Innovative Applications of O.R.

The wisdom of amateur crowds: Evidence from an online community of sports tipsters☆

Alasdair Brown [a,*], J. James Reade [b,c]

[a] *School of Economics, University of East Anglia, Norwich NR4 7TJ, UK*
[b] *Department of Economics, University of Reading, Reading RG6 6AA, UK*
[c] *Institute for New Economic Thinking, University of Oxford, Oxford, UK*

A R T I C L E   I N F O

A B S T R A C T

We analyse the accuracy of crowd forecasts produced on Oddsportal, an online community of amateur sports tipsters. Tipsters in this community are ranked according to the betting return on their tips, but there are no prizes for accuracy. Nevertheless, we find that aggregated tips in this community contain information not in betting prices. A strategy of betting when a majority predict an outcome produces average returns of 1.317% for 68,339 events. The accuracy of these forecasts stems from the wisdom of the whole crowd, as selecting sections of the crowd based on experience or past forecast accuracy does not improve betting returns.

1. Introduction

Predicting the outcome of sporting events, particularly in a way unanticipated by bookmakers, is of prime interest to gamblers. Sports betting is estimated to be worth somewhere between 700 billion and 1 trillion worldwide per annum[1], which, even allowing for some exaggeration in that figure, clearly demonstrates that there are substantial sums at stake.

Traditionally, bettors may have decided to devise a model to forecast outcomes and see if these models produced information not in betting prices. Examples of academic work in this area are numerous and include Dixon and Coles (1997), Klaassen and Magnus (2003), Dixon and Pope (2004), Goddard and Asimakopoulos (2004), Easton and Uylangco (2010), and McHale and Morton (2011). More recently, and particularly after the publication of Surowiecki (2005) and the revival of the Galton (1907) 'wisdom of crowds' idea, there has been an interest in crowd-sourcing predictions. The wisdom of crowds operates on the premise that an averaging of forecasts eliminates individual prediction errors, and leads to greater accuracy.

Recent evidence suggests that there is indeed wisdom in the crowd when it comes to sports forecasting. Schumaker, Jarmoszko, and Labedz (2016) and Brown, Rambaccussing, Reade, and Rossi (2018) found that Twitter sentiment, or tone, could be leveraged to amass positive returns in English Premier League soccer betting. Peeters (2018) found that information from Transfermarkt valuations – where online users submit transfer valuations of soccer players – could be used to generate sizeable betting returns in matches.

In this paper we analyse predictions collected on Oddsportal, a betting comparison website which also hosts an online community of sports tipsters. Members of the Oddsportal community are ranked according to the betting return on their tips. The crowds on Oddsportal are smaller than Twitter, for example, but because of the ranking criteria these crowds are specifically tasked with identifying when betting markets are mispriced (i.e. when there is information not in betting prices). This setting therefore provides small, but highly-targeted, crowd-sourced predictions.

We set out to answer two questions. (1) Can Oddsportal tips be used to improve betting returns? And (2) does the informational content of these crowd-sourced predictions stem from the full crowd, or a subset of skilled or experienced individuals?

We find that Oddsportal tips can be used to generate positive betting returns. A strategy of betting when a majority of tipsters predict an outcome produces average returns of 1.317% for 68,339 events. This shows that even amateur crowds, with no tangible

prizes for accurate forecasting, still produce information not in market prices. In further analysis we find that limiting the crowd to tipsters with more experience (more past tips) or more skill (higher historical returns on their tips) does not improve betting returns. This suggests that the accuracy of these crowd forecasts stems from the whole crowd, rather than just a select few tipsters.

2. Related literature

Our work firstly contributes to the literature on sports tipsters. Deschamps and Gergaud (2005, 2008) and Forrest and Simmons (2000) found evidence that tipsters produced highly dispersed forecasts which ignored public information. The professional tipsters in Forrest and Simmons (2000) and Spann and Skiera (2009) performed poorly when considered against betting markets, but did improve these market forecasts when used in conjunction. For example, in Spann and Skiera (2009) the betting market predicted the correct winner 53.69% of the time in isolation, but 56.52% of the time when used in combination with aggregated tips. However, because of the large margins in the market considered, these combined forecasts did not produce positive returns (−9.08%). Reade (2014) considered the accuracy of Oddsportal soccer tips, but considered these tips as stand-alone forecasts rather than predictions of betting market mispricing.

Our paper also contributes to a literature on the wisdom of experts and laypeople (people without professional or specialist knowledge) in forecasting sports events. Experts outperformed laypeople in Pachur and Biele (2007), in part because laypeople forecasted based simply on name recognition (Goldstein & Gigerenzer, 2002). O'Leary (2017) found that a crowd of laypeople were more accurate than a smaller (n = 5) group of experts, but did not evaluate whether these layperson predictions could produce positive betting returns. Herzog and Hertwig (2011) examined whether laypeople predictions could add to sports betting prices, and found not. Amateur tipsters on Oddsportal, on the other hand – who are not professionals but have self-selected themselves into predicting these events – produce information not in betting prices and, due to lower margins, can yield positive returns (1.317%).

Our paper is also related to a more general literature on the efficiency of betting prices, surveyed in Vaughan Williams (2005). For example, Ma, Tang, McGroarty, Sung, and Johnson (2016) found that horse race betting markets failed to incorporate an important variable – the time since a horse last ran – into betting prices. Hwang and Kim (2015) found that betting market prices were poorly calibrated for extreme probabilities. The betting returns generated by Oddsportal predictions in our study are also indicative of market inefficiency, as information contemporaneously available to individuals and the crowd is not incorporated into betting prices.

This may be surprising to some, as Hayek (1945) argued that markets are well-suited to aggregate dispersed information, and the efficient market hypothesis of Fama (1970) may imply that markets should hold primacy in matters of forecasting. Part of the reason of the success of Oddsportal may be due to the payoff structure of forecasting contests. (Oddsportal, in effect, run an infinite-horizon forecasting contest with rankings determined by the betting returns on tips, albeit with no prizes). Pfeifer, Grushka-Cockayne, and Lichtendahl (2014), Ottaviani and Sørensen (2005; 2006) and Lichtendahl, Grushka-Cockayne, and Pfeifer (2013) modelled forecasting contests and showed that forecasters will overweight their private information in a bid to win the contest. Put another way, the convexity of the prize schedule (e.g. winner-takes-all) encourages individuals to take risks and base their forecasts (solely) on their private information. As a result, individual

forecasting errors may be large (as public information is ignored), but aggregated crowd forecasts will be more accurate as there is less repeated counting of public information ('public knowledge bias'). In markets, on the other hand, there is perhaps less incentive to ignore public information – as payoffs are not convex or dictated by relative rank – and therefore this may explain why forecasting contests can add information to that produced within markets.

What is perhaps most striking about Oddsportal is that there are no tangible prizes – only the intangible esteem of ones' online peers – and yet there is still information contained in the forecasts made by these amateur tipsters.

Our results remind us of the findings in Servan-Schreiber, Wolfers, Pennock, and Galebach (2004), where play prediction markets performed as well as real-money prediction markets. Tipsters in our setting have weaker incentives than bookmakers and other participants in betting markets. Nevertheless, important information is produced in this low-stakes tipster community. This suggests that high-powered incentives are not the only consideration when generating accurate crowd forecasts.

In relation to the recent literature on crowd-sourced predictions of sporting events (e.g. Schumaker et al., 2016; Brown et al., 2018; and Peeters, 2018), the Oddsportal setting we study in this paper allows for two innovations. Firstly, we can examine whether targeted forecasts – on whether the bets are mispriced – can offset smaller crowd sizes and still produce profitable crowd forecasts. (It appears that they can). Secondly, as we have rich data on the full history of tipster predictions, we can analyse whether smaller crowds, made up of only the most experienced or skillful tipsters, can outperform the forecasts produced by the whole crowd. (It would appear not).

3. Data

The setting for our study is oddsportal.com, a website founded in 2008. The website serves two functions. Firstly it has an odds comparison function, providing the quoted odds from more than 80 bookmakers plus two betting exchanges, Betfair and Matchbook. The odds relate to 22 different sports from soccer (Association Football) to mixed martial arts. A screenshot of the odds comparison for the 5th February 2017 English Premier League match between Leicester City and Manchester United can be found in Fig. 1. For illustration, we display only the first 18 bookmakers. The remaining bookmakers and the two betting exchanges were to be found underneath. In addition to the odds on the match outcome (home win/draw/away win), which we display, the website collates odds from the same bookmakers on the correct score and a range of other exotic bets.

The second function of oddsportal.com is the hosting of an online community of sports tipsters. Registered users of the site can predict sporting outcomes, and they are then ranked according to the betting return on their tips. (The average bookmaker price is used when calculating the return). In Fig. 2 we display a screenshot of the leaderboard of tipsters. There are various ways to filter this list, but in this particular case the tipsters are sorted by ROI (return on investment), and the list is restricted to those users with at least 50 historical predictions across all sports. Users may choose to keep their picks secret (with an eye indicating that this choice has been made), but the majority of users choose to share their picks with other users. This brings us to the Tips Feed (see Fig. 3), where the most recent tips by all users (or just users you are following) are displayed. Users can comment on or 'Like' the tips made by other users, and this facility creates a social network feel to the website.