0.049)					(0.056)	
Support weighted value					−0.013					0.050***
					(0.018)					(0.018)
Number of players	1.775**	1.778**	2.076**	1.789**	1.735**	2.323**	2.440***	2.724***	2.410***	2.386***
	(0.880)	(0.880)	(0.898)	(0.879)	(0.881)	(0.918)	(0.917)	(0.916)	(0.916)	(0.915)
Home advantage	0.281***	0.281***	0.280***	0.281***	0.281***	0.408***	0.406***	0.407***	0.408***	0.408***
	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)
Model	Ordered probit					OLS				
Observations	1,020	1,020	1,020	1,020	1,020	1,020	1,020	1,020	1,020	1,020
(Pseudo-)R^2	0.247	0.247	0.250	0.247	0.247	0.482	0.485	0.490	0.485	0.486
AIC	1648.4	1650.3	1642.9	1649.6	1649.9	3866.5	3862.7	3852.1	3862.3	3860.6
BIC	1668.1	1675.0	1667.5	1674.2	1674.5	3886.2	3887.4	3876.7	3887.0	3885.2

Notes: Standard errors are given in parentheses. Significance levels are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

predicts the performance of a national team better than more traditional predictors of soccer results, namely the FIFA ranking and ELO ratings. Furthermore, comparing the returns to different betting strategies indicates that the gain in predictive power is economically meaningful. I also exploit data on crowd member preferences in order to examine wishful thinking bias in the valuations, but fail to find evidence that the valuations of players at popular clubs are biased upward. Taken together, these findings suggest that crowd opinions can be a rich source of information for forecasting purposes, even when they are not generated in controlled environments.

There are two natural extensions to the analysis presented in this paper. First, it remains to be seen whether the Transfermarkt valuations suffer from other biases. It is well documented that betting odds may suffer from favorite-longshot and sentiment biases (e.g., Franck, Verbeek, & Nuesch, 2011; Simmons & Forrest, 2008). It is clear that wishful thinking can be thought of as a type of sentiment bias. However, in the betting literature sentiment bias may express itself in the behaviors of both odds-setters and bettors. Such is not the case for Transfermarkt valuations, where the platform aggregators have little incentive to react strategically to potential biases in the crowd's preferences. Still, one could wonder to what extent Transfermarkt values are driven by media attention, for example. Second, it would be interesting to check whether the reported results carry over to other settings in soccer. For example, can Transfermarkt values also generate accurate predictions for games between clubs?

Acknowledgments

The author gratefully acknowledges financial support from the EUR fellowship program and the Flanders Research Foundation (FWO) (FB 6924), and thanks Tom Peeters for his assistance in gathering data. This project benefited from presentations at the University of Reading, the Université de Liege HEC School of Business, the Universidad de Cantabria, the Mathematics in Sports Conference in Leuven, the European Conference on Sports Economics in Esbjerg and the WEAi meeting in Denver; as well as from comments by Bernd Frick, Scott Page, Ignacio Palacios-Huerta, James Reade, Rob Simmons, Stefan Szymanski and Bram van den Bergh. I thank the editor, an associate editor and two referees for their invaluable feedback and Cathy Morgan for careful copyediting.

References

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., et al. (2016). Distilling the wisdom of crowds: Prediction markets versus prediction polls. *Management Science*, *63*(3), 691–706.

Bloomberg, (2016).Ronaldo is worth four times the entire Hungarian national team. *Bloomberg*. Retrieved May 31st, 2016, from http://www.bloomberg.com/news/articles/2016-05-30/ronaldo-is-worth-four-times-the-entire-hungarian-national-team.

Brown, A., Rambaccussing, D., Reade, J. J., & Rossi, G. (2016) Using social media to identify market inefficiencies: evidence from Twitter and Betfair. Working Paper.

Bryson, A., Frick, B., & Simmons, R. (2013). The returns to scarce talent: Footedness and player remuneration in European soccer. *Journal of Sports Economics*, *14*(6), 606–628.

Budescu, D. V., & Chen, E. (2015). Identifying expertise to extract the wisdom of crowds. *Management Science*, *61*(2), 267–280.

Camara, F., & Dupuis, N. (2014) Structural estimation of expert bias: The case of movie critics. Working paper.

Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, *27*(5), 1367–1403.

Coates, D., Frick, B., & Jewell, T. (2016). Superstar salaries and soccer success: The impact of designated players in major league soccer. *Journal of Sports Economics*, *17*(7), 716–735.

Cowgill, B., & Zitzewitz, E. (2015). Corporate prediction markets: Evidence from Google, Ford, and Firm X. *Review of Economic Studies*, *82*(4), 1309–1341.

Economo, E., Hong, L., & Page, S. E. (2016). Social structure, endogenous diversity, and collective accuracy. *Journal of Economic Behavior and Organization*, *125*, 212–231.

Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, *21*, 551–564.

Forsythe, R., Rietz, T. A., & Ross, T. W. (1999). Wishes, expectations and actions: A survey on price formation in election stock markets. *Journal of Economic Behavior and Organization*, *39*, 83–110.
Franck, E., & Nüesch, S. (2010). The effect of talent disparity on team productivity in soccer. *Journal of Economic Psychology*, *31*, 218–229.
Franck, E., Verbeek, E., & Nuesch, S. (2010). Prediction accuracy of different market structures—bookmakers versus a betting exchange. *International Journal of Forecasting*, *26*, 448–459.
Franck, E., Verbeek, E., & Nuesch, S. (2011). Sentimental preferences and the organizational regime of betting markets. *Southern Economic Journal*, *78*(2), 502–518.
Frick, B., & Wicker, P. (2016). Football experts versus sports economists: Whose forecasts are better? *European Journal of Sports Science*, *16*(5), 603–608.
Garicano, L., Palacios-Huerta, I., & Prendergast, C. (2005). Favoritism under social pressure. *The Review of Economics and Statistics*, *87*, 208–216.
Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, *21*(2), 331–340.
Goldstein, D. G., McAfee, P. R., & Suri, S. (2014) The wisdom of smaller, smarter crowds. In *Proceedings of the fifteenth ACM conference on economics and computation* (pp. 471-488). New York.
Herm, S., Callsen-Backer, H.-M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sport Management Review*, *17*, 484–492.
Herzog, S. M., & Hertwig, R. (2011). The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. *Judgement and Decision Making*, *6*(1), 58–72.
Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, *26*, 460–470.
Lambertson, P. J., & Page, S. E. (2012). Optimal forecasting groups. *Management Science*, *58*(4), 805–810.
Li, C., & Liu, N. (2015) What to tell? Wise communication and wise crowd. Working Paper.
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*, *108*(22), 9020–9025.
Massey, C., Simmons, J. P., & Armor, D. A. (2011). Hope over experience: Desirability and the persistence of optimism. *Psychological Science*, *22*(2), 274–281.
Mollick, E., & Nanda, R. (2015). Wisdom or madness? Comparing crowds with expert evaluation in funding the arts. *Management Science*, *62*(6), 1533–1553.
Servan-Schreiber, E., Wolfers, J., Pennock, D. M., & Galebach, B. (2004). Prediction markets: Does money matter? *Electronic Markets*, *14*(3), 243–251.
Simmons, J. P., Nelson, L. D., Galak, J., & Frederick, S. (2011). Intuitive biases in choice versus estimation: Implications for the wisdom of crowds. *Journal of Consumer Research*, *38*, 1–15.
Simmons, R., & Forrest, D. (2008). Sentiment in the betting market on spanish football. *Applied Economics*, *40*(1), 119–126.
Smith, M. A., Pateon, D., & Williams, L. V. (2009). Do bookmakers possess superior skills to bettors in predicting outcomes? *Journal of Economic Behavior and Organization*, *71*, 539–549.
Stekler, H., Sendor, D., & Verlander, R. (2010). Issues in sports forecasting. *International Journal of Forecasting*, *26*, 606–621.
Strumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting*, *30*(4), 934–943.
Strumbelj, E. (2016). A comment on the bias of probabilities derived from betting odds and their use in measuring outcome uncertainty. *Journal of Sports Economics*, *17*(1), 12–26.
Surowiecki, J. (2004). *The wisdom of crowds*. New York: Doubleday.
Szymanski, S. (2000). A market test for discrimination in the english professional soccer leagues. *Journal of Political Economy*, *108*(3), 590–603.
Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *The Journal of Economic Perspectives*, *18*, 107–126.

Thomas Peeters is an assistant professor of applied industrial organization at the Erasmus School of Economics in Rotterdam, Netherlands. He is a research fellow at Tinbergen Institute and an associate member of the Erasmus Research Institute in Management (ERIM). He obtained his Ph.D. from the University of Antwerp in Belgium. Before joining Erasmus University, he worked at the Flanders Research Foundation and the University of Michigan. His work has previously been published in journals such as Economic Policy, the International Journal of Industrial Organization, Economic Inquiry and the *Journal of African Economies*.