Analyzing Information Efficiency in the Betting Market for Association Football League Winners

Lars Magnus Hvattum

Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Norway, lars.m.hvattum@iot.ntnu.no

Abstract

Sports betting markets have attracted a fair amount of research over the years. For association football, most of this research has focused on predicting the outcome of single matches and hence on the evaluating the efficiency of the match results betting markets. This paper presents a study on the betting market for league winners, a market that operates for almost a full year and therefore operates under different conditions than the relatively short lived match results markets. Attempts are made to analyze both weak and semi-strong forms of information efficiency, with results indicating that the market is inefficient with respect to both forms of information.

Keywords: Simulation, Regression, Forecasting, Soccer

1 Introduction

This paper presents a study on information efficiency in the betting market for association football league winners. Some indications of inefficiencies are found: First, an ordered logit regression model coupled with a Monte Carlo simulation produces predictions that are able to outperform the market over two seasons of the five biggest leagues in Europe. Second, simple betting rules based on price data can be seen to produce returns that are better than expected from blind betting. Third, arbitrage opportunities appear over significant periods of time, and fourth, certain apparently irrational changes of odds are observed, possibly to take advantage of irrational punters.

Hill [1974] suggested that match results in association football are governed in part both by chance and by skill. Unpredictability is therefore central to association football, and as Dobson and Goddard [2001] states, a key characteristic of the product that is sold to the spectators. In this paper, the predictability of final league standings is examined indirectly by developing a simulation model that produces probabilities that are then compared to market odds.

Vaughan Williams [2005] gives an introduction to information efficiency in betting markets. While most studies on market efficiency are performed for conventional financial markets, there are certain features of betting markets that make them particularly relevant when studying information efficiency: There is a large number of investors, that is bettors, with access to large quantities of information, and each asset, that is bet, has a well defined point in time where its final value becomes revealed. In contrast, while there are large number of investors in conventional financial markets, there is typically no fixed point in time where a true value is revealed, except in derivatives markets. Most studies on information efficiency in betting have focused on markets where the end time lies very near, such as betting on match outcomes. The current study considers seasonal bets, for which the end may be almost a year ahead. Thus, league winner betting is more similar to financial markets, where the value of an asset in the present is dependent both on the present value of future cash flows and also on the uncertain price at which it can be sold at some future point in time.

It is useful to differentiate between different forms of information efficiency. Vaughan Williams [2005] describes weak form information as information contained in the set of historical prices. If it is impossible to make economic profits by trading on the basis of information contained in historical prices, the market is said to be weakly efficient. Semi-strong information includes all information that is publicly known. In a semi-strong efficient market, it is impossible to gain economic profits by using only public information. Finally, strong information includes privately and monopolistically held information, and a market is strongly efficient if there are no possibilities for economic profits even when holding this information. Typically, when analyzing betting markets, a less strict concept is used, where an inefficiency is said to exist when it is possible to trade upon information to get greater than normal returns, which may correspond to smaller than typical economic losses.

For financial markets, Fortune [1991] concludes that the efficient market hypothesis (that is, the market is efficient) does not hold up, and cites well established anomalies giving above average returns. However, there is less evidence that this can be used to formulate profitable trading rules, and when there is such evidence, these possibilities tend to disappear quickly.

For the fixed odds betting market, Pope and Peel [1989] investigated prices of four bookmakers and found that the odds on the outcome draw had no significant predictive content. They found that positive pre-tax returns could be possible, but that neither of two examined betting strategies could provide profit after tax. Later, Dixon and Coles [1997] also succeeded at finding a strategy that yielded positive pre-tax returns. Dobson and Goddard [2001, Sec. 8.3] used weighted least squares estimation and rejected a condition for semi-strong form of efficiency.

However, the inefficiency was, again, not large enough to make a profit. Forrest et al. [2005] compared market odds to an ordered probit model. They found that the predictive power of the market odds improved over time, and that the odds setters probably used information not captured by the probit model. An ordered logit model with just a single covariate was tested in Hvattum and Arntzen [2010] and failed to show betting strategies that would fare better than placing bets at random. Constantinou et al. [2012] described a Bayesian network model, and found that their model failed to produce a profit when using only objective historical data. However, they observed that including subjective information (such as key player availability and psychological effects) yielded a model that generated profits, and hence indicated an inefficiency with respect to semi-strong information. Constantinou and Fenton [2012] went on to suggest that data-only approaches to prediction are limited since they fail to adjust quickly enough to new evidence. They also argued that the accuracy of the bookmakers had not improved, as opposed to the findings of Forrest et al. [2005] Furthermore, they found that there were some limited arbitrage opportunities, which would indicate that the market is inefficient even with respect to weak information.

Another indication of market inefficiency in betting markets is the favorite-longshot bias. In association football it is generally found that betting on a favorite gives a better return on investment than betting on a longshot. Cain et al. [2000] reported this longshot-bias in markets for both match scores and results. This conclusion has been supported by later studies on match results such as [Forrest et al., 2005, Graham and Stott, 2008, Constantinou and Fenton, 2012]. For studies on market efficiency in betting markets on other types of sports, we refer to references in [Constantinou and Fenton, 2012, Vaughan Williams, 2005]

Compared to fixed odds markets for match results, little work has been done to study the efficiency of betting markets for football league winners. However, some work has been done in relation to predicting winners of cup tournaments. Koning et al. [2003] devised a simulation model for football championships, which was assessed using the World Cups of 1994 and 1998 and the European Cups of 1996 and 2000. The method to produce match results was based on the weighted average number of goals scored and conceded for each team and on the assumption that goals scored can be described by a Poisson distribution. The simulation could handle both cup tournaments and league tournaments, although experiments on the latter were not reported. Min et al. [2008] proposed a framework for sports prediction using both Bayesian inference and rule-based reasoning. The framework was evaluated on the 2002 World Cup. Leitner et al. [2010] used simulation for the European Cup of 2008, where match results were predicted using Elo ratings as well as the FIFA ratings. Both approaches using ratings or ranks were outperformed by the odds provided by bookmakers prior to the start of the tournament. Suzuki et al. [2010] used FIFA ratings and specialists' opinions as input, assuming that goals scored are independent and Poisson distributed, and illustrated the method on the 2006 World Cup. In Rue and Salvesen [2000] one season of the English Premier League was simulated using posterior probabilities of match outcomes, but only to assess whether the final league table was surprising based on the final model output.

No inefficiencies were discovered in any of the studies on predicting cup tournament winners. Furthermore, market odds were shown to be superior to predictions based on ratings. In this paper a simulation model for league tournaments is proposed where match results are predicted using the ordered logit model of Hvattum and Arntzen [2010]. This logit model uses a single covariate that is based on the Elo rating, and which performs worse that the betting market in predicting the outcome of a single match. However, when simulating a whole season, the effects that influence the outcome of a single match (such as short term injuries and suspensions, current form, fatigue, and psychological effects) may become less important. Furthermore, the negative results for using ratings to predict cup tournament winners may be related to the fact that these are for national teams and not club teams: there is much less data available for national teams, and having good ratings may obviously depend on having sufficient amounts of data.

This paper continues in Section 2 by discussing betting rules and the simulation model used in the empirical study. The results of the empirical study are reported in Section 3, and Section 4 contains concluding remarks.

2 Experimental Setup

With an aim of searching for market inefficiencies, data was collected for two seasons (2008/09 and 2009/10) and for five different leagues (The English Premier League, the German Bundesliga, the Spanish Primera Divisíon, the Italian Serie A, and the French Ligue 1). Historic data about match results (and odds associated to match results) are freely available on the internet, and data used here was downloaded from [Football-Data.co.uk, 2010]. Historical odds from the league winner markets are not as easily available, and were collected semi-manually by the author using [BetBrain.com, 2010]. Data about matches and league winner odds were collected once per week, every week from season start to season end. This gives slightly more than 40 data points per league per season.

The analysis can be split in two parts. In the first part, simple rules based on odds data development are used to look for weak-form inefficiencies (Section 2.1). In the second part, a prediction model is built to look for semi-strong-form inefficiencies (Section 2.2). Since much of the analysis is based on simulated betting, the staking strategy may have some influence. Realizing that some betting rules may place a relatively large number of bets per week per league, using a unit bet size may be too volatile. Hence, the staking strategy is chosen to be such that a winning bet would return one unit, or equivalently that a winning bet would yield a profit equal to one minus the stake (for that bet). This means that a profit will be obtained every week that the eventual winner is included among the teams on which a bet is placed, and that every other week would result in a loss.

2.1 Simple betting rules

Two simple betting rules are considered as means to look for weak-form information inefficiency. The first is based on the observations of favorite-longshot biases in other sports, that is, that returns when betting on favorites are typically higher than returns when betting on long-shots. This leads to a simple rule stating that one should bet on the team that is the favorite to win the league (the team with the smallest decimal odds). If there is a similar favorite-longshot bias as in match results betting, this rule would be expected to perform slightly better than placing random bets.

The second rule can be used to see if the market responds rationally to new information, and is based on placing a bet on all teams for which the decimal odds is observed to be decreasing. This corresponds to a situation where the market thinks that the probability of the team winning the league has increased. If the market responds correctly to new information, the expected returns of such a betting scheme would be expected to be equal to that of random betting. If the observed returns are better than this, it could be an indication that the market reacts too slowly to new information, whereas if the returns are worse, it could be an indication that the market overreacts to new information.

2.2 Prediction model

Should the two betting rules above fail to indicate weak-form information inefficiencies, a more involved prediction model may be used to test for semi-strong form of information efficiency. This model uses historical data about match results to predict the final league standings, and consists of three steps: first, a dynamic calculation of ratings for each team; second, a regression model to convert rating differences into match results probabilities; third, a Monte Carlo simulation that simulates the remainder of the season using the probabilities obtained in the second step. Each of the three steps are now described in more detail.

The first step is to dynamically assign Elo ratings to each team involved in the league. Elo [1978] originally developed the rating system now bearing his name for assessing the relative strength of chessplayers. It was given an adaptation to association football by Hvattum and

Arntzen [2010], and the latter procedure is included in the prediction model presented here. Assuming that teams A and B has ratings r_A and r_B prior to playing a match. Let m_{AB} denote the result of the match, where $m_{AB} = 1$ if A won the match, $m_{AB} = 0.5$ if the match was drawn, and $m_{AB} = 0$ if B won the match. Let g_A be the number of goals scored by team A and g_B be the number of goals scored by team B. The new rating r'_A for team A after the match becomes:

$$r'_A = r_A + k\left(1 + |g_A - g_B|\right)\left(m_{AB} - \frac{1}{1 + c^{(r_B - r_A)/d}}\right)$$

where c and d are parameters that influence the scaling of the ratings and k is a parameter that influence how quickly new match results affect ratings of the teams. A low k would yield ratings that are stable over time, whereas a high k will yield ratings that fluctuate quickly. The parameters were calibrated in [Hvattum and Arntzen, 2010] to optimize the prediction power of the ratings for the English Premier League up to the 2007/08 season, and the same values are used here: $k = 10$, $c = 10$, and $d = 400$. A mechanism to calculate initial ratings is also needed. For this a bootstrapping procedure is used, where two or more full seasons (across two or more divisions of the league system) are used for initialization only. All teams are initially assigned a rating of 0, and the dynamic rating update is then run for the set of matches selected for the bootstrapping procedure. If the final ratings after running the ratings updates are sufficiently different from the start ratings, the final ratings are taken as new initial ratings and the procedure is repeated.

The second step is to convert rating differences into match results probabilities. Assume a large amount of historical match results is available, and that the Elo ratings have been calculated up to the point of each match being played. A large number of pairs are then available of the type (x, y) where x is the rating difference between the home team and the away team and y is the outcome of the match, encoded as $y = 1$ if the home team won, $y = 2$ if the match was drawn, and $y = 3$ if the away team won. As in [Hvattum and Arntzen, 2010] an ordered logit regression model [Greene, 1999] can then be used to find a transition from a rating difference x to probabilities over the three possible match outcomes $y \in \{1, 2, 3\}$. While in principle the regression model may be recalculated when new match results are obtained, the effect of this is minor and for the league standings prediction model presented here, the regression is only run once, just prior to the start of the league.

The final step is to use simulation to find the probability that a team will win the league. The simulation procedure is straightforward and only considers double round robin tournaments, where each team plays the other teams twice (once at home and once away). Since the regression model in step two only produces a probability distribution over the match outcomes and disregards victory margins, the simulation procedure will also disregard goals differences. At any given point, each team has a rating. For each unplayed match, the rating difference is used to find a probability distribution over the match outcomes based on the regression model calculated in step two. In a single simulation, each unplayed match is simulated by drawing from this probability distribution and noting the simulated match outcome. At the end of the simulation, the total number of points is calculated for each team, with 3 points given to a win (either in an already played match or in a simulated match), with 1 point given for a draw, and with 0 points given for a loss. A count is kept over how many times each team wins the league (that is, obtains the highest number of points) according to the simulation. If, in a single simulation run, n teams should end up with the same number of points, and this amount of points is the maximum amount of points that any team achieves, the count of each of these teams is increased by $1/n$. By running the simulation a large number of times, the count of a team divided by the total number of simulation runs will represent the probability that the team wins the league.

3 Results

This section reports on the results of examining two seasons (2008/09 and 2009/10) for five European top leagues (The English Premier League, the German Bundesliga, the Spanish Primera Divisíon, the Italian Serie A, and the French Ligue 1), using the methods described in Section 2.

First, Table 1 and 2 shows results of simulated betting for seasons 2008/09 and 2009/10 respectively. There are three betting strategies considered: Elo, where a bet is placed whenever market odds is more generous than indicated by the probabilities obtained by the prediction model described in Section 2.2, Imp, where a bet is placed whenever the market odds of a given team is reduced (compared to the previous week), and Fav, where a bet is placed on the team that is favorite to win according to the best odds available from the bookmakers. The bet size is always set equal to the inverse of the decimal odds.

As there are essentially only 10 data points, the results are not significant in a statistical sense. However, it is seen that Elo produces overall profits for both seasons, whereas Imp and Fav both have one season with profits and one season with losses. There is much more variance in the overall seasonal result for Fav than the other methods, which may be related to the fact that it only places one bet per league per week, whereas the other methods may place bets on several teams.

The hypothesis that the market is weakly efficient must be considered weakened. The Fav method gives overall profit after two seasons, and even Imp shows a result that is better than should be expected from blind betting: Based on the calculated overround of the best odds available, blind betting should return 97.0 % for the 2008/09 season and 98.0 % for the 2009/10 season. Imp, however, returns 102.4 % for the 2008/09 season and 96.6 % for the 2009/10 season, or combined for the two seasons: 99.6 %. As Fav returns an overall profit, this is an indication of a favorite-longshot bias, similar to that which has been observed in match results betting markets.

	Elo			Imp			Fav		
	Staked	Profits	ROI	Staked	Profits	ROI	Staked	Profits	ROI
England	25.59	-1.59	0.938	20.86	0.14	1.007	24.52	-2.52	0.897
France	18.41	7.59	1.413	20.01	0.99	1.049	27.71	-23.71	0.144
Germany	17.95	-5.95	0.669	20.55	-2.55	0.876	23.86	-21.86	0.084
Italy	29.52	10.48	1.355	21.95	2.05	1.093	26.67	13.33	1.500
Spain	19.00	-8.00	0.579	19.20	1.80	1.094	29.43	6.57	1.223
Total	110.47	2.53	1.023	102.58	2.42	1.024	132.18	-28.18	0.787

Table 1: Summary of results for simulated bets in the 2008/09 season.

	Elo			Imp			Fav		
	Staked	Profits	ROI	Staked	Profits	ROI	Staked	Profits	ROI
England	20.60	8.40	1.408	19.81	-1.81	0.909	19.34	15.66	1.810
France	21.33	2.67	1.125	19.29	0.71	1.037	21.50	-16.50	0.233
Germany	15.14	-9.14	0.396	18.87	-0.87	0.954	21.49	16.51	1.768
Italy	27.13	9.87	1.364	20.28	-1.28	0.937	27.21	10.79	1.397
Spain	24.64	14.36	1.583	18.06	-0.06	0.997	25.68	12.32	1.480
Total	108.84	26.16	1.240	96.31	-3.31	0.966	115.21	38.79	1.337

Table 2: Summary of results for simulated bets in the 2009/10 season.

Another factor that may weaken the hypothesis that the market is weakly efficient is an observation of arbitrage opportunities. While the average overround implies that blind betting should return around 97.5 % of the amount staked, the overround based on best odds is not stable, and for a surprisingly high number of weeks during the seasons, the best odds in the data set implies an arbitrage opportunity. This holds true in particular for the Bundesliga in the 2008/09 season where the observed overround was negative for 23 out of 42 weeks, and for all the five leagues in the 2009/10 season where between 5 and 12 weeks had negative overround for each of the leagues examined. For Bundesliga in the 2008/09 season, all the three betting strategies tested ended with losses despite the favorable overround. One may speculate that the reason for the prevalent arbitrage opportunities were due to some bookmakers using superior information and insight to set better odds, without being able to influence the odds in the rest of the market (the eventual winner, Wolfsburg, was somewhat surprising, and bets could be placed on the team for odds as high as 150 well after the league was half played).

There are some arguments as to why the observation of arbitrage opportunities should not

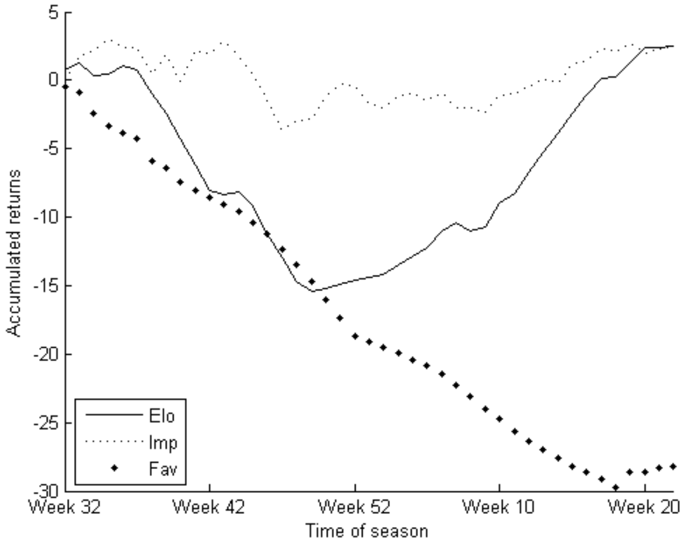

Figure 1: Returns on simulated betting for the 2008/09 season.

be seen as evidence of an inefficient market. First, when they are observed early in the season, the argument is that a better return on investment could be had by simply investing the money in another safe asset, since a winning bet on the league winner would not pay out until more than half a year later. Second, when they are observed late in the season, most of the time the bookmakers are quoting odds only for selected teams. This means that it is impossible to place bets on some teams that only has a (near) zero chance of winning, and hence it is not really an arbitrage opportunity anyway since there is always a non-zero chance that administrative changes (such as other teams receiving penalties that affect the final standings) may lead to any team being awarded the trophy. Disregarding this, if all ostensible arbitrage opportunities were followed, one could have had a return on investment of 101.8 % in the 2008/09 season and 102.0 % in the 2009/10 season (notably worse, though, than the return on investment obtained by the Elo betting strategy).

While there seems to be some indications of weak-form inefficiency, it is still interesting to examine the prediction model described in Section 2.2. The results of betting according to this strategy was already summarized in Tables 1 and 2. Figures 1 and 2 illustrates at which point in time profits or losses were accumulated. The most notable observation is that Elo would end with a loss of up to 15 units if betting was ended near the middle of the 2008/09 season, but that bets placed after that were sufficient to end with a small profit.

Having positive returns on betting based on the model is a clear indication of a semi-strong

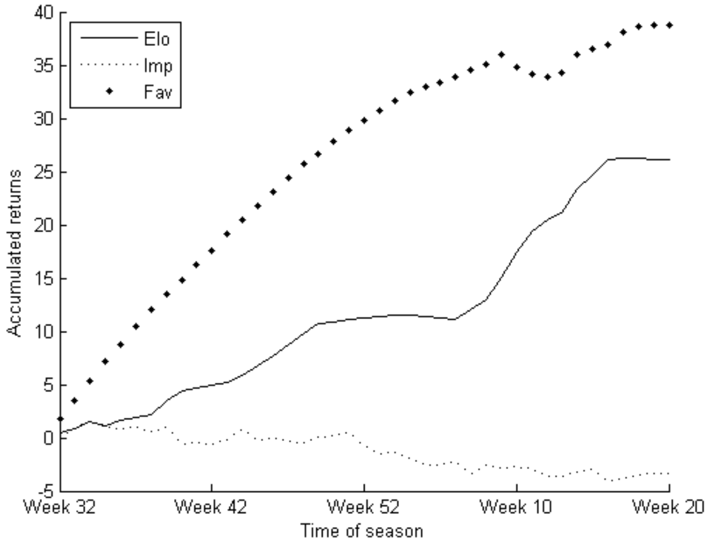

Figure 2: Returns on simulated betting for the 2009/10 season.

information inefficiency. However, historical betting returns often have high variance and are not very reliable as a means to predict future returns. Johnstone [2007] suggested that forecasters selected their methods based on mathematical loss functions rather than historical profits. In the following, a quadratic loss function [Witten and Frank, 2005] also known as a Brier score [Brier, 1950] is used to compare the market odds with the odds generated through the Monte Carlo simulation.

Let p_A be the probability (at a given time using a given prediction method) that team A wins the league. If T is the set of teams and the eventual winner is W, the quadratic loss can be expressed as

$$L^2 = (1 - p_W)^2 + \sum_{A \in T \setminus \{W\}} (p_A)^2$$

Figures 3 and 4 shows the average quadratic loss calculated for each week of seasons 2008/09 and 2009/10, respectively. The best market odds are converted to probabilities by taking the inverse of the decimal odds and then normalizing so that the sum of the inverses equals 1. It is evident that the the loss decreases towards the end of the season, as more information becomes available and it becomes easier to predict the eventual winner. It is also clear that the loss calculated based on market odds follows the same development as the loss calculated based on the probabilities of the prediction model. However, the simulation model produced predictions with lower loss in 49 out of 85 weeks, when considering the average loss over all five leagues. Also, the average loss over all weeks is lower for the simulation model probabilities than the implied market probabilities, with 0.608 against 0.614 for the 2008/09 season, and 0.380 against 0.404 for the 2009/10 season (weighing all weeks equally). This is thus another indication of an information inefficiency.

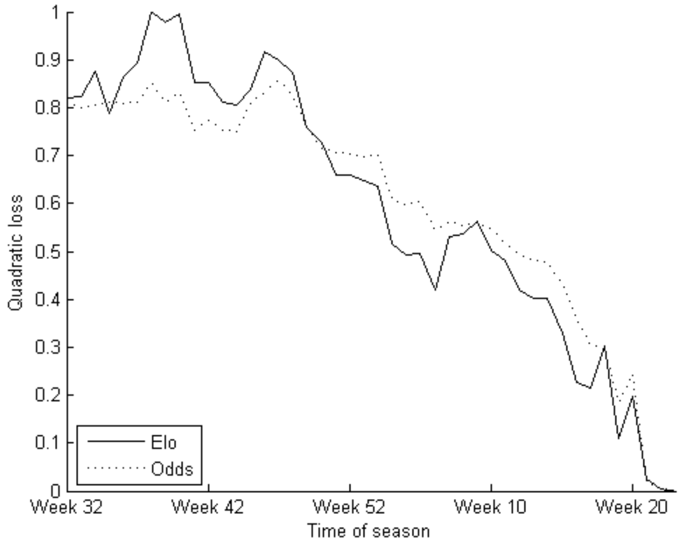

Figure 3: Average quadratic loss over the 2008/09 season.

Finally, this section ends by presenting an anecdotal evidence of another market inefficiency. This happened when the announcement was made that there were new owners of Manchester City that made promises of spending lots of cash on player transfers. The news broke around week 37 in 2008, and the effect on the odds of Manchester City winning the league in 2008/09 is shown in Figure 5: the odds suddenly fell from a level of more than 600 to a level of just above 50. It then took more than 10 weeks for the odds to recover. The new owners did eventually spend much on new players; in fact the club spent around £160M, £77M, and £75M on new players at the start of seasons 2009/10, 2010/11, and 2011/12, according to Constantinou et al. [2012]. It seems like an overreaction, though, to think that the new owners would be able to influence the winning chances in the ongoing season, especially since the transfer window would not re-open until the following January. In the 2011/12 season Manchester City won the Premier League.

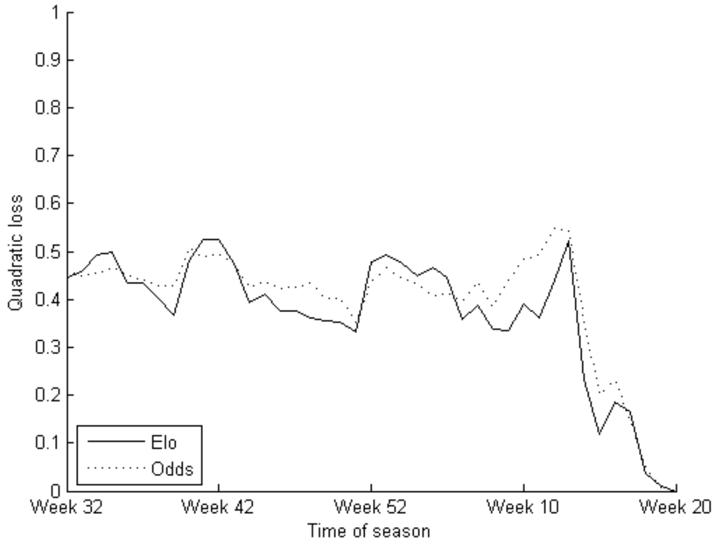

Figure 4: Average quadratic loss over the 2009/10 season.

4 Concluding Remarks

The betting market for league winners is small compared to other football betting markets. While there have been several studies focusing on the efficiency of the betting market for match results and some studies focusing on markets for cup winners, the betting market for league winners has not been extensively studied. This paper shows several indications of inefficiencies in that market. A model based on Monte Carlo simulation provides an overall profit for two seasons of the top five leagues of Europe. An evaluation using a quadratic loss function indicates that the probabilities generated by the simulation model are more accurate than the probabilities implied by market odds. Two simple betting rules also give returns that are better than one should expect based on the odds available, and one of them even provides an overall profit. Finally, some persistent arbitrage opportunities are found, possibly indicating that some bookmakers are able to generate better probabilities than others while not being able to instantly influence the odds provided by other less informed bookmakers.

References

BetBrain.com. http://www.betbrain.com/, 2010. Accessed between 2008 and 2010.

G.W. Brier. Verification of weather forecasts expressed in terms of probability. *Monthly Weather*

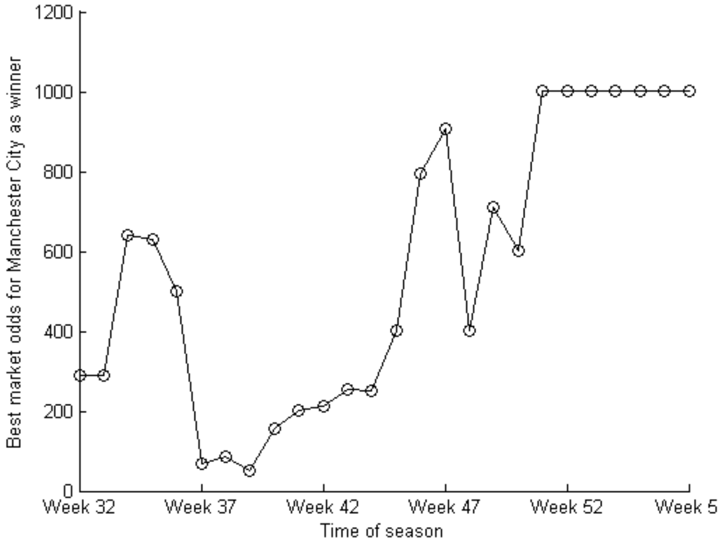

Figure 5: Best market odds for Manchester City as the winner of the 2008/09 season.

Review, 78:1–3, 1950.

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

A.C. Constantinou and N.E. Fenton. Evidence of an (intended) inefficient association football gambling market. 2012.

A.C. Constantinou, N.E. Fenton, and M. Neil. pi-football: a Bayesian network model for forecasting association football match outcomes. *Knowledge-Based Systems*, 2012. doi: 10.1016/j.knosys.2012.07.008.

M.J. Dixon and S.C. Coles. Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics*, 46:265–280, 1997.

S. Dobson and J. Goddard. *The Economics of Football.* Cambridge University Press, Cambridge, 2001.

A.E. Elo. *The Rating of Chessplayers, Past and Present.* Arco Publishing, New York, 1978.

Football-Data.co.uk. http://www.football-data.co.uk/, 2010. Accessed between 2008 and 2010.

D. Forrest, J. Goddard, and R. Simmons. Odds-setters as forecasters: the case of English football. *International Journal of Forecasting*, 21:551–564, 2005.

P. Fortune. Stock market efficiency: An autopsy? *New England Economic Review*, pages 17–40, March 1991.

I. Graham and H. Stott. Predicting bookmaker odds and efficiency for UK football. *Applied Economics*, 40:99–109, 2008.

W.H. Greene. *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ, 4th edition, 1999.

I.D. Hill. Association football and statistical inference. *Applied Statistics*, 23:203–208, 1974.

L.M. Hvattum and H. Arntzen. Using elo ratings for match result prediction in association football. *International Journal of Forecasting*, 26:460–470, 2010.

D. Johnstone. Economic darwinism: who has the best probabilities? *Theory and Decision*, 62: 47–96, 2007.

R.H. Koning, M. Koolhaas, G. Renes, and G. Ridder. A simulation model for football championships. *European Journal of Operational Research*, 148:268–276, 2003.

C. Leitner, A. Zeileis, and K. Hornik. Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008. *International Journal of Forecasting*, 2010.

B. Min, J. Kim, C. Choe, H. Eom, and R.I. McKay. A compound framework for sports results prediction: A football case study. *Knowledge-Based Systems*, 21:551–562, 2008.

P.F. Pope and D.A. Peel. Information, prices and efficiency in a fixed odds betting market. *Economica*, 56:323–341, 1989.

H. Rue and Ø. Salvesen. Prediction and retrospective analysis of soccer matches in a league. *The Statistician*, 49:399–418, 2000.

A.K. Suzuki, L.E.B. Salasar, J.G. Leite, and F. Louzada-Neto. A Bayesian approach for predicting match outcomes: the 2006 (association) football world cup. *Journal of the Operational Research Society*, 61(10):1530–1539, 2010.

L. Vaughan Williams, editor. *Information Efficiency in Financial and Bettimg Markets*. Cambridge University Press, Cambridge, 2005.

I.H. Witten and E. Frank. *Data mining: practical machine learning tools and techniques*. Elsevier, San Francisco, CA, 2005.