Betting market efficiency and prediction in binary choice models

Ruud H. Koning[1] · Renske Zijm[1]

Abstract

Implied winning probabilities are usually derived from betting odds by the normalization: inverse odds are divided by the booksum (sum of the inverse odds) to ensure that the implied probabilities add up to 1. Another, less frequently used method, is Shin's model, which endogenously accounts for a possible favourite-longshot bias. In this paper, we compare these two methods in two betting markets on soccer games. The method we use for the comparison is new and has two advantages. Unlike the binning method that is used predominantly, it is based on match-level data. The method allows for residual favourite-longshot bias, and also allows for incorporation of match specific variables that may determine the relation between the actual probability of the outcome and the implied winning probabilities. The method can be applied to any probabilistic classification problem. In our application, we find that Shin's model yields unbiased estimates for the actual probability of outcome in the English Premier League. In the Spanish La Liga, implied probabilities derived from the betting odds using either the method of normalization or Shin's model suffer from favourite bias: favourites tend to win their matches more frequently than the implied probabilities suggest.

Keywords Betting odds · Normalization · Shin's model · Logit model · Splines · Favourite-longshot bias

1 Introduction

One of the defining characteristics of a sports match is that the outcome is uncertain when the match is started. Betting odds offered by bookmakers are a good predictor of the probability of a certain outcome in a sport match (Stekler et al. 2010; Štrumbelj 2014). An advantage of this predictor is that it is easily available. Two possible alternatives that give an estimate of the probability of outcome are model-based predictions (see for instance Goddard and Asimakopoulos 2004; Goddard 2005; Reade et al. 2020) and predictions based on the aggregation of beliefs of many agents (as opposed to a single bookmaker). An example of the latter

✉ Ruud H. Koning
r.h.koning@rug.nl

Renske Zijm
Renske.Zijm@hotmail.com

[1] Department of Economics, Econometrics and Finance, Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands

approach would be predictions based on a betting exchange such as Betfair, where market participants can both offer bets and accept bets (see for instance Croxson and Reade 2013; Dobson and Goddard 2017). Another example of this approach to aggregate information based on many agents is to let the probability of outcome depend on the activity on social media (see for example Brown et al. 2018; Ramirez et al. 2021). In this paper, we restrict ourselves to the fixed odds betting market.

Betting odds are the payout for an outcome including the stake, where a high payout indicates that an outcome is unlikely to occur and a low payout indicates that an outcome is likely to occur. We take odds to be decimal (European) odds. However, bookmakers take a profit by offering betting odds that are too low to be fair, where betting odds are defined as fair if the expected profit equals zero. If odds were fair, the implied probability of the event would be equal to one over the odds. As a result of bookmakers taking profit, the sum of the inverse odds of all possible outcomes (called the booksum) is more than 1 and, therefore, the inverse odds cannot be directly interpreted as the probabilities of different match outcomes. As an example, assume a match with two possible outcomes, i.e., home win and away win, and the odds of a home win are 5 and the odds of an away win are 1.16. The odds are translated into implied winning probabilities of 0.20 (= 1/5) and 0.86 (= 1/1.16), respectively, and the booksum is 1.06, which is more than 1. To remove this excess probability, called overround (here, 0.06), the odds require an adjustment to obtain applicable winning probabilities that add up to one.

The most common adjustment is basic normalization, where the inverse odds are divided by their sum (Štrumbelj 2014). However, scaled probabilities do not lead to unbiased estimates of the true winning probabilities (Deschamps and Gergaud 2007; Koning and Boot 2020). The biased estimators of the winning probabilities tend to be too low for favourites and too high for underdogs, which is called the favourite-longshot bias. The concept of favourite-longshot bias was first documented by Griffith (1949) for horse racing and the existence of this bias has been found in the betting odds of several other sports including soccer (Cain et al. 2000). Under-estimation of high probability events and over-estimation of low probability evens has also been documented in non-sports contexts (see for example Kip Viscusi 1998).

In the literature, most papers use the method of basic normalization to derive implied probabilities from published odds. Basic normalization, though, is not the only way to convert observed odds into implied winning probabilities. Shin (1993) developed a model to account endogenously for the favourite-longshot bias, based on the assumed existence of insider traders who have superior knowledge to bettors and bookmakers. In his model, which is developed in the context of horse racing, he assumes that insiders know the identity of the winning horse before the race starts. His conversion of odds into implied winning probabilities is more complex than basic normalization, but it seems to provide better implied probabilities than the method of basic normalization (see for instance Clarke et al. 2017; Koning and Boot 2020). Štrumbelj (2014) is one of the very few papers that applies Shin's model to soccer matches, and he shows that implied probabilities derived from Shin's model are a better predictor of outcomes than implied probabilities derived from basic normalization.

To compare implied probabilities derived from basic normalization and Shin's model, we need to relate these implied probabilities to the actual outcomes. This is, of course, a much more general problem. Basically, it is the question how to assess the fit of a probabilistic classifier. We tackle this problem with a novel approach, that allows us to use match-level data. As our application is betting on sports, we need to allow for a possible favourite-longshot bias by estimating a flexible functional form relation between the outcome (dependent variable) and the implied probability (independent variable). We do so using restricted cubic splines. This method to assess the quality of a classifier can be applied elsewhere: it can be used to

assess whether an estimated probability is an unbiased estimator for the probability of the outcome in other settings as well. Hence, our paper contains two contributions: it compares basic normalization and Shin's probabilities as estimates of the probability of the outcome of a soccer match, and it does so using a simple and general approach that allows for local deviations of actual probability of outcomes from the estimated probabilities that are derived from the betting odds. In particular, this approach allows for the favourite-longshot bias.

In some cases, implied probabilities (or measures derived from those) are used as covariates to model, for example, match attendance or television demand. It is important that such probabilities are good estimates of the actual probability of outcome. Moreover, the analysis of the informational efficiency of the sports betting market has a broader relevance. If it can be shown that the sports betting market is efficient, then it is more likely that similar markets also process information efficiently. Thaler and Ziemba (1988) conclude that betting markets are more relevant for analyzing market efficiency and rationality than stock markets, since each bet includes a well-defined point in time at which its value becomes certain, in contrast to financial assets.

The paper begins with a review of the literature on the favourite-longshot bias. Then, the data and methodology are described, with special focus on measuring classification. In Sect. 4 the results are presented and discussed. The paper ends with a conclusion.

2 Favourite-longshot bias

The most widely documented inefficiency within sport betting markets is the favourite-longshot bias (Deschamps and Gergaud 2007). The favourite-longshot bias implies that underdogs are overvalued and favourites are undervalued, indicating that underdogs lose more often and favourites win more frequently than the bookmakers' probabilities suggest. In other words, bettors overbet underdogs more than expected given their low winning frequency, while favourites are underbet given how often they win. Hence, bettors who bet systematically on underdogs receive lower returns than bettors who bet on favourites.

The existence of the favourite-longshot bias is explained in several ways. Shin (1993) explains this bias in a theoretical model that assumes that bookmakers face a percentage of (hypothetical) insiders with perfect knowledge of the outcome. The information asymmetry exposes the bookmaker to large losses because of the high payouts and, therefore, the bookmaker reduces this risk by decreasing the odds on the underdogs and thereby increasing the odds on the favourites. Alternatively, Quandt (1986) explains that this bias arises since bettors are risk-loving. In other words, bettors are willing to give up some expected return in exchange for the additional risk. The third explanation refers to the assumption that the favourite-longshot bias is caused by overconfidence instead of risk-loving behaviour, which leads to individuals misinterpreting probabilities because they underestimate the error variance (Golec and Tamarkin 1995; Snowberg and Wolfers 2010; Woodland and Woodland 1994). Additionally, Franke (2020) suggests that bettors bias odds due to misperception of the probabilities independently of the number of possible outcomes.

Most early studies that analyze the efficiency of the fixed odds betting sports market focused on horse racing, although more recent studies show that the favourite-longshot bias found in horse racing occurs within several other gambling sports markets, including tennis and soccer (Abinzano et al. 2016; Clarke et al. 2017; Koning and Boot 2020). Deschamps and Gergaud (2007) explore the favourite-longshot bias in English soccer data and show that this bias depends on the odds status. For the odds status home win and away win, the authors

find a clear favourite-longshot bias, however, there is a reversed favourite-longshot bias for draw odds. Cain et al. (2000) also indicate the presence of favourite-longshot bias in betting odds of soccer results. Cain et al. (2000) note that the existing literature is unclear whether the bias only exists in the low and in the high probability cases and how these extreme cases extend. Relevant recent contributions to this literature that concern a similar betting market to ours are Angelini and De Angelis (2019) and Elaad et al. (2020). In both cases, implied probabilities are derived using the method of basic normalization. Angelini and De Angelis (2019) consider many more leagues than we do, and they find different degrees of efficiency among markets. They find evidence of a significant favourite-longshot bias in three out of eleven markets considered. Neither the English Premier League or the Spanish La Liga exhibit a favourite-longshot bias in their analysis. Elaad et al. (2020) document odds to be unbiased in general, both in terms of the favourite-longshot bias or outcome type.

Angelini et al. (2022) analyse the efficiency of a betting exchange, which is a market where punters both offer and accept bets. There is no bookmaker and negligible overround. Punters can trade in bets both before the match and during the match. These betting markets show large turnover, suggesting that they are liquid and that new information is processed quickly. They find a material reverse favourite-longshot bias both when considering before match odds and within match odds. The latter differ from before match odds because punters trade bets during the match and the odds reflect information available at the moment of trading. In general, one would expect that liquid betting markets where punters provide both demand for and supply of bets would be more efficient than the fixed-odds betting market in our paper, where the bookmaker is the only supplier of bets. The bookmaker charges an overround, so any inefficiency may be hard to exploit by punters.

To obtain an effective betting strategy, the winning probabilities should account for the favourite-longshot bias. Clarke et al. (2017) compare the four most popular methods to transform betting odds for tennis, horse racing, and greyhound racing and conclude that the Shin (1993) model is a more accurate approach than basic normalization. Additionally, Štrumbelj (2014) also suggests that probabilities estimated using the Shin (1993) model are, on average, better than probabilities based on basic normalization. Furthermore, Shin's model provides unbiased estimates of the winning probability, while basic normalization does not account for the favourite-longshot bias (Koning and Boot 2020). According to Cain et al. (2003), there is no direct effect of the presence of insiders when using Shin's model within soccer. If the bookmakers create a bias in their odds because of the presence of insiders, this would suggest that there is no favourite-longshot bias when the Shin (1993) model is used. However, Cain et al. (2003) suggest that further research for soccer is required due to the possible indirect effect of the presence of insiders and the possibility of more than two outcomes. Only a few studies use Shin's model for transforming the betting odds in winning probabilities of soccer games and much more empirical research is required.

3 Methodology

3.1 Pricing methods

Similar to previous research, basic normalization and Shin's model are used to transform betting odds into winning probabilities (Clarke et al. 2017; Koning and Boot 2020; Štrumbelj 2014). In this paper, the efficiency of basic normalization and Shin's method is compared

by analyzing whether these implied winning probabilities are unbiased indeed for the actual probability of outcomes.

Basic normalization A simple way to obtain winning probabilities that add up to 1 is to divide the implied probabilities by their sum. Suppose $\boldsymbol{o} = (o_1, o_2, \ldots, o_k)$ are the offered (European) decimal betting odds for a match with $k \geq 2$ outcomes and $o_l > 1$ for all $l = 1, 2, \ldots, k$. The inverse odds $\boldsymbol{b} = (b_1, b_2, \ldots, b_k)$ can be obtained by

$$b_l = \frac{1}{o_l}. \tag{1}$$

The sum of these inverse odds over all possible outcomes exceeds 1. To standardize the inverse odds, they are divided by the booksum B, where

$$B = \sum_{l=1}^{k} b_l. \tag{2}$$

The excess probability of the booksum $\lambda \equiv B - 1$ is called the overround. The standardized implied probabilities, given by

$$p_l = \frac{b_l}{B}, \tag{3}$$

sum up to 1 and can be interpreted as applicable probabilities. The winning probabilities determined from betting odds using basic normalization are referred to as scaled probabilities. This method is also known as (simple) standardization.

Shin's model An alternative to basic normalization is Shin's model that is based on the following theoretical model. Suppose one assumes a hypothetical group of insider traders with perfect knowledge of the outcome. These traders with superior information are called 'insiders' and significantly influence the outcome of the betting market (Crafts 1985; Shin 1993). Nowadays, more advanced approaches still suggest the existence of insider trading in betting markets and other financial markets, see, for example Coleman (2007), Schnytzer et al. (2012), and Deng et al. (2019). In the model, bookmakers want to limit their exposure to insiders, especially in the case of low probability-high payout events. They do so by reducing the odds offered for such events. In Shin's model, bookmakers set odds to maximize profit, knowing in advance that they have to pay the insiders. The insiders are a fraction z of the population, which is the measure of the incidence of insider trading. Then, z equals 0 in the absence of insiders, while $z > 0$ indicates a deviation of prices from the true probabilities due to insider trading. Shin (1993) and Clarke et al. (2017) show that the implied fraction of insiders and the corresponding implied probabilities can be calculated from the inverse odds b_l and the booksum B. The implied Shin probabilities of outcome (that sum to 1) and the implied fraction of insiders are obtained by solving:

$$p_l = \frac{\sqrt{z^2 + 4(1-z)\frac{b_l^2}{B}} - z}{2(1-z)}, \tag{4}$$

and

$$z = \frac{\sum_{l=1}^{k} \sqrt{z^2 + 4(1-z)\frac{b_l^2}{B}} - 2}{k - 2}. \tag{5}$$

Hence, given the profit maximizing inverse odds b_l, Eqs. (4) and (5) can be used to calculate Shin probabilities.

3.2 Test strategies

To test whether standardized probabilities or Shin probabilities provide a better estimate for the actual probability of the outcome, we extend the test strategy in Koning and Boot (2020). For simplicity, we take a home win in a soccer match as the leading example. The actual probability of the outcome in match i is denoted by $\Pr(H_i)$. The implied probability p_{H_i} of a home win in match i can be calculated through basic normalization (Eq. (3)), or through Shin's model (Eqs. (4) and (5)). In both cases, the implied probabilities depend on the stated decimal betting odds. For simplicity of notation, we drop the index i, and from the context it will be clear if the implied probability is obtained through normalization or through Shin's model, and we denote this implied probability by p_H. Our test strategy can be used to test whether any probability would be an unbiased predictor of $\Pr(H)$. This predictor could also be derived from an extensive regression model (see for example, Goddard and Asimakopoulos 2004; Goddard 2005), or from subjective beliefs of a particular fan.

If p_H would be an unbiased predictor for $\Pr(H)$, we would have

$$\Pr(H) = p_H, \tag{6}$$

and consequently, we rewrite this as

$$\Pr(H) = p_H = \frac{1}{\frac{1}{p_H}} = \frac{1}{1 + \frac{1}{p_H} - 1} = \frac{1}{1 + \exp\left(\log\left(\frac{1}{p_H} - 1\right)\right)}. \tag{7}$$

This suggests a very simple test of the validity of Eq. (6): estimate the logistic regression model

$$\Pr(H) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 \log\left(\frac{1}{p_H} - 1\right)\right)\right)}. \tag{8}$$

and test the joint null hypothesis $\beta_0 = 0$ and $\beta_1 = -1$.

Usual tests for calibration of binary outcome models are based on grouping the data (based on p_H) in a number of buckets. In each bucket, the average implied probability (the average p_H) is compared to the actual frequency of the event occurring. Some examples of this approach are Blochwitz et al. (2006) in credit risk default modeling, Štrumbelj (2014) in the context of analyzing Shin probabilities, Kuypers (2000) in an analysis of fixed odds betting markets for football in England, and Guo et al. (2017) in exploring classification decisions in health and medicine. This binning approach has two major disadvantages: it does not allow for observation specific covariates that may cause Eq. (6) to be invalid, and it reduces the effective number of observations basically to the number of bins. A potential favourite-longshot bias would show up as a poor fit in bins that correspond to high and low values of p_H.Considering the literature on the existence of the favourite-longshot bias, we would like to allow for deviations from Eq. (6) for low probability events and high probability events.

We take model (8) as our point of departure, and extend it to address these two disadvantages. Instead of estimating Eq. (8) directly, we estimate the relation between the probability of the outcome and $\log(\frac{1}{p_H} - 1)$ in a flexible way. In the context of informational efficiency of betting markets, we need to test whether the implied probability is an unbiased estimator for the actual probability of outcome. Also, other covariates should not determine the probability of the outcome. In other words, the coefficients of any additional covariates in Eq. (8) should be 0. Adding covariates will improve the power of the test (see Sauer et al. 1988).

The model we estimate is

$$\Pr(H) = \frac{1}{1 + \exp\left(-g\left(\log\left(\frac{1}{p_H} - 1\right)\right) - \gamma' w\right)}. \tag{9}$$

In this equation, the vector w contains other variables, for example the day when the match is being played, measures of form of both teams, etc. The function $g(\cdot)$ is a flexible function. We take it to be a restricted cubic spline (see for example Harrell 2015). These splines are characterized by a number of knots, where the curvature of the relation may change. The restricted cubic splines are defined by (1) being cubic splines (piecewise third-order polynomials) between the knots in the range of the independent variable, (2) being a linear function beyond the boundary knots and (3) being continuous and forcing the first and second derivative of the function to agree at the knots. We assume p knots, and use the shorthand notation $x \equiv \log(\frac{1}{p_H} - 1)$. The knots (in terms of x) are at $t_1, t_2, \ldots t_p$. The restricted cubic spline can then be represented as

$$g(x) = \beta_0 + \beta_1 x + \beta_2 (x - t_1)_+^3 + \beta_3 (x - t_2)_+^3 + \cdots + \beta_{p+1} (x - t_p)_+^3, \tag{10}$$

with $(x - t_j)_+^3 = \max(0, (x - t_j)^3)$. This representation is overparametrized: if we impose differentiability at each knot and linearity for $x > t_p$, two linear restrictions on the parameters are obtained. The model that is actually estimated is based on a linear transformation of the matrix with all covariates in Eq. (10) and a linear transformation of the parameters. With a slight abuse of notation, we only present the estimation results of the transformed model, as that model is exactly identified.

Summarizing, we propose to estimate the logit model

$$\Pr(H) = \frac{1}{1 + \exp(-g(x) - \gamma' w)}. \tag{11}$$

By choosing enough knots $t_1, \ldots, t_p$, we can detect a favorite-longshot bias. Moreover, any information not incorporated in the standardized probabilities or Shin probabilities can be added as the vector with additional covariates w. Then, a test whether these probabilities incorporate all information amounts to testing $\gamma = 0$. Additionally, the implied probabilities are unbiased estimates for the probability of the outcome, if $\beta_1 = -1$, $\beta_0 = \beta_2 = \ldots = \beta_{p+1} = 0$, and $\gamma = 0$.

Related to our approach is the one in Angelini and De Angelis (2019), Elaad et al. (2020), and Angelini et al. (2022). These papers follow the Mincer–Zarnowitz (Mincer and Zarnowitz 1969) regression based forecast evaluation framework. Using our notation, they estimate the following model

$$I_H - b_H = \beta_0 + \beta_1 b_H + \gamma' w + \eta, \tag{12}$$

with I_H an indicator taking value 1 if the match ends in a home win, b_H the inverse decimal odds of a home win and 0 otherwise, and w other covariates. If markets are efficient, we should have $\mathcal{E}(I_H - b_H) = \beta_0$, over all matches. In this approach, no explicit conversion to implied winning probabilities is proposed, so we would expect $\beta_0 < 0$, reflecting the overround of the bookmakers. The model is estimated using weighted least squares, reflecting heteroscedasticity of the dependent variable.

Our approach differs in some respects: we compare winning probabilities derived using two different models (basic normalization and Shin's model), and we allow deviations from the efficient market hypothesis to be local, that is to depend on the level of the implied winning probability. For example, our approach should be able to distinguish between a favourite bias

and a favourite-longshot bias separately. Also, our estimation approach (a logit model) reflects the discrete outcome of a match. This, and the fact that we allow for a nonlinear relationship between the outcome and implied probability, should give our test higher power to the one based on the linear probabilty model in Eq. (12).

Note that our methodology to assess the precision of predictions is very general. The leading case in this paper are predictions based on fixed odds betting market, but such predictions could also have been derived from the two other methods discussed in the Introduction: p_H could have been derived from some kind of regression model, or from aggregating information of many agents.

4 Empirical results

We illustrate our approach to testing informational efficiency of betting odds using data from football-data.co.uk. We use published odds by a major bookmaker (bet365) on home wins in the English Premier League and Spain's La Liga, from the season 2002/2003 onwards. Both leagues represent the highest level of professional soccer in these countries. Both samples of odds were offered by this bookmaker in the UK (and not in Spain in the case of La Liga). During most of the time period considered there was no open European market market for offering betting opportunities on sports matches. Even though Spanish bettors could have arranged access to these betting opportunities, such access would require technical proficiency and knowledge of English language. For them, local betting alternatives will be easier to access. Press coverage of the English Premier League is much better in English media than coverage of La Liga. We expect English bettors to be better informed about the English Premier League than about La Liga. The datasets for the Premier League and La Liga contain 6840 observations each (18 seasons (2002/2003 to 2019/2020), with 380 matches per season). Implied home win probabilities according to either basic normalization (Eq. (3)) or Shin's model (Eqs. (4) and (5)) are calculated for each match using the R-library `implied` (Lindstrøm 2020). All calculations are done in R (R Core Team 2020).

We start with an analysis of the English Premier League. First, we estimate the logit model without nonlinear terms (that is, the only covariate is $\log(\frac{1}{p_H} - 1)$). Results for both models to derive implied home win probabilities from betting odds are presented in Table 1. For both models, the estimated slope coefficient $\hat{\beta}_1$ is not significantly different from -1. However, the intercept of the model for the scaled probabilities is different from 0 ($p = 0.015$), which is not the case for the implied probabilities according to the Shin model. The last line of Table 1 gives the Wald statistics for the joint null hypothesis $\beta_0 = 0$ and $\beta_1 = -1$, and the p value corresponding to that statistic. Because of the size of the dataset, we take a 0.01 significance level. We conclude that in both cases the hypothesis that betting odds implied probabilities are unbiased for the probability of outcome cannot be rejected.

In a second step, we extend the basic logit model with nonlinear terms, so that a potential favourite-longshot bias can be accommodated. In the end, deviations from $\Pr(H) = p_H$ may depend on the level of p_H, and if Eq. (8) is misspecified, the test may falsely conclude that implied probabilities are unbiased estimates for the probability of the outcome. For this reason, we proceed with the semi-parametric approach by adding restricted cubic spline terms to the basic specification. We choose knots to be at the 0.05, 0.275, 0.5, 0.725, and 0.95 quantiles of the observed $\log(\frac{1}{p_H} - 1)$-distribution. This choice of quantiles differs from the one suggested in Harrell (2015), since we have moved the first and last knot a bit more to the beginning and end of the distribution. Since the restricted cubic splines regression

Table 1 Estimation results, basic logit model (English Premier League)

		Scaled prob.		Shin prob.	
		Estimate	St.error	Estimate	St.error
β_0		0.068	0.028	0.038	0.028
β_1		−1.021	0.037	−0.975	0.035
		Scaled prob.		Shin prob.	
		Test statistic	p value	Test statistic	p value
W		5.982	0.050	3.036	0.219

Table 2 Estimation results, model with restricted cubic spline (English Premier League)

	Scaled prob.		Shin prob.	
	Estimate	St.error	Estimate	St.error
β_0	0.292	0.167	0.278	0.170
β_1	− 0.839	0.177	− 0.792	0.167
β_2	− 1.144	0.778	− 1.124	0.733
β_3	8.876	6.350	8.596	6.009
β_4	−13.581	11.605	−13.094	11.128
	Scaled prob.		Shin prob.	
	Test statistic	p value	Test statistic	p value
W (nonlinear terms 0)	4.201	0.241	4.533	0.209
W (overall)	10.137	0.071	7.677	0.175

assumes a linear function beyond the boundary knots, a favourite-longshot bias can be better accounted for using these fixed quantiles. In Table 2, we present the results of the extended logit model. The magnitudes of the estimated coefficients are difficult to interpret, except for $\hat{\beta}_0$ and $\hat{\beta}_1$, those are comparable to the estimates in Table 1. More interesting are the last two lines of Table 2. The first of those gives the Wald statistics corresponding to the null hypothesis that all coefficients of the nonlinear terms are jointly 0. This hypothesis cannot be rejected in both cases, judging from the p values that are reported (0.241 and 0.209, respectively). Consequently, adding nonlinear terms has not improved the fit of the model. The very last line gives the Wald statistics corresponding to $\beta_0 = 0$, $\beta_1 = -1$, and the coefficients of all nonlinear terms are also 0. Also in this case, the null hypotheses cannot be rejected in both models: there is no evidence of a favourite-longshot bias in the market for home wins of the English Premier League.

The first and second step models are shown graphically in Fig. 1. We graph the relation between the scaled and Shin probabilities of outcome (horizontal) and the probability of outcome (vertical). In each panel, a 45-degree line is drawn which reflects the hypothesis to be tested ($\beta_0 = 0$ and $\beta_1 = -1$). The red line in the left two panels is the estimated logit model from Table 1, and in the right two panels the models with the nonlinear terms are drawn. The knots are indicated by the dashed vertical lines. Even though the red line in the right two panels suggest some nonlinearity, this nonlinearity is statistically not significant.

In a final step, we add other covariates to Eq. (8). Since nonlinearity is not significant, we do not incorporate the restricted cubic splines terms. Even though one can think of many different
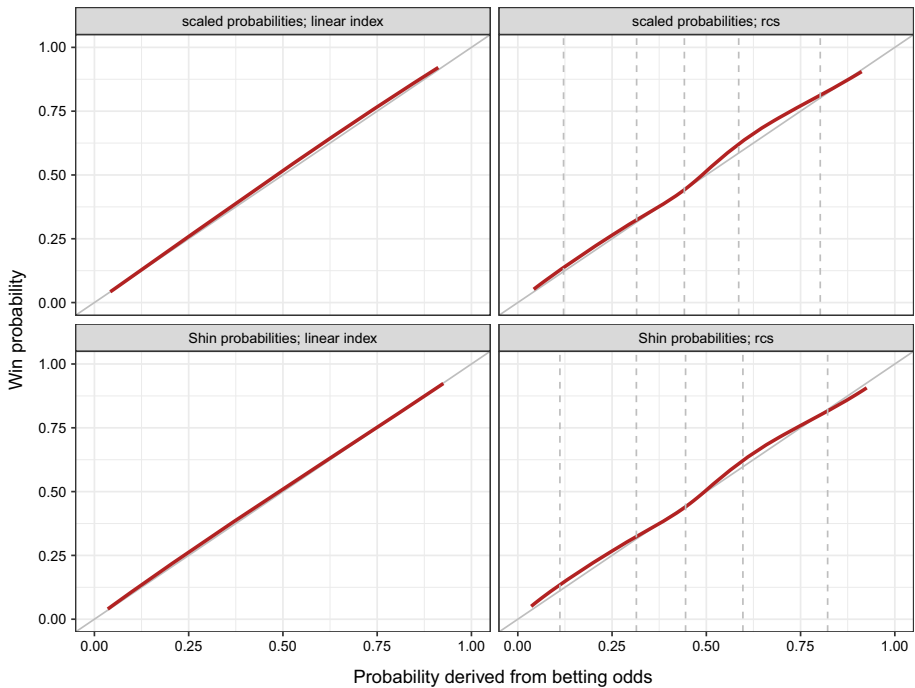
Fig. 1 Fit of the probabilities implied by betting odds of the English Premier League

covariates that are not priced into the betting odds, we consider only two, as an example of the method. The first covariate is the moving average of the number of points obtained in the last three matches by the home team (coefficient γ_1). This measure will vary between 0 (last three matches are lost) and 3 (last three matches have been won). The second covariate is based on the hypothesis that teams perform differently on weekdays from weekends (Krumer and Lechner 2018), so we include a dummy variable for whether or not the the match is played in the weekend (Friday, Saturday, or Sunday) (coefficient γ_2). Results of this specification are given in Table 3. Again, the Wald tests are most interesting, and we find that the additional covariates are not jointly significant. The overall test ($\beta_0 = 0$, $\beta_1 = -1$, $\gamma_1 = \gamma_2 = 0$), however, indicates there is some evidence that scaled probabilities are biased estimators for the probability of outcome. This is our best assessment of the unbiasedness of standardized and Shin probabilities, as the last test allows both for a potential favourite-longshot bias, and influence of other covariates. There is some evidence that implied probabilities derived according to Shin's model are unbiased, while the ones derived from standardisation are biased.

In a second example, we have applied the same approach to betting data for matches of the Spanish La Liga. The results of the logit model without linear terms are shown in Table 4. The intercept and the estimated slope coefficient for the scaled probabilities are significantly different from 0 and -1, respectively. We cannot reject the null hypothesis ($\beta_0 = 0$, $\beta_1 = -1$) for Shin probabilities at a 0.01 significance level. Expanding the model by adding restricted cubic spline terms to the basic specification leads to the results given in Table 5. The hypothesis that all nonlinear terms are jointly 0 is rejected for both normalization and Shin's model ($p < 0.01$ for both models), implying that adding the nonlinear terms has

Table 3 Estimation results, model with additional covariates (English Premier League)

	Scaled prob.		Shin prob.	
	Estimate	St.error	Estimate	St.error
β_0	0.122	0.099	0.096	0.099
β_1	−1.041	0.045	−0.995	0.043
γ_1	−0.062	0.052	−0.064	0.052
γ_2	0.084	0.073	0.083	0.073
	Scaled prob.		Shin prob.	
	Test statistic	p value	Test statistic	p value
$W\ (\gamma = 0)$	2.799	0.247	2.899	0.235
W (overall)	14.073	0.007	10.336	0.035

Table 4 Estimation results, basic logit model (Spanish La Liga)

	Scaled prob.		Shin prob.	
	Estimate	St.error	Estimate	St.error
β_0	0.115	0.029	0.074	0.029
β_1	−1.094	0.044	−1.028	0.041
	Scaled prob.		Shin prob.	
	Test statistic	p value	Test statistic	p value
W	16.945	0.000	6.544	0.038

improved the fit of the model. Furthermore, the overall hypothesis testing whether $\beta_0 = 0$, $\beta_1 = -1$, and the coefficients of all nonlinear terms are equal to 0, is also rejected for both methods. This shows that there is evidence that basic normalization and Shin's model provide biased estimates of the probability of the outcome. The models are graphically shown in Fig. 2, and it is clear that both methods suffer from a favourite bias since home wins are actually more frequent than implied by the derived probabilities. For longshots (unlikely winners), the estimated models are close to the 45-degree line, so there is no evidence of a longshot bias. Even though Shin's model allows endogenously for a favourite-longshot bias in published betting odds, it seems that a residual favourite bias remains in the implied probabilities.

In the analysis of this section we use data from eighteen seasons, and it is possible that the relation between implied probabilities and outcomes has changed over time. This could be due to changes in regulation, the profile of internet bettors, or the advent of betting exchanges. In this paper we want to give an overal picture which model for the calculation of the implied probabilities provides the best fit (either standardization or Shin's model). The statistical tests of this section can be easily extended to incorporate heterogeneity over time by including time effects in the vector of covariates w.

5 Conclusion

In this paper we have developed a general approach to assess the quality of probabilistic classification. Using a simple model-based approach, we have shown how it is possible to

Table 5 Estimation results, model with restricted cubic spline (Spanish La Liga)

	Scaled prob.		Shin prob.	
	Estimate	St.error	Estimate	St.error
β_0	− 0.460	0.186	− 0.491	0.194
β_1	− 1.802	0.216	− 1.650	0.201
β_2	2.307	0.799	1.980	0.725
β_3	−17.567	7.971	−16.556	7.870
β_4	29.029	15.827	27.524	15.309

	Scaled prob.		Shin prob.	
	Test statistic	p value	Test statistic	p value
W (nonlinear terms 0)	12.992	0.005	12.047	0.007
W (overall)	27.020	0.000	17.713	0.003
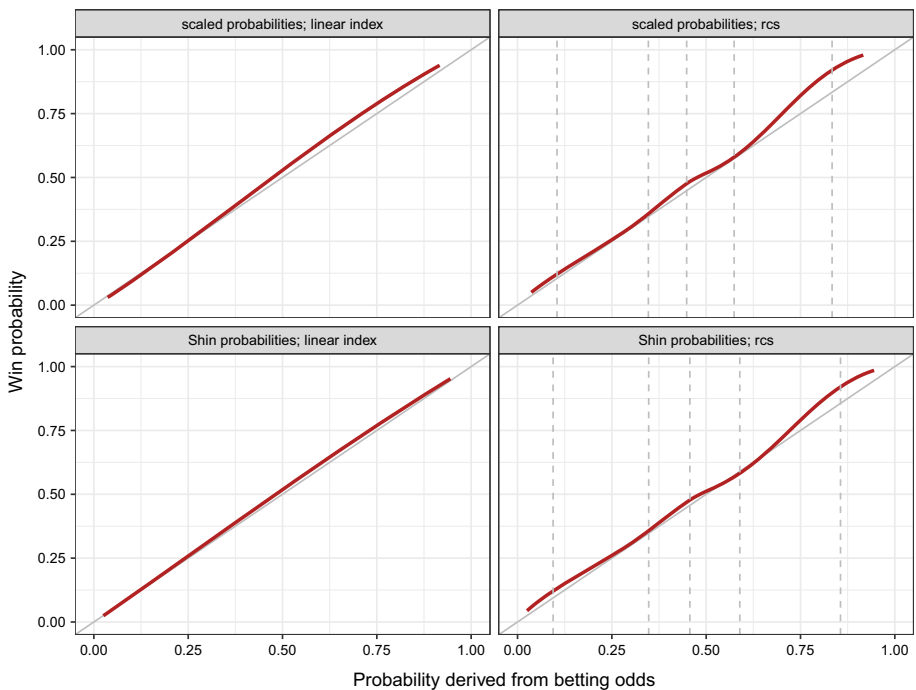
Fig. 2 Fit of the probabilities implied by betting odds of the Spanish La Liga

assess whether or not implied probabilities are unbiased for the probability of outcome. Our approach allows for nonlinearities in a relationship that may, for example, capture a favourite-longshot bias in the case of betting markets. This novel method allows us to use individual level data which, unlike binned data, allow for local deviations. Also, in our approach it is easy to assess whether the relation between implied probabilities and probability of outcome is moderated by other covariates.

The approach has been illustrated in a simple example, where two models to derive probabilities from betting odds are compared: basic normalization and Shin's model. We applied the methods to betting data from both the English Premier League and the Spanish La Liga. For the English Premier League, we have shown that Shin's model, unlike the case of basic standardization, yields unbiased implied probabilities. Using Shin's model, we did not find any evidence of a favourite-longshot bias in the English Premier League, nor that the information in current form or day of the week is priced incorrectly in published betting odds. The conclusion is different for the Spanish league: implied probabilities suffer from a favourite bias, also in the case of Shin's model. Consequently, informational efficiency of betting odds varies between countries. This may be due to lack of access to the English betting market or the availability of local alternatives.

References

Abinzano, I., Muga, L., & Santamaria, R. (2016). Game, set and match: The favourite-long shot bias in tennis betting exchanges. *Applied Economics Letters, 23*(8), 605–608.

Angelini, G., & De Angelis, L. (2019). Efficiency of online football betting markets. *International Journal of Forecasting, 35,* 712–721.

Angelini, G., De Angelis, L., & Singleton, C. (2022). Informational efficiency and behaviour within in-play prediction markets. *International Journal of Forecasting, 38*(1), 282–299.

Blochwitz, S., Martin, M. R. W., & Wehn, C. S. (2006). Statistical approaches to PD validation. In B. Engelmann & R. Rauhmeier (Eds.), *The basel II risk parameters, chapter XIII* (pp. 206–289). Springer.

Brown, A., Rambaccussing, D., Reade, J. J., & Rossi, G. (2018). Forecasting with social media: Evidence from tweets on soccer matches. *Economic Inquiry, 56*(3), 1748–1763.

Cain, M., Law, D., & Peel, D. (2000). The favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy, 47*(1), 25–36.

Cain, M., Law, D., & Peel, D. (2003). The favourite-longshot bias, bookmaker margins and insider trading in a variety of betting markets. *Bulletin of Economic Research, 55*(3), 263–273.

Clarke, S., Kovalchik, S., & Ingram, M. (2017). Adjusting bookmaker's odds to allow for overround. *American Journal of Sports Science, 5*(6), 45–49.

Coleman, L. (2007). An evaluation using thoroughbred wagering markets. *The Journal of Gambling Business and Economics, 1*(1), 31–55.

Crafts, N. F. R. (1985). Some evidence of insider knowledge in horse race betting in Britain. *Economica, 52*(207), 295–304.

Croxson, K., & Reade, J. J. (2013). Information and efficiency: Goal arrival in soccer betting. *The Economic Journal, 124,* 62–91.

Deng, S., Wang, C., Li, J., Yu, H., Tian, H., Zhang, Y., et al. (2019). Identification of insider trading using extreme gradient boosting and multi-objective optimization. *Information, 10*(12), 367.

Deschamps, B., & Gergaud, O. (2007). Efficiency in betting markets: Evidence from English football. *The Journal of Prediction Markets, 1*(1), 61–73.

Dobson, S., & Goddard, J. (2017). Evaluating probabilities for a football in-play betting market. In P. Rodríguez, B. R. Humphreys, & R. Simmons (Eds.), *The economics of sports betting, chapter 4* (pp. 52–70). Edward Elgar Publishing.

Elaad, G., Reade, J. J., & Singleton, C. (2020). Information, prices and efficiency in an online betting market. *Finance Research Letters, 35,* 101291. https://doi.org/10.1016/j.frl.2019.09.006.

Franke, M. (2020). Do market participants misprice lottery-type assets? Evidence from the European soccer betting market. *The Quarterly Review of Economics and Finance, 75,* 1–18.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting, 21,* 331–340.

Goddard, J., & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting, 23,* 51–66.

Golec, J., & Tamarkin, M. (1995). Do bettors prefer long shots because they are risk-lovers, or are they just overconfident? *Journal of Risk and Uncertainty, 11*(1), 51–64.

Griffith, R. M. (1949). Odds adjustments by American horse-race bettors. *The American Journal of Psychology, 62*(2), 290–294.

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In D. Precup, & Y. W. Teh (Eds.), *Proceedings of the 34th international conference on machine learning, volume 70 of Proceedings of machine learning research* (pp. 1321–1330). PMLR.

Harrell, F. E., Jr. (2015). *Regression Modelling Strategies* (2nd ed.). Springer.

Kip Viscusi, W. (1998). *Rational Risk Policy*. Clarendon Press.

Koning, R. H., & Boot, T. (2020). Informational content of tennis betting odds. In C. Ley & Y. Dominicy (Eds.), *Science meets sports: When statistics are more than numbers*. Cambridge Scholars Publishing.

Krumer, A., & Lechner, M. (2018). Midweek effect on soccer performance: Evidence from the German Bundesliga. *Economic Inquiry, 56*(1), 193–207.

Kuypers, T. (2000). Information and efficiency: An empirical study of a fixed odds betting market. *Applied Economics, 32,* 1353–1363.

Lindstrøm, J. C. (2020). *Implied: Convert bookmaker odds to probabilities.* https://CRAN.R-project.org/package=implied.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In J. A. Mincer (Ed.), *Economic forecasts and expectations: Analysis of forecasting behavior and performance, chapter 1* (pp. 3–46). NBER.

Quandt, R. E. (1986). Betting and equilibrium. *The Quarterly Journal of Economics, 101*(1), 201–207.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ramirez, P., Reade, J. J., & Singleton, C. (2021). Betting on a buzz, mispricing and inefficiency in online sportsbooks. Discussion Paper 2021-10, Department of Economics, University of Reading.

Reade, J. J., Singleton, C., & Brown, A. (2020). Evaluating strange forecasts: The curious case of football match scorelines. *Scottish Journal of Political Economy, 68,* 261–285.

Sauer, R. D., Brajer, V., Ferris, S. P., & Marr, W. M. (1988). Hold your bets: Another look at the efficiency of the gambling market for National Football League games. *Journal of Political Economy, 96*(1), 206–213.

Schnytzer, A., Makropolou, V., & Lamers, M. (2012). Pricing decisions and insider trading in horse betting markets. Working Paper 772, Universiteit Gent, Faculteit Economie en Bedrijfskunde.

Shin, H. S. (1993). Measuring the incidence of insider trading in a market for state-contingent claims. *The Economic Journal, 103*(420), 1141–1153.

Snowberg, E., & Wolfers, J. (2010). Explaining the favorite-long shot bias: Is it risk-love or misperceptions? *Journal of Political Economy, 118*(4), 723–746.

Stekler, H. O., Sendor, D., & Verlander, R. (2010). Issues in sports forecasting. *International Journal of Forecasting, 26*(3), 606–621.

Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting, 30*(4), 934–943.

Thaler, R. H., & Ziemba, W. T. (1988). Anomalies: Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives, 2*(2), 161–174.

Woodland, L. M., & Woodland, B. M. (1994). Market efficiency and the favorite-longshot bias: The baseball betting market. *The Journal of Finance, 49*(1), 269–279.