Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 26 (2010) 482–488
international journal of forecasting
www.elsevier.com/locate/ijforecast

Online bookmakers' odds as forecasts: The case of European soccer leagues

E. Štrumbelj*, M. Robnik Šikonja

Faculty of Computer and Information Science, Tržaška 25, 1000 Ljubljana, Slovenia

Abstract

In this paper we examine the effectiveness of using bookmaker odds as forecasts by analyzing 10,699 matches from six major European soccer leagues and the corresponding odds from 10 different online bookmakers. We show that the odds from some bookmakers are better forecasts than those of others, and provide empirical evidence that (a) the effectiveness of using bookmaker odds as forecasts has increased over time, and (b) bookmakers offer more effective forecasts for some soccer leagues for than others.

Keywords: Sports forecasting; Brier score; Statistical tests; Soccer; Betting

1. Introduction

Betting on sports and other types of events attracts the attention of casual bettors, pundits and scientists alike. Betting markets are of particular interest to researchers because there are many similarities between wagering in betting markets and trading in financial markets. These markets consist of different bookmakers, who offer odds for the outcomes of uncertain events, and bettors, who decide whether or not to bet on events. The actual value of the bet is known once the uncertainty has been resolved (i.e., the actual outcome of an event is known). The main goal of bookmakers is to make a profit. This drives them to set the odds high enough to be competitive, but low enough that betting on them is not profitable. Therefore, though the posted odds may not reflect the bookmakers' true probabilistic beliefs, they can still be viewed as probabilistic assessments of a sporting event's outcome, or, in other words, as *forecasts*. So, how effective are these forecasts? Are the odds from some bookmakers better forecasts than those from other bookmakers? Are bookmakers' odds equally effective for different leagues, and does this change over time? In this paper we try to answer these and other similar questions by analyzing soccer matches from six major European soccer leagues, together with the fixed-odds offered for the matches by ten different online bookmakers.

* Corresponding author.
E-mail addresses: erik.strumbelj@fri.uni-lj.si (E. Štrumbelj), marko.robnik@fri.uni-lj.si (M.R. Šikonja).

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2009.10.005

1.1. Related work

Most related work on sports betting has focussed on the efficiency of betting markets. Earlier work was predominantly about parimutuel horse-race betting. Unlike fixed-odds betting, where a price is fixed at the moment of betting, parimutuel prices are not known until the market closes. Most authors find evidence that the parimutuel horse-racing market is not efficient and provide profitable strategies, for example Bolton and Chapman (1986), Hausch, Ziemba, and Rubinstein (1981), Hausch and Ziemba (1995), and Lo (1995); although some add that the results should be verified on larger data sets to provide more confident results, for example Asch, Malkiel, and Quandt (1984). Hausch and Ziemba (1990), Ali (1998), and Swindler and Shaw (1995) argue that parimutuel markets are indeed weakly efficient. Also relevant is the work of Dixon and Pope (2004), which is a continuation of the work of Pope and Peel (1988). Dixon and Pope focus on the UK fixed-odds soccer betting market data from 3 bookmakers (years 1993–1996) and find several market inefficiencies, including a reverse favorite-longshot bias. Note that we do not focus on market efficiency in this paper. Instead, we focus on treating bookmaker odds as forecasts and investigating their quality. Related work has provided evidence that bookmaker odds are a very good source of match outcome forecasts. For example, Forrest, Goddard, and Simmons (2005) analyze English soccer games and find that bookmakers' forecasts cannot be outperformed by statistical models. Several related forecasting publications analyze and incorporate knowledge other than betting odds into forecasting match outcomes. For example, Scheibehenne and Broderb (2007) and Pachur and Biele (2007) show that even name recognition by laymen offers some information about the outcome of a sports event, though less than expert knowledge or betting odds. Andersson, Edman, and Ekman (2005), Song, Boulier, and Stekler (2007), and Forrest and Simmons (2000) show that both lay and expert predictions are outperformed by statistical models, which are in turn usually worse than bookmaker odds.

1.2. Notation

Let $\mathcal{X}$ and $\mathcal{B}$ be sets of matches and bookmakers, respectively. A regular *soccer match*, $x \in \mathcal{X}$, has three possible *outcomes*. Either the *home* or the *away* team wins, or it ends in a *draw*. The outcome of a match can be described using the triplet $r : \mathcal{X} \to \{0, 1\}^3$, a mapping from the set of all matches to a binary triplet, with the restriction that exactly one of the numbers equals 1 in each triplet. A *bookmaker*, $b \in \mathcal{B}$, may or may not offer odds for a specific match. When odds are offered, they are in the form of a *triplet*, with one number for each possible match outcome. This can be described by $o : \mathcal{X} \times \mathcal{B} \to \{\} \cup (1, \infty)^3$. We use decimal odds. For example, here are the match odds for a match between Chelsea and Manchester United: Chelsea (2.10), Draw (3.2), and Man Utd (3.75). Therefore, if we bet on Chelsea and they win, we get 2.1 times the amount we bet. On the other hand, the odds for betting on Manchester United are higher, which implies that Chelsea is a slight favorite. Let this example be a match x where the odds are offered by bookmaker b. We can write $o(x, b) = \{2.1, 3.2, 3.75\}$; or, if we break down the triplet into the home win, away win, and draw components, $o_H(x, b) = 2.1$, $o_A(x, b) = 3.75$, and $o_D(x, b) = 3.2$. These bookmaker odds can also be viewed as probabilistic forecasts of the match outcome. In simplified form, the odds 2.1 imply that the probability of Chelsea winning is $\frac{1}{2.1} = 0.48$. Similarly, the probability of Manchester United winning is 0.27 and the probability of a draw is 0.31. However, when we sum these probabilities, we get 1.06, which is more than 1.00. The extra 6% is the result of bookmakers lowering the odds in order to ensure a profit, which is known as the *bookmaker margin*. The bookmaker margin of bookmaker b for match x is: $\text{mrg}(x, b) = \frac{1}{o_H(x,b)} + \frac{1}{o_D(x,b)} + \frac{1}{o_A(x,b)} - 1$. We eliminate the margin by normalizing the forecasts so that the probabilities sum up to 1 (i.e., dividing the odds-implied probabilities by the margin). All of the results presented in this paper are for normalized probabilities.

2. Data description

The data used in this paper consist of 10,699 matches across several seasons between the years 2000 and 2006 from the following six major European soccer leagues: English Premier League (*E0*), English Championship (*E1*), Scottish Premiership League (*SC0*), Italian Serie A (*I1*), French Ligue 1 (*F1*), and Spanish La Liga (*SP1*). From this point