

A STATISTICAL DEVELOPMENT OF FIXED ODDS BETTING RULES IN SOCCER

*Ian Milliner, Department
of Mathematics and
Statistics, University of the
West of England, Bristol,
UK
ian.milliner@tiscali.co.uk*

*Paul White, Department
of Mathematics and
Statistics, University of
the West of England,
Bristol, UK
paul.white@uwe.ac.uk*

*Don J. Webber,
School of Business
Economics, Auckland
University of
Technology, Auckland,
New Zealand
don.webber@aut.ac.nz*

I. ABSTRACT

Two simple but seemingly profitable betting rules for betting on the away win in association football are developed. One rule is consistent with avoiding those games in which there is a clear favourite. The second rule is based directly on modelling bookmaker odds and assessing the residuals under the fitted model. Contrary to previous research the betting rule using the residuals suggests avoiding betting on those games where there are large discrepancies between bookmaker odds and predicted-model odds.

Keywords: Fixed odds betting rules; away win; bookmakers' probabilities

JEL Classification: D81

* Corresponding author

II. A STATISTICAL DEVELOPMENT OF FIXED ODDS BETTING RULES IN SOCCER

Betting on the outcome of UK association football matches (soccer matches) is big business with in excess of 3% of the UK adult population regularly placing bets on the football fixed odds market (Department for Culture, Media and Sport, 2007). It has also been the subject of recent academic research (see, for example, Archontakis and Osborne, 2007). In the football fixed odds market bookmakers offer betting odds for each of the three mutually exclusive and exhaustive outcomes which are for the home team to win or the away team to win or for the game to end in a draw. In the UK, it is customary practice to quote odds in the form ‘ $a - \text{to} - b$ ’ for the home win, ‘ $c - \text{to} - d$ ’ for the away win, and ‘ $e - \text{to} - f$ ’ for the draw, where a, b, c, d, e and f are integers. Thus, for instance, if a bettor wagers b pounds on a home win and the outcome is a home win then the total return would be ‘ $a + b$ ’ pounds (profit a pounds) but otherwise would result in a loss to the bettor of b pounds.

The precise procedures with which bookmakers derive odds are best viewed as a commercially guarded secret although it is widely known that the odds are essentially judgement forecasts by panels of experts employed by bookmakers (Sharpe, 1997). The odds offered are usually made available approximately one-week prior to the game taking place. Although bookmakers always reserve the right to alter the odds on offer they seldom do so irrespective of betting volumes and irrespective of new information that may come to light during the course of the week (Forrest *et al.*, 2005). Typically odds would only alter in the rare event of incorrect odds being posted through a typographical error. It is in these senses that the odds are considered fixed.

Bookmakers’ odds may be converted readily into decimal odds, whereby the decimalised odds for the home win (d_h), away win (d_a) and draw (d_d)

are given by $d_h = \frac{a}{(a+b)}$; $d_a = \frac{c}{(c+d)}$; $d_d = \frac{e}{(e+f)}$ respectively. The

total of these decimalised odds, d , will always exceed unity and $(d - 1) \times 100$, known as the ‘over-round’, reflects an anticipated inbuilt profit for the bookmaker to offset cost of running the market assuming liabilities are evenly spread over the three outcomes. In the UK fixed odds football market the

over-round is currently about 10% per game. The values $d_{B,h} = d_h / d$;

$d_{B,a} = d_a / d$ and $d_{B,d} = d_d / d$ may be thought of as estimated bookmaker

probabilities of match outcomes. A small but precious point noteworthy of further explanation is the terminology “estimated bookmaker probability”. The estimated bookmakers’ probabilities are a simple multiplicative rescaling

of the decimal odds under the assumption that the over-round is spread proportionally amongst the decimalised odds (which may not necessarily be the case). In addition different bookmakers are at liberty to offer different odds giving rise to different estimated probabilities. Bookmakers' odds are set with commercial objectives in mind and as such the derived probabilities may not reflect their best estimates of the probabilities that they might otherwise derive. Figlewski (1979) and Knight (1965) make the important point that in games such as Roulette the probability of winning are known in advance and is therefore a game of risk but with no uncertainty. In the case of betting on a football match outcome the bettor is presented with both risk and uncertainty since although each team can be thought of as having a certain chance of winning the true chance will not be known.

There are large individual differences amongst people relating to gambling. Some will not gamble. Some will be risk-positive being attracted by large odds with the attendant bragging rights should a long shot pay dividends (Woodland, 1994). Avery and Chevalier (1999) found evidence of sentimental betting such as betting on well-known teams or on teams that have been covered in the media recently. Others may be attracted to betting on so-called "certainties" irrespective of the odds on offer. Others may bet on the football team that they support to win out of loyalty, or bet on their team to not win so as to have some financial comfort should their team not win. The behaviour of gamblers cannot be defined on a simple continuum.

In contrast a prevalent theory amongst statisticians and econometricians is the idea of developing betting strategies on "value bets" by systematically identifying football games in which there has been a perceived incorrect setting of bookmakers' odds through an inefficient use of information available i.e. "the market has failed to capture the information used by a superior analyst" Pankoff (1968, p.204). Our modelling approach partly revolves around the prediction of bookmaker estimated probabilities.

A number of authors have operationalised the idea of a value bet by building statistical models to predict the probabilities of match outcomes (say p_h, p_a, p_d) and to compare these estimated model probabilities with estimated

bookmakers probabilities by considering the ratios $r_h = p_h / d_{B,h}$;

$r_a = p_a / d_{B,a}$ and $r_d = p_d / d_{B,d}$. Betting strategies then take the form, for

example, to only bet on a home win outcome if $r_h > r_h^*$ where r_h^* is some determined constant chosen to balance risk and profitability. These strategies have been successfully implemented by Dixon and Coles (1997) and by Rue and Salvesen (2000) amongst others. In passing we comment that there are other ways of quantifying the degree of mismatch.

The statistical models for predicting match outcomes either model the match outcome directly using discrete choice models such as binary or ordinal logistic regression models or through models predicting probabilities of match

scores and then aggregating to predict match outcomes. The Poisson distribution or negative Poisson distribution are invariably used in this latter approach. In an early attempt to model match scores Reep *et al.* (1971) concluded that it was difficult to predict match outcomes confidently through this route and that chance dominates the game. Since then others have successfully used more computationally intensive complicated models (e.g. Dixon and Coles, 1997).

It is our view that working directly with goals scored and conceded may not be the correct stance. In league football a home win with a four-nil score line will be awarded the same number of points as a two-nil home win. During a match a team winning by two clear goals may change tactics or personnel and be content with that margin of victory rather than aiming to beat the opposition by the greatest possible score line. In other cases a very strong club may field a comparatively weakened team against some opposition with a view to resting star players for future games at the expense of a comparatively low winning margin. The primary aim for teams involved in a football match is the match outcome and (except in very restricted circumstances) the extent of goal difference is secondary. The primary aim of betting on the fixed odds market revolves around the match outcome and financial returns are not linked to the margin of victory. In a comparative study Goddard (2005) found very little difference in predictive ability between the discrete choice approach and models directly modelling goals scored and conceded. There is an argument that modelling exact scores may be more susceptible to the effect of outliers than a model based on match outcomes. For all of these reasons we have opted to work directly with discrete choice models and to model the match outcome directly.

Betting odds on the draw outcome invariably fall over a narrow range relative to the range of odds on offer for the home win or the away win. Pope and Peel (1989) suggest that this reflects a lack of ability of experts to forecast draws and Archontakis and Osborne (2007) argue that this could simply reflect a general inability to predict draw outcomes with any degree of reliability. Taken at face value this suggests that there may be inefficiencies in the draw odds market. However the prediction of the draw outcome is notoriously difficult. Prior to the development of fixed odds a common form of gambling amongst UK football fans was the treble chance football pools whereby bettors would try to identify draws from games to be played; monies staked would go into a pool and dividends paid on relative performance. Although highly popular this form of gambling was viewed by most as essentially a lottery (Forrest, 1999). Accordingly, although inefficiencies may exist in the draw odds, it does not necessarily follow that they may be systematically exploitable and for these reasons we will not consider the development of a betting strategy for the draw odds in this note.

In league football there is a well established home advantage effect (Clarke and Norman, 1995), and Dixon and Coles (1997) report that approximately 46 percent of games in the English football leagues result in a

home win. For these reasons it naturally follows that the odds offered for a home win are typically lower than those on offer for an away win. Consequently a betting strategy based on the away win outcome, although occurring less often than one based on the home win outcome, may have the potential for greater profits than a betting strategy for the home win simply because of the greater odds often found for the away win. For these reasons our modelling strategy will focus on the away win only.

Most of the published betting strategies in the statistical literature that have a positive expected return are based on models with estimated team specific parameters that are continually updated. This high dimensionally adversely impacts on the development of a practical betting rule. For these reasons we consider a low dimensional parsimonious model specification to underpin the betting rule.

Strategies based around betting on long shots or underdogs have been reported to be troublesome. For instance Thaler and Ziemba (1988) report on the favourite - long shot bias in racetrack betting whereby favourites tend to be under-backed and long shots tend to be over-backed. Bird and McCrae (1987) reported the strategy of betting on favourites or on long shots in horse racing would not yield a positive return. Likewise Woodland and Woodland (1994) report that the favourite – long shot bias is reversed in baseball betting and that no simple favourite or long shot betting strategy would produce a positive return. It is our contention that betting rules based on the form “bet on the home win if $r_h > r_h^*$ ” may be particularly susceptible to the inclusion of too many matches where there is a clear favourite. Our approach is based around a modification of this form of betting rule so as to allow the possibility of avoiding placing bets on matches involving clearly identified long-shot / favourite pairs.

Section 1 gives an overview of the methodological approach utilised to develop betting rules for the away win. A brief account of the data used and the rationale for the variables used in the model is given in Section 2. Derived statistical models and betting rules are given in Section 3 and the utility of the models is discussed in Section 4.

III. MODELLING APPROACH

Our modelling approach is based around sample data ($i = 1, \dots, I$) for deriving a betting rule and a second sample ($j = 1, \dots, J$) for out of sample assessment of the efficacy of the rules derived. Our first model is to directly estimate the probability of an away win using a discrete choice model.

Let $p_{a,i}$ denote a model based estimated probability that a match indexed by i will result in an away win, and let $d_{B,a,i}$ denote the corresponding estimated probability derived from the bookmaker’s fixed odds. Let $r_{a,i} = p_{a,i} / d_{B,a,i}$. We will consider a betting rule of the form “bet on the away win

for match i if, and only if, $r_l \leq r_{a,i} \leq r_u$ ” where r_l and r_u are constants chosen on criteria such as maximum profit per game or maximum profit. This form of rule is a more general version of the structure used by Dixon and Coles (1997) and Rue and Salvesen (2000) which use a one-sided limit utilising r_l only.

One approach to determine optimum values for r_l and r_u is to consider all possible values for r_l and r_u ($r_l \leq r_u$) and to apply these to sample values and choose the estimated parameters to be those values that maximise within sample profit per game or maximise within sample profit. This approach however may produce many small seemingly good profitable intervals but which not may be replicated on unseen data (particularly so if profit per game is considered). For this reason we consider determining r_l and r_u separately. The optimum value for r_l is chosen so that the betting rule “bet on the away win for game i if $r_{a,i} \geq r_l$ ” results in maximum within sample profit. Similarly, r_u is chosen so that the betting rule “bet on the away win if $r_{a,i} \leq r_u$ ” results in maximum profit (and not maximum profit per game). This approach is intended to produce a betting rule with greater relative robustness which does not overly capitalise on chance idiosyncratic sample characteristics.

In addition we consider the development of a betting rule of the form “bet on the away win for match i if and only if $p_l \leq p_{a,i} \leq p_u$ ”. The rationale behind this rule is to determine whether profitable rules can be developed whereby the away team is a strong favourite (in which case the rule would default to betting on the away win if $p_{a,i} \geq p_l$), or in opposing a strong home team (in which case the rule would default to betting on the away win if $p_{a,i} \leq p_u$) or whether it is better to focus betting on the away win when there is no seemingly clear favourite.

Our second approach is to use ordinary least squares regression with bookmakers’ odds as the dependent variable. Residuals (e_i) under the model may be used to assess the relative extent of disagreement between the bookmakers’ odds for the away win and the predicted bookmakers odds under the model. Following the reasoning given above, we consider a betting rule of the form “bet on the away win for match i , if and only if, $e_l \leq e_i \leq e_u$ ”.

IV. SAMPLE DATA

Data was recorded on the 194 league football games that took places between the 2nd October 2007 and 22nd October 2007 from the games played in the top four English football leagues and the top four Scottish football

leagues. The outcome of each game was recorded (home win, draw, away win) along with fixed odds for each outcome offered by Ladbrokes plc, the UK's largest bookmaker.

Fixed odds are set with commercial and financial gains in mind and may not necessarily reflect the best assessment of match outcomes since they may be set with anticipated betting volumes in mind or indeed set to influence betting volumes. For these reasons we consider as predictor variables the home and away team performance ratings published weekly by the Racing and Football Outlook (RFO) which is a weekly newspaper published by Trinity Mirror plc, dedicated to betting on horseracing and association football. The RFO index is an index based on the results of the past 60,000 games and provides a form rating on a scale of 0 to 1000 for each team in the English and Scottish football leagues. Increasing ratings are intended to reflect increasing ability of a team and the difference in RFO ratings between two teams is intended to reflect the extent of the degree of mismatch between the two chosen teams. The RFO produces a separate index for home and away performance to account for the home advantage effect and the extent of club specific home advantage effect (the home effect cannot be considered to be of the same influence for all teams). We therefore consider the RFO home rating for the home team and the RFO away rating for the away team as predictor variables for match outcomes.

Our second approach is to use a good predictor of betting odds which utilises information that might not be used by bookmakers in deriving odds. For this reason, for each team in each game, we consider the average proportion of time that the team was winning, irrespective of goal margin, in their previous three league games as a predictor variable. This choice of predictor is partly informed by the ready availability of the data and partly informed by the idea that the margin of victory is not of primary importance but that the percentage of time winning in previous games will still provide an indication of relative dominance in recent games against teams from the same league. We therefore consider average measure of time winning in previous three games and RFO ratings as predictor variables of estimated bookmaker probabilities.

A second data set comprising all of those matches held in the English and Scottish divisions (63 games) between 15th January 2008 and 21st January 2008 was used to assess independently the out of sample usefulness of the derived betting rules.

V. DERIVED BETTING RULES

Table 1 summarises the discrete choice complementary log-log model for predicting the probability of an away win. The complementary log-log model provided a statistically significant better overall fit than a logit specification. For an account of the complementary log-log model see Collett (1991, chapter 3). Overall the complementary log-log model is statistically

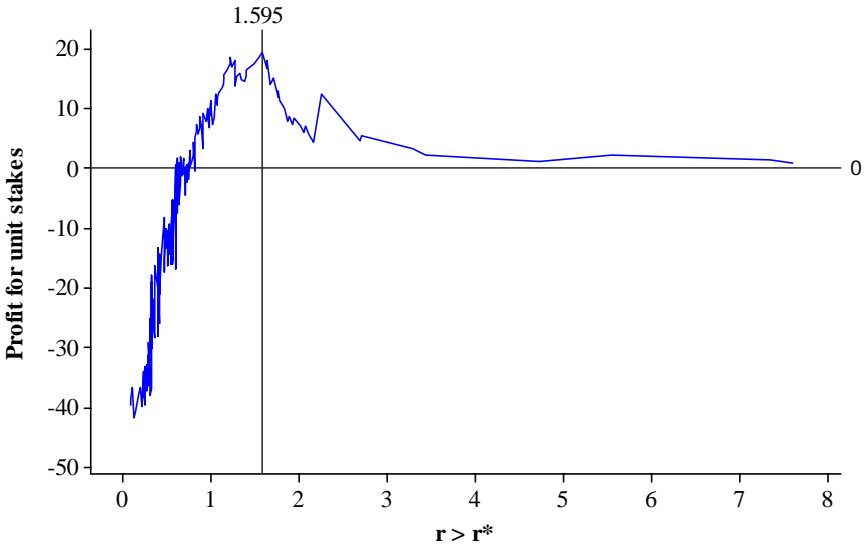
significant (Log-likelihood chi-square = 17.50, $df = 2$, $p < 0.001$), the individual predictors are statistically significant ($p < 0.001$) and the direction of effects for the RFO ratings for the home team at home (RFO HH) and the RFO ratings for the away team playing away (RFO AA) make good conceptual sense. The model adequately captures the structure in the data (percentage concordant pairs between model predictions and outcomes is 66.6%) and goodness-of-fit tests using Pearson's residuals ($p = 0.246$) and deviance residuals ($p = 0.106$) do not cast doubt on the appropriateness of the model specification. Inspection of delta beta and delta deviance graphics indicate that model does not suffer from the presence of overly influential observations. Prior to fitting this model we did consider a simple logistic specification however application of Brown's test indicated that a model with a non-symmetric link function would be more appropriate.

Table 1 Complementary log-log model for the probability of the away win

Variable	Coefficient (B)	SE(B)	Z	p
Constant	-1.1897	1.7051	-0.70	0.485
RFO HH	-0.0157	0.0038	-4.12	<0.001
RFO AA	0.0165	0.0042	3.90	<0.001

Figure 1 is a plot of within sample profit against possible choices for r_l for the betting rule “bet on the away win in match i if and only if $r_{a,i} \geq r_l$ ” with $r_{a,i}$ estimated for match i in the data set using the complementary log-log regression equation and with a one pound bet wagered each time the rule is fired. In this way the optimal value for r_l was found to be $r_l^* = 1.596$. In a similar way the value for the upper bound r_u^* was determined to be 7.597, which is the largest observed ratio in the data set. For the within sample data the rule “bet on the away win in match i if, and only if, $1.596 \leq r_{a,i} \leq 7.597$ ” effectively defaults to “bet on the away win if $r_{a,i} \geq 1.596$ ” and fired 29 times yielding an absolute profit of £19.43 giving a 67% profit on monies staked. When applied to the test data, the rule fired on 13 occasions giving an essentially break-even return of £0.80.

Figure 1: Profit from rule “bet on away win if $r > r^*$ ”



Applying the same procedure but using the predicted probabilities for an away win from the complementary log-log model gives the betting rule “bet on match i to be an away win if $0.4470 \leq p_{a,i} \leq 0.7146$ ”. This rule fired on 22 occasions and with £1 staked on each game an overall profit of £16.34 was obtained (i.e. a 74% return). When applied to the test data the rule fired on five occasions giving an overall percentage profit of 62.8%.

Table 2 summarises the fitted ordinary least squares model with the estimated bookmaker odds for the away win as the dependent variable. The overall model is statistically significant ($R^2 = 41.9\%$, $F(4, 195) = 34.04$, $MSE = 1.186$, $p < 0.001$), each predictor provides a unique statistically significant contribution to the model and the direction of the effects in the model make good conceptual sense. The model does not suffer with problems associated with multicollinearity (all variance inflation factors are less than 4). A visual examination of the residuals under the model suggests that the assumption of independence of errors has not been grossly violated although there is some evidence of a small departure from normality (Kolmogorov-Smirnov test statistics for normality has a p -value of 0.01). Adopting the same procedure as earlier, but using the residuals, gives a betting rule of the form “bet on the away win in match i if and only if $-0.1489 \leq e_i \leq 0.2042$ where e_i is the residual for match i . Application of this rule to the sample data gives rise to placing 28 bets yielding an overall profit of £9.75 (34.8% profit). Applying the rule to the out of sample data gives a percentage profit of 19.8%.

Table 2 OLS regression model with bookmaker odds of away win as the dependent variable

Variable	Coefficient <i>B</i>	SE(<i>B</i>)	<i>t</i>	<i>P</i>
Constant	1.057	1.0060	1.05	0.295
RFO HH	0.0202	0.0022	9.32	<0.001
RFO AA	-0.0194	0.0025	-7.79	<0.001
Time 1+ Home Team	0.0164	0.0047	3.50	0.001
Time 1+ Away Team	-0.0129	0.0048	-2.66	0.008

VI. DISCUSSION AND CONCLUSIONS

The preceding analyses indicate that a profitable betting strategy based on gambling on the away win may be possible. The results of the discrete choice model indicate that profitability may be obtained by avoiding those matches where there is a large estimated probability of an away win or a small estimated probability of an away win. This finding is consistent with previous research cautioning against a betting strategy based on a long-shot or on a clear favourite. Instead the derived rule suggests that it may be profitable to wager on the away win outcome on those seemingly difficult to call matches. This may be a reasonable finding if the extent of the home effect advantage has been incorrectly estimated by the bookmaker. The results from the “value bet” approach which considers the ratio of model estimated probability of the away win to the derived bookmaker probability of an away win as a betting trigger seem to be less spectacular.

Distinct from other approaches we considered the direct modelling of bookmaker odds using OLS regression. This analysis supported our prior reasoned hypothesis that average time winning in previous games is associated with the odds on offer. Distinct from other approaches we considered the residuals under the regression as quantifying the extent of mismatch between the bookmaker odds for the away win and the model predicted odds. The derived betting rule from this approach suggests avoiding betting on the away win when there is a large discrepancy between predicted values and bookmaker values and this finding is quite contrary to the usual stance of betting on so called “value” matches.

The results presented relate to league football only and due to the small sample size should be treated with caution. However we have only fitted prior reasoned models and have not undertaken a data “dredging” exercise which otherwise may have lead to too many false findings. In deriving and assessing the betting rules we have simply placed a one-unit stake per game. In practice it might be favourable to vary the stake in some optimal way (e.g. betting stakes in proportion to perceived risk) and on this basis the percentage returns quoted might be optimistically considered as an understatement. Likewise in practice a bettor will be in a position to shop around the different bookmakers for best prices for the away win and doing so would give a non-

trivial positive impact on the percentage returns offered. We chose to consider average winning times in the past three games a predictor variable although there may be further merit in extending this predictor variable over a different number of previous games.

A similar strategy could be considered for betting on the home win, however if betting rules for both the home win and away win are to be considered then some additional thought would have to be given to the possibility or prevention, of both rules firing on the same game.

REFERENCES

- Archontakis, F. & Osborne, E. (2007). Playing it safe? A Fibonacci strategy for soccer betting. *Journal of Sports Economics*, 8(3), 295-308.
- Avery, C. & Chevalier, J. (1999). Identifying investor sentiment from price paths: The case of football betting. *Journal of Business*, 72(4), 493-521.
- Bird, R. & McCrae, M. (1987). Tests of the efficiency of racetrack betting using bookmaker odds. *Management Science*, 33, 1552-1562.
- Clarke, S. R & Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. *Statistician*, 44, 509-521.
- Collett, D (1991) Modelling Binary Data, Chapman and Hall, London, Chapter 3
- Department of Culture, Media & Sport (2007). *Taking Part: The National Survey of Culture, Leisure and Sport*, Chapter 9, Gambling available from http://www.culture.gov.uk/images/research/TPMay2007_9_Gambling.pdf
- Dixon, M. J. & Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market, *Applied Statistics*, 46(2), 265-280.
- Figlewski, S. (1979). Subjective information and market efficiency in a betting market. *Journal of Political Economy*, 87(1), 75-88.
- Forrest, D. (1999). The past and the future of British football pools. *Journal of Gambling Studies*, 15(2), 161-176.
- Forrest, D., Goddard, J. & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, 21, 551-564
- Goddard, J. A. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21, 331-340.
- Knight, F. H. (1965). *Risk, Uncertainty and Profit*. New York: Harper Torchbooks.
- Pankoff, L. D. (1968). Market efficiency and football betting, *The Journal of Business*, 41(2), 203-214.
- Pope, P. F. & Peel, D. A. (1989). Prices and efficiency in a fixed-odds betting market. *Economica*, 56(223), 323-341
- Reep, C., Pollard, R. & Benjamin, B. (1971). Skill and chance in ball games. *Journal of the Royal Statistical Society, Series A*, 134, 623-629.
- Rue, H. & Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *The Statistician*, 49(3), 399-418.
- Sharpe, G. (1997). *Gambling on goals: A century of football betting*. Mainstream, Edinburgh
- Thaler, R. H. & Ziemba, W. T. (1988). Parimutuel betting markets: racetracks and lotteries. *Journal of Economic Perspectives*, 2, 161-174.
- Woodland, L. M. & Woodland, B. M. (1994). Market efficiency and the favorite-longshot bias: the baseball betting market. *Journal of Finance*, 49, 269-279.
- Woodland, L. M. (1994). Market efficiency and the favourite-longshot bias: The baseball betting market. *Journal of Finance*, 49(1), 269-279