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.