A Multinomial Logit-based Statistical Test of Association Football Betting Market Efficiency

Henri Nyberg
University of Helsinki and HECER

A Multinomial Logit-based Statistical Test of Association Football Betting Market Efficiency*

Abstract

This paper presents a novel statistical test for the informational efficiency of bookmakers' odds in the association football betting market. The new test has several advantages in terms of interpretation and statistical properties over the previously considered match outcome-specific tests. According to simulation experiments, the empirical size of the classical likelihood-based test statistics is satisfactory, whereas the power is dependent on the degree of market inefficiency and the sample size. A large English football dataset reveals that the efficient market hypothesis cannot be statistically rejected except for the Premier League games.

Henri Nyberg

Department of Political and Economic Studies
University of Helsinki
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: henri.nyberg@helsinki.fi

1   Introduction

In applied econometric and statistical research increasing interest has been paid on the predictability of the results of association football (soccer) matches and the consequent implications to the efficiency of the betting market. Due to the similarity with financial markets, betting markets provide an interesting field to examine the efficient market hypothesis. The betting odds offered by different bookmakers (odds-setters) produce information on the underlying probabilities of the football match results in a similar way as traditional financial markets reflect the future expectations on the fundamentals determining asset prices. In fact, as pointed out by Thaler and Ziemba (1988), bettors usually know the future payoffs and each bet has a well-defined termination point at which its value becomes certain. Thus, the betting markets may suit even better for testing market efficiency than financial markets.

In finance, testing the efficient market hypothesis and its three forms (weak, semi-strong and strong efficiency) has been a major research topic already several decades (see, e.g., Summers, 1986; Malkiel, 2003; Lim and Brooks, 2011). Similarly as in the stock market, we can evaluate whether the implicit probability estimates for different match outcomes (i.e. home win, draw and away win) obtained from the quoted betting odds of different bookmakers rationally reflect the underlying fundamentals. If the implicit probabilities do not reflect all the relevant information, the weak-form efficiency condition is not met (see the discussion, e.g., in Goddard and Asimakopoulos (2004) and Vlastakis *et al.* (2009)) and it would be possible to obtain systematically more accurate probability forecasts.

In this study, we propose a novel statistical test for the (weak-form) association football betting market efficiency. The test is based on the multinomial logit model where the informational efficiency of the implicit probabilities can be straightforwardly tested. This is not generally the case in the traditional, and so far dominant, statistical modeling approach where different count data models have been used to predict the number of goals scored by the home and away teams (see, e.g., Dixon and Coles, 1997; Rue and Salvesen, 2000; Crowder *et al.*, 2002; Karlis and Ntzoufras, 2003; Skinner and Freeman, 2009; Bastos and da Rosa, 2013; Koopman and Lit,

2014). In addition to the market efficiency interpretation, our test can also be interpreted as a rationality or unbiasedness test of the implicit probabilities. Although we are interested in aggregated market level data, the proposed test can analogously be used to check whether the probability forecasts of one bettor or the odds offered by a single bookmaker are informationally efficient.

This study appears to be the first one to consider the statistical properties of the efficiency tests using Monte Carlo simulation experiments. The main interest is in the proposed multinomial logit-based test but we will also make comparisons to the previously examined approach where independent binary logit (and linear probability) models have been used to test the informational efficiency of each match outcome separately (see, e.g., Pope and Peel, 1989; Goddard and Asimakopoulos; 2004; Franck *et al.*, 2010; Koning, 2012). An advantage of our approach is that the three possible match outcomes are modeled simultaneously within one model. Efficiency testing can thus be based on one test statistic instead of three obtained with the match outcome-specific testing procedure. Furthermore, it is not even possible to test the informational efficiency of the implicit probabilities explicitly with the count data models (see the references above) or the ordered logit and probit models (see Forrest and Simmons, 2000; Koning, 2000; Goddard and Asimakopoulos, 2004; Goddard 2005; Forrest et al. 2005) examined in the past research.

In our simulation experiments, the small-sample properties of the classical likelihood-based (Likelihood ratio, Wald and Lagrange multiplier) test statistics show that the test based on the multinomial logit model outperforms the previously used outcome-specific testing approach. The size of the multinomial logit-based tests is satisfactory and there are no large differences between different test statistics. We also find that the power of the test is strongly dependent on the degree of inefficiency and the sample size. The power is high when clear inefficiencies exist but under the hypothesis of only marginally inefficient betting market the sample size must be very large before the test has reasonable power. This is in line with the findings of Gandar *et al.* (1988), Golec and Tamarkin (1991), Gray and Gray (1997) and Goddard and Asimakopuolos (2004), among others. They have found, without explicit simula-

tion evidence and concentrating mainly on the NFL matches, that the power of the previously considered efficiency tests has been low.

In our empirical application, we consider the efficiency of the English football betting market using the data containing the matches played in the top four league divisions between the years 2000 and 2013. Overall, the results show that statistically there are no large deviations from the efficient market hypothesis. The only exception is the Premier League (the highest tier of the English football league system) where the efficient market hypothesis is strongly rejected although, as suggested by Goddard and Asimakopoulos (2004), the testing results are also in that case partly dependent on the stage of the season.

The rest of the paper is organized as follows. In Section 2, we briefly explain how the bookmakers' implicit probabilities can be obtained from the betting odds before introducing the multinomial logit model and the new efficiency test in Section 3. The results of the Monte Carlo simulation experiments and English football data are reported in Section 4. Finally, Section 5 concludes.

2 Implicit Probabilities as Predictors

In this study, we aim to test the informational efficiency of the underlying implicit probabilities derived from the bookmakers' odds as predictors of the outcomes of association football matches. For clarity, to make a distinction to the goals scored by the home and away teams (which eventually determine the final score of the match), we refer the result of the association football match (home win, draw or away win) to as an outcome.

Before introducing the new statistical test in Section 3, in this section we will briefly consider how to obtain the implicit probabilities from the betting odds. Throughout this paper, we concentrate purely on the statistical side of market efficiency testing. In addition to the statistical tests and other statistical procedures, various economic-based tests and betting rules have also been considered in the past research when trying to formulate profitable betting strategies (see, e.g., Vlastakis *et al.*, 2009; Direr, 2013, and the references therein).

In our application to English football data (see Section 4), the dataset contains bookmakers operating in the fixed odds betting markets. In these markets, a bookmaker acts as a market maker determining the odds (payouts) of the football match that can be taken by a bettor by staking some amount of money on that bet. As the odds are fixed at the time when the bettor and bookmaker enter the contract, this type of betting is called fixed odds betting.

Goddard and Asimakopoulos (2004), Franck et $al.$ (2010) and Koning (2012), among others, have explained how the bookmakers' odds for the association football matches can be converted to the implicit probability estimates. To this end, denote the kth bookmaker's (decimal) odds for the match i by O_{ik}^H, O_{ik}^D and O_{ik}^A, respectively. The implicit probabilities for the home win ($P_{ik}^{ip,H}$), draw ($P_{ik}^{ip,D}$) and away win ($P_{ik}^{ip,A}$) can be obtained as follows:

$$
\begin{aligned}
P_{ik}^{ip,H} &= \frac{1}{O_{ik}^H} \times \frac{1}{\xi_{ik}+1} \\
P_{ik}^{ip,D} &= \frac{1}{O_{ik}^D} \times \frac{1}{\xi_{ik}+1} \\
P_{ik}^{ip,A} &= \frac{1}{O_{ik}^A} \times \frac{1}{\xi_{ik}+1},
\end{aligned}
\tag{1}
$$

where the bookmaker's implicit marginal ("over-round") $\xi_{ik} > 0$ can be written

$$
\xi_{ik} = \frac{1}{O_{ik}^H} + \frac{1}{O_{ik}^D} + \frac{1}{O_{ik}^A} - 1.
$$

The resulting probabilities (1) will sum up to one. As our aim is to explore the efficiency of the betting market as a whole, in line with the previous literature, we assume that the averages of implicit probabilities (1) (over K bookmakers)

$$
\bar{P}_i^{ip,H} = \frac{1}{K}\sum_{k=1}^K P_{ik}^{ip,H}, \quad \bar{P}_i^{ip,D} = \frac{1}{K}\sum_{k=1}^K P_{ik}^{ip,D}, \quad \bar{P}_i^{ip,A} = \frac{1}{K}\sum_{k=1}^K P_{ik}^{ip,A},
\tag{2}
$$

reflect the market consensus. To test the hypothesis that the probabilities (2) are in statistical sense informationally efficient, those should be unbiased predictors of the actual match outcomes. In Section 3, we show that this hypothesis can straightforwardly be tested by using the multinomial logit model.

Throughout this study, in accordance with Goddard and Asimakopoulos (2004) and Koning (2012), among others, we refer our test to as a statistical (weak-form)

efficiency test. Alternatively, this test can also be interpreted as an unbiasedness or rationality test of the average implicit probabilities (2). It is, however, important to point out that in the bookmakers' perspective the implicit probabilities (1) are not necessarily their "true" or the best unbiased predictions complicating the efficiency interpretation of the results. The bookmakers' objective is to maximize their profits. Therefore, as formalized by Levitt (2004) and discussed in Vlastakis *et al.* (2009) and Franck *et al.* (2010), they may set their odds (slightly) inefficiently if they are influenced by both their true probability estimates and the bettors' demand (i.e. expected betting volumes) for each match outcome given the level of the quoted odds. Although the odds O_{ik} are publicly available, this is not the case for the distribution of betting volumes and, hence, the constructed implicit probabilities (1) may deviate from the bookmakers' true probability estimates.

For simplicity and following the standard approach employed in the previous literature (see, e.g., Forrest *et al.*, 2005; Franck *et al.*, 2010), to obtain the implicit probabilities (1), it is reasonable to assume that the betting volumes for different bets are (approximately) uniformly distributed across the outcomes. If this assumption is not valid, then one potential reason for the inefficiency of the implicit probabilities is the bookmakers' intention to predict also betting volumes when setting their odds. However, as we are interested in market averages (2), the role of this irrationality is expected to be small at most. The individual bookmakers are exposed to substantial risks if their odds deviate systematically from their underlying true probabilities (Franck *et al.*, 2010). This is the case especially since the emergence of internet betting which has substantially increased competition (Forrest *et al.*, 2005) and increased incentives for bookmakers to set their odds such that deviations from their true probability estimates are small.

3 Statistical Model and Market Efficiency Test

3.1 Multinomial Logit Model

In the previous academic sports betting literature, a multinomial logit model has been used to evaluate the market efficiency in horse racing (Figlewski, 1979) and predict the results of test cricket matches (Akhtar and Scarf, 2012). To the best of our knowledge, Vlastakis *et al.* (2009) is the only previous study where the multinomial logit model has been used to analyze (association) football games and related betting markets. However, their objective was not to test the market efficiency directly using a similar kind of explicit statistical test as in this study. Instead, Vlastakis *et al.* (2009) used the multinomial logit model as an alternative to the goal-based statistical models (see, e.g., Dixon and Coles, 1997; Rue and Salvesen, 2000; Crowder *et al.*, 2002; Dixon and Pope, 2004; Skinner and Freeman, 2009; Bastos and da Rosa, 2013; Koopman and Lit, 2014) to predict the match outcomes using the implicit market probabilities (2) as predictive variables. They also concentrated mainly on different betting strategies and their profitability as well as potential arbitrage opportunities between different bookmakers.

As we have seen in Section 2, there are three possible outcomes in an association football match implying a natural use of the multinomial logit model. Let us denote the outcome of the match i as $y_i = j$, where j is the result of the match. A home win (denoted by H in Section 2) is denoted by $y_i = 1$ while draw (D) $y_i = 0$ and away (A) win $y_i = -1$ are other possible outcomes. For the subsequent notation, let us introduce three binary indicator variables $y_{i,j}$ where one gets the value one for each i depending on the outcome of the match. For example, if the result of the game i is a home win, then $y_{i,1} = 1$ and $y_{i,0} = y_{i,-1} = 0$. It is worth noting that in contrast to the ordered multiresponse models, such as the ordered logit and probit models employed by Koning (2000), Goddard and Asimakopoulos (2004), Goddard (2005) and Forrest *et. al.* (2005), the ordering of the match results is not critical in the multinomial logit model.

Following the notation above, let us contain the average implicit probabilities (2)

6

to the vector

$$\bar{\mathbf{P}}_i^{ip} = [\bar{P}_i^{ip,1} \quad \bar{P}_i^{ip,0} \quad \bar{P}_i^{ip,-1}]'. \tag{3}$$

The multinomial logit model is specified when determining the conditional probabilities $P_{i,j}$ $(j = -1, 0, 1)$ of the match outcomes $y_i = j$ conditional on the relevant predictive information that is, under the efficient market hypothesis, completely included in (3). The model can be written using the log-odds ratios (see, e.g., Greene 2012, pp. 803–805)

$$\log\left(\frac{P_{i,j}}{P_{i,0}}\right) = \pi_{i,j}, \tag{4}$$

where the linear functions $\pi_{i,j}$, $j = -1, 1$, should be determined to complete the model. Using the implicit probabilities (3), the log-odds ratios reduce to

$$\log\left(\frac{\bar{P}_i^{ip,j}}{\bar{P}_i^{ip,0}}\right) = \bar{\pi}_{i,j}^{ip}. \tag{5}$$

Therefore, as we are interested in the predictive power of the implicit probabilities (3), the linear functions given in (4) are specified as

$$\pi_{i,j} = \alpha_j + \bar{\pi}_{i,j}^{ip}\beta_j, \quad j = -1, 1, \tag{6}$$

where the parameters α_j and β_j are match outcome-specific which is not the case, for example, in the ordered probit model complicating efficiency testing therein (see Section 3.3). Following Vlastakis *et al.* (2009), the draw ($y_i = 0$) is used as a benchmark category in (4) indicating that the linear function (6) is not determined to that outcome.

Solving for the conditional probabilities $P_{i,j}$ from (4), we get

$$\begin{aligned}
P_{i,1} &= P(y_i = 1|\bar{\mathbf{P}}_i^{ip}) = \frac{\exp(\pi_{i,1})}{1 + \exp(\pi_{i,1}) + \exp(\pi_{i,-1})} \\
P_{i,0} &= P(y_i = 0|\bar{\mathbf{P}}_i^{ip}) = \frac{1}{1 + \exp(\pi_{i,1}) + \exp(\pi_{i,-1})} \\
P_{i,-1} &= P(y_i = -1|\bar{\mathbf{P}}_i^{ip}) = \frac{\exp(\pi_{i,-1})}{1 + \exp(\pi_{i,1}) + \exp(\pi_{i,-1})},
\end{aligned} \tag{7}$$

where $\sum_{j=-1}^{1} P_{i,j} = 1$. Expression (7) shows that the linear functions (6) completely determine the conditional probabilities of the match outcomes which are sufficient for describing the conditional probability mass function of y_i (see Appendix A).

The parameters of the multinomial logit model can conveniently be estimated by the method of maximum likelihood (ML). The ML estimator is obtained by maximizing the log-likelihood function derived in Appendix A by numerical methods. The method of maximum likelihood also facilitates to use the conventional likelihood-based statistical test statistics when examining the market efficiency test proposed in the next section.

3.2   Efficiency Test

In previous research, different statistical tests and related procedures have been considered when evaluating the betting market efficiency. Most of the previous studies have concentrated on the National Football League (NFL, American football) games (see, e.g., Pankoff, 1968; Gandar *et al.*, 1988; Sauer *et al.*, 1988; Golec and Tamarkin, 1991; Gray and Gray, 1997). In the case of association football, Pope and Peel (1989), Goddard and Asimakopoulos (2004) and Koning (2012) have examined separate efficiency tests for each match outcome $y_i = j$. A modified version of their testing approach will be considered in Section 3.3 as an alternative to the multinomial logit-based test presented below.

If the bookmakers are setting their odds (in statistical sense) efficiently, the average implicit probabilities (3) should be informationally efficient (unbiased) predictors of the match outcomes (assuming approximately uniformly distributed betting volumes). This hypothesis can be tested by using a restricted multinomial logit model where the linear functions (6) are determined as

$$\pi_{i,1} = \bar{\pi}_{i,1}^{ip}, \quad \pi_{i,-1} = \bar{\pi}_{i,-1}^{ip}. \tag{8}$$

In other words, if the null hypothesis

$$H_0 : \alpha_j = 0, \, \beta_j = 1, \quad j = -1, 1, \tag{9}$$

imposed on the unrestricted model (6) cannot be rejected, the implicit probabilities (3) are statistically unbiased and informationally efficient predictors. If the null hypothesis (9) is rejected, then the match results deviate systematically from the im-

plicit probabilities and it would be possible to get more accurate probability forecasts
(7) using the unrestricted model.

The rejection of the hypothesis (9) also implies that some additional predictive
power can possibly be obtained by augmenting the unrestricted model with some
predictive variables. These variables may reflect, for example, the teams recent and
long-term performance (past match results), the importance of the match for cham-
pionship, promotion or relegation at the end of the season, F.A. Cup involvement or
geographical distance (see Goddard and Asimakopoulos, 2004; Forrest *et al.*, 2005).
This possible extension is, however, out of the scope of this study (as we are interested
in the implicit probabilities (3)) and it is left for the future research.

In Appendix A, we show how the classical likelihood-based test statistics, i.e.
Likelihood ratio (LR), Wald (W) and Lagrange multiplier (LM) can be used to test
the hypothesis (9). Although the tests are asymptotically equivalent (under the
correctly specified model), their small-sample properties may differ. In Section 4.2,
we consider Monte Carlo simulation experiments to examine the empirical size and
power of the test statistics.

3.3 Comparison to Alternative Models and Tests

Instead of the multinomial logit model, Pope and Peel (1989), Goddard and Asi-
makopoulos (2004) and Koning (2012), among others, have used linear probability
and binary logit models to test the (weak-form) market efficiency separately for each
outcome $y_{i,j}$. We concentrate on the former i.e. independent binary logit models.
Koning (2012) formulated the logit model as

$$P(y_{i,j} = 1 | \bar{\mathbf{P}}_i^{ip}) = \frac{1}{1 + \exp(-a_j - b_j \log(1/\bar{P}_i^{ip,j} - 1))}, \quad j = -1, 0, 1, \qquad (10)$$

where the implicit probability $\bar{P}_i^{ip,j}$ (see (2)) is used as a single predictor for the
outcome $y_{i,j}$. The (weak-form) market efficiency condition is hence

$$H_0 : a_j = 0, \, b_j = -1, \quad j = -1, 0, 1. \qquad (11)$$

Under this hypothesis, $\bar{P}_i^{ip,j}$ is an unbiased predictor of $y_{i,j}$ in model (10). In the
linear probability model (see, e.g., Greene 2012, p. 727), the test itself is similar as

above but the parameters are estimated by using the Ordinary Least Squares instead of the method of maximum likelihood employed in the logit model.

The main difference between this outcome-specific testing procedure and our approach in Section 3.2 is that the hypothesis (11) must be examined for each match outcome separately. This leads to several potential disadvantages. As the logit models (10) are (implicitly) assumed to be independent, this approach does not take into account the restriction that one of the possible outcomes will occur for each match i. Hence, the resulting probabilities (10) may not sum up to one. Another potentially more fundamental difficulty is that there is a different value of the test statistic for each j and, thus, the conclusions on the market efficiency may differ for different match outcomes. Both of these issues are avoided in the multinomial logit model as the response probabilities (7) are constructed simultaneously within one model leading to a single test statistic instead of three obtained with the binary logit models (10).

In Section 4, we compare the outcome-specific and the multinomial logit-based testing approaches using simulation experiments. To facilitate that, we have to first determine how to combine the information of the independent logit models and their respective test statistics. Pope and Peel (1989), Goddard and Asimakopoulos (2004) and Koning (2012), among others, have not explicitly specified when their testing procedure imply enough evidence against the efficient market hypothesis. Therefore, as the logit models are treated independent, we consider a simple and intuitively plausible approach (see details at Appendix B) where the total value of the log-likelihood function is the sum of the independent logit models. A joint test for the hypothesis (11) for all $j = -1, 0, 1$, can thus be based on the Likelihood ratio test in the usual way.

In addition to the binary logit and linear probability models, Forrest et $al.$ (2005) have used the ordered probit model to evaluate bookmakers' implicit probabilities as predictors for the match outcomes (see also Koning, 2000; Goddard and Asimakopoulos, 2004; Goddard, 2005, and the references therein). However, their main interest was in the predictive power of different covariates including a transformed predictor

constructed from the implicit probabilities (3). In particular, their model cannot be used as a direct statistical test of the informational efficiency of the implicit probabilities. The reason is that the ordered logit and probit models (see details at Greene 2012, p. 827–829) are based on a single latent linear index and the same parameter vector for all the outcomes $y_i = j$ (cf. the linear functions (6)). This does not enable to construct an unambiguous hypothesis similar to (9) which can be tested directly with the multinomial logit model. This is also the case with goal-based count data models (see the references in Introduction).

4 Empirical Results

4.1 Dataset

In our application, the dataset comprises the results and betting odds offered by different bookmakers for the football matches played in the top four English football league divisions. The sample period contains the results of the past 13 football seasons between 2000–2001 and 2012–2013. English football matches have been the most commonly used source of data in the previous literature (see, e.g., Dixon and Coles, 1997; Cain *et al.* 2000; Rue and Salvesen, 2000; Dixon and Pope, 2004; Goddard, 2005; Forrest *et al.* 2005; Koopman and Lit, 2014). Our testing approach can, of course, easily be extended to other leagues and international datasets (see, e.g., Vlastakis *et al.*, 2009; Franck *et al.*, 2010; Koning, 2012; Direr, 2013).

Our data covers the matches played in the English Premier League (denoted by PL, the highest tier in English football) and The Football League consisting of The Championship (CH), League One (L1) and League Two (L2). Like Franck *et al.* (2010), Koning (2012) and Direr (2013), the source of all data is Football-Data.co.uk where the data can be downloaded for free. In the raw data, the odds offered by five different bookmakers are recorded for the first season 2000–2001 and the number of bookmakers increases gradually about ten for more recent seasons. Odds for weekend games are collected Friday afternoons, and on Tuesday afternoons for midweek games. The odds of several competing bookmakers allow us to construct the average implicit

probabilities (3) using the procedures explained in Section 2.

The basic descriptive statistics of the dataset are presented in Table 1. The number of teams in the Premier League is 20 while 24 in other leagues indicating that there are 380 and 552 matches played in one season, respectively. The number of observations in the sample for different leagues is thus about 7 000 except for the Premier League (N=4 940) making it total N=26 463. A few matches have been withdrawn (about ten) from the sample as there were no recorded betting odds for those games.

It appears that the Premier League differs somewhat from other leagues as the share of home wins is higher which equivalently means that there are more draws and away wins in the lower leagues. Although the differences are rather small, this and the fact that there is typically much more publicly available information on the Premier League teams (like information on the key players, their possible injuries as well as other important factors affecting the teams' performance levels), it is reasonable to carry out statistical analysis for the full dataset but also separately for each league.

4.2   Simulation Study

In this section, the small-sample properties of the efficiency test based on the multinomial logit model (Sections 3.1 and 3.2) are examined by simulation experiments. In particular, we consider the size and power properties of the likelihood-based test statistics introduced more detail in Appendix A. We will also make comparisons to the outcome-specific testing procedure considered in Section 3.3 and Appendix B. Throughout this study, the maximum likelihood estimates of the parameters are obtained by the BHHH optimization algorithm implemented via the cmlmt library in Gauss (version 10).

In size simulations, we employ a bootstrap resampling method where the full sample (all leagues) introduced in Section 4.1 is used as a sampling population where the bootstrap samples obtained with replacement are generated. Under the null hypothesis (9), the data generating process (DGP) implies that the implicit probabilities (3) are efficient predictors of the match outcomes. Therefore, we generate bootstrap

samples of size N_B from that population and simulate the match outcomes y_b from the corresponding conditional multinomial distribution (i.e. $y_b | \bar{\mathbf{P}}_b^{ip}$, $b = 1, \ldots, N_B$). Using the simulated bootstrap sample, after the parameters of the multinomial and binary logit models have been estimated, the values of the test statistics and their p-values based on the asymptotic χ^2-distribution are computed. This procedure is repeated 10 000 times.

In Table 3, we report the empirical size of the test statistics. In addition to the very small bootstrap sample size of N_B=100, N_B=500 and N_B=2 000 correspond approximately the number of matches played during one season in one league and four leagues altogether, respectively. The last selection (N_B=10 000) is used to describe the asymptotic properties of the tests. The empirical rejection rates are reported at the 10%, 5% and 1% statistical significance levels. We also present the rejection rates of the testing procedure relying on the independent binary logit models and the LR_{sum}^{IL} test (see Section 3.3 and Appendix B).

Table 3 shows that the empirical size of all three test statistics (LR, W and LM) are close to the nominal levels when the efficiency testing is based on the multinomial logit model. In other words, the asymptotic critical values coming from the χ^2-distribution turn out to be rather accurate. This is the case even with the smallest sample sizes considered. Above all, the multinomial logit-based test statistics outperform the test based on the independent logit models. The latter test was found to suffer from size distortions being clearly oversized demonstrating the importance of using the multinomial logit model.

In the past research, various authors have argued that the power of the previously considered tests has generally been rather low. Most of these studies have concentrated on the NFL games (see, e.g., Gandar *et al.*, 1988; Golec and Tamarkin, 1991; Gray and Gray, 1997) but similar conclusions have also been made in the case of association football (see, e.g., Pope and Peel, 1989; Goddard and Asimakopoulos, 2004). To the best of our knowledge, this is the first study to consider the power of the tests using Monte Carlo simulation experiments. Overall, likely due to the low power, a general conclusion in the past research has typically been that the statistical tests

cannot reject the weak-form market efficiency. At the same time economic tests, such as different betting rules, show that it is possible to form profitable betting strategies using the predictive power provided by the statistical models which is against the idea of the efficient market hypothesis.

In power simulations, we need to first obtain informationally efficient probabilities for different match outcomes. To this end, we use the estimated parameters presented in Table 2 as the DGP parameters (the results of Table 2 will be discussed more detail in Section 4.3). Next, we simulate bootstrap samples where the constructed (efficient) DGP probabilities (based on (3)) are used to generate match results. Following the linear functions (6), as we are examining the efficiency of the implicit probabilities (3), those are used as predictors in the multinomial logit model to predict simulated match results. As in power simulations the implicit probabilities (3) are expected to deviate from the DGP probabilities, we should reject the null hypothesis (9).

Because the estimation results in Table 2 are somewhat different for different leagues, we consider two Monte Carlo experiments. In the first one, the DGP parameters are based on the estimation results of the Premier League (PL) games (third column in Table 2) where the deviations from the parameter values implied by the null hypothesis (9) are rather large. In the second experiment, the DGP parameters are the ones obtained for The Championship (CH, 4th column in Table 2) which are closer to the null hypothesis (9) than in the first experiment. Thus, this latter one can be seen as a simulation experiment where the market is only marginally inefficient.

The power of different test statistics reported in Tables 4 and 5 is studied by the same sample sizes and statistical significance levels as in size simulations. To facilitate a comparison between the multinomial logit and the outcome-specific testing approach, we consider size-adjusted power where the critical values are obtained from the size simulations. In the multinomial logit-based tests, the power is almost exactly the same irrespective of the selection of critical values (asymptotic or bootstrap) while in the latter case bootstrap critical values are clearly needed as LR_{sum}^{IL} was found to be oversized.

The (size-adjusted) power of the tests appears to be closely dependent on the

DGP and the sample size. When the degree of inefficiency is large (Table 4), the tests have high power, especially when the sample size is rather large (at least 2 000 observations). However, in Table 5 the power is generally much lower when the DGP parameters are closer to the values implied by the null hypothesis (9). In this latter case, the sample size has to be very large before the tests have reasonable power. When the sample size is rather large, then the tests based on the multinomial logit model are substantially more powerful than the outcome-specific tests. Overall, in the previous literature different authors have used much smaller sample sizes than, say, 2 000 observations in their analyses. Given the simulation evidence obtained here, it is not hence surprising that most of the previous tests cannot statistically reject market efficiency.

When comparing the three classical test statistics, the Lagrange multiplier (LM) test turns out to be the most powerful test. The Likelihood ratio (LR) and the Wald (W) tests have also reasonable power without large differences between them. The LR test based on the independent logit models (LR_{sum}^{IL}) is also rather powerful but the above-mentioned LM test based on the multinomial logit model is still even more powerful also in this comparison.

Overall, we can thus conclude that when the objective is to test the informational efficiency of the implicit probabilities as a whole (the outcome-specific testing can be the objective in other type of analysis), the LM test based on the multinomial logit model is the best one in terms of size and power properties. As the construction of the LM test requires only the restricted model (8), the possible uncertainty coming from the parameter estimation is hence circumvented. This is one of the possible explanations for its superior performance over the alternative test statistics.

4.3 Efficiency in English Football Betting Market

We test the informational efficiency of the English football betting market in several steps. First, we estimate the multinomial logit models and construct the values of the test statistics for the full sample and separately for each league. The estimated parameter coefficients presented in Table 2 have already been used in simulations in

Section 4.2. Later on, we will extend the analysis by using the observations of the matches played in different parts of the football season and separately for different seasons. As already emphasized in Sections 2 and 3.2, we will concentrate on statistical efficiency tests and, for example, the possibilities to construct profitable betting strategies are left for the future research.

Table 2 presents the estimated parameter coefficients of the multinomial logit model using the full sample and league-specific subsets of the dataset. The goodness-of-fit of the models is evaluated in terms of the pseudo-R^2 of Estrella (1998) and the Brier score. The pseudo-R^2 can be written as

$$\text{pseudo} - R^2 = 1 - \Big(\frac{l_u}{l_c}\Big)^{-(2/N)l_c}, \tag{12}$$

where l_u is the maximum value of the estimated unconstrained log-likelihood function (see Appendix A) and l_c is its constrained counterpart in a restricted model where only an intercept term is included. Estrella (1998) showed that (12) has a closer connection to the interpretation of the coefficient of determination ("R^2") of linear models than the alternative pseudo-R^2 expressions. Furthermore, the Brier score for the outcome $y_{i,j}$ is

$$\text{Brier}_j = \frac{1}{N} \sum_{i=1}^{N} (y_{i,j} - P_{i,j})^2, \quad j = -1, 0, 1. \tag{13}$$

The smaller the Brier score, the more accurate are the probability forecasts $P_{i,j}$ constructed in (7)).

The testing results in Table 2 show that the Premier League (PL) appears to differ from other leagues. In the former, the null hypothesis (9) is rejected at all the conventional significance levels whereas the p-values are rather high for other leagues. In other words, except the Premier League, we cannot statistically reject the weak-form market efficiency. The effect of the Premier League games to the full sample results can be seen in the second column of Table 2 where the p-values are smaller than 1%. The results of the three classical test statistics (LR,W,LM) are essentially the same which is consistent with the size simulation results reported in Section 4.2.

In line with the testing results, the estimated parameter coefficients show large deviations from the efficiency hypothesis (9) predominantly only for the Premier

League. Interestingly, the multinomial logit model can predict the Premier league games much more accurately than the matches played in other leagues. This can be seen as higher (smaller) values of the pseudo-R^2 (Brier score) and more accurately estimated β_j coefficients while in other leagues those are much closer to the ones (i.e. $\beta_j = 1$, $j = -1, 1$) implied by the null hypothesis (9). According to the simulation results presented in Section 4.2, the power is expected to be rather low when the degree of inefficiency is small. However, the values of the test statistics are now so low that we can rather safely conclude that there is not statistically significant evidence against the efficient market hypothesis for The Championship (CH), League One (L1) and League Two (L2).

Similarly as in Koning (2012) (see also the linear probability models employed by Pope and Peel (1989) and Goddard and Asimakopoulos (2004)), Table 6 presents the estimation and efficiency testing results of the independent binary logit models. The reported Likelihood ratio tests (LR_1^{IL}, LR_0^{IL} and LR_{-1}^{IL}) are carried out separately for each match outcome $y_{i,j}$ when testing the hypothesis (11). The results of the joint LR test (LR_{sum}^{IL}) are reported in the lower panel.

As a whole, the results in Table 6 are essentially the same as obtained with the multinomial logit model in Table 2. In accordance with the outcome-specific results of Koning (2012), there is statistically significant inefficiency in the odds of the Premier League based on the separate (although for the draw outcome ($y_i = 0$) the p-value is 0.10) and the joint test (LR_{sum}^{IL}). For other leagues, the p-values of the separate and the joint LR tests are rather large without notable differences between different match outcomes showing that the efficient market hypothesis cannot be rejected in those cases.

Table 7 reports the results of the matches played in four different parts of the football season. The classification of months is the same as in Goddard and Asimakopoulos (2004). We report the results of the multinomial logit-based LM test (the most powerful test in Section 4.2) but the results of the Likelihood ratio and Wald tests were essentially the same (available upon request). According to their outcome-specific statistical and economic tests, Goddard and Asimakopoulos (2004)

concluded that the market efficiency is typically strongest at the mid and partly also at the beginning of the season. We find similar results with our test. In fact, the efficient market hypothesis cannot be rejected for any league, not even in the case of the Premier League, for the games played between August and December. However, for the second half of the season (January–May), the p-values for the Premier League are smaller than 1%. There is also a similar tendency in other leagues: the values of the test statistics are higher for the second half of the season although the p-values are not as small as for the Premier League. In particular, we can reject market efficiency for the matches played between March and May also for The Championship and League Two at the 5% and 10% significance levels, respectively.

We also made comparisons between different football seasons in a similar way as in Goddard (2005) and Forrest *et al.* (2005). Details on the estimation and efficiency testing results are available upon request. In general, in contrast to the stage of the season (Table 7), there were no large and systematic differences in the market efficiency between different seasons.

In the previous research, substantial amount of attention has also been paid on the various stylized empirical regularities of the betting odds. The most notable one is the longshot-bias (favourite-longshot bias) implying that the bets placed on favourites yield higher returns than the bets for the longshots (underdogs). This phenomenon has been identified in a variety of sports betting markets including association football (see, e.g, Cain *et al.*, 2000; Koning, (2012). On the contrary to these findings, Forrest *et al.* (2005) have found only scant evidence and Dixon and Pope (2004) even a reverse longshot-bias relationship in the association football betting market. Vlastakis *et al.* (2009) confirmed the existence of the longshot-bias and they find also an "away-favourite" bias (i.e. higher betting market returns when betting on away favourites) as a result of overestimation of the home field advantage and the longshot-bias.

To examine these regularities, in Figures 1 and 2, we compare the implicit probabilities (3) and the predicted probabilities (solid line) from the multinomial logit model (7) whose estimation results are reported in Table 2. The depicted solid lines

are based on the robust loess curves obtained with Matlab function *smooth* (see also Koning, 2012). The dashed 45-degree line corresponds the case where the implicit probabilities are efficient, that is the null hypothesis (9) cannot be rejected. Figure 1 depicts the results for the Premier League and The Championship while Figure 2 consists of the League One and League Two.

In accordance with the results presented in Table 2, the largest deviations from the 45-degree line are obtained with the Premier League data. In this case, we can confirm the findings of Vlastakis *et al.* (2009) that there is evidence of longshot-bias as well as away-favourite bias. In other words, strong home and away favourites will win their games more often than expected based on the implicit probabilities. The evidence for other leagues is somewhat similar but there the deviations from the 45-degree line are not as large as in the Premier League games.

5   Conclusions

In this study, we have proposed a new statistical test for the informational efficiency in the association football betting market. The test is based on the multinomial logit model where parameter estimation can be carried out using the method of maximum likelihood facilitating the use of the classical likelihood-based test statistics for the efficient market hypothesis. We concentrate on the betting market efficiency in the whole market level but the proposed test can also be applied as a rationality check of the probability forecasts of the association football matches provided by any single bookmaker or bettor.

It appears that this is the first study to consider the properties of the statistical football betting market efficiency tests using Monte Carlo experiments. Simulation results show that the proposed test has reasonable empirical size, even in relatively small samples, while the power is dependent on the sample size and the degree of inefficiency. If the inefficiency in the betting odds is large, then the power is high, especially for rather large sample sizes. In the opposite case of only marginally inefficient odds, the sample size has to be very large (thousands of observations) before the test has some power. We also find that independent of the considered

simulation experiment the Lagrange multiplier (LM) test is the most powerful among the alternative test statistics. Overall, the proposed multinomial logit-based LM test outperforms the previously considered tests built on separate binary logit (or linear probability) models for each match outcome used in the previous research.

In our empirical application to English football data, there is statistically significant evidence of market inefficiency only in the English Premier League (the highest league division). In the lower leagues, the efficient market hypothesis cannot be rejected at the conventional statistical significance levels. As a whole, the results indicate that the betting odds are typically informationally more efficient at the beginning of the season while the degree of inefficiency increases when coming close to the end of the season.

Appendix A: Log-likelihood Function and Trinity of Likelihood-Based Test Statistics

Following the notation employed in Section 3.1, the outcome of an association football match is either a home win ($y_i = 1$), draw ($y_i = 0$) or an away win ($y_i = -1$). The corresponding conditional probabilities $P_{i,j}$ obtained with the multinomial logit model, conditional on the implicit probabilities (3), are constructed in (7). Using the binary indicator variables $y_{i,j}$, $j = -1, 0, 1$, introduced in Section 3.1 the conditional likelihood function can be written

$$L(\boldsymbol{\theta}) \;=\; \prod_{i=1}^{N} \prod_{j=-1}^{1} P_{i,j}^{y_{i,j}},$$

where the parameters of the model are included in the vector $\boldsymbol{\theta} = \begin{bmatrix} \alpha_1 & \alpha_{-1} & \beta_1 & \beta_{-1} \end{bmatrix}'$ and N is the sample size. The log-likelihood function is hence

$$l(\boldsymbol{\theta}) \;=\; \log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} l_i(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j=-1}^{1} y_{i,j} \log(P_{i,j}),$$

where $y_{i,j}$ is equal to one for the realized outcome $y_i = j$ and zero otherwise. The ML estimate is obtained by maximizing the log-likelihood function $l(\boldsymbol{\theta})$ using numerical methods. Under reasonable regularity conditions, the ML estimator $\widehat{\boldsymbol{\theta}}$ is consistent

and asymptotically normally distributed

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, I(\boldsymbol{\theta}_0)^{-1}),$$

where the true (DGP) value of the parameter vector is denoted by $\boldsymbol{\theta}_0$,

The null hypothesis (9) can be written as

$$A\boldsymbol{\theta} = \mathbf{c},$$

where $A = I_4$ and $\mathbf{c} = \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix}'$. The Wald test statistic is

$$W = (A\widehat{\boldsymbol{\theta}} - \mathbf{c})'[AJ(\widehat{\boldsymbol{\theta}})^{-1}A']^{-1}(A\widehat{\boldsymbol{\theta}} - \mathbf{c}),$$

where $J(\widehat{\boldsymbol{\theta}})$ is a consistent estimate of $I(\boldsymbol{\theta}_0)$ (estimated here by using the negative numerical Hessian). The Wald as well as the LR and LM tests below follow an asymptotic χ_4^2-distribution under the null hypothesis (9).

The Likelihood ratio (LR) test is

$$LR = 2\big(l(\widehat{\boldsymbol{\theta}}) - l(\widetilde{\boldsymbol{\theta}})\big),$$

where $l(\widehat{\boldsymbol{\theta}})$ and $l(\widetilde{\boldsymbol{\theta}})$ denote the values of the log-likelihood function given above evaluated at the unrestricted $\widehat{\boldsymbol{\theta}}$ and the restricted ML estimates $\widetilde{\boldsymbol{\theta}}$ (under the null hypothesis (9)), respectively.

The Lagrange multiplier (LM) test is based on the score of the log-likelihood function

$$s_i(\boldsymbol{\theta}) \;\; = \;\; \frac{\partial}{\partial \boldsymbol{\theta}} l_i(\boldsymbol{\theta}) = \Big[\frac{\partial l_i(\boldsymbol{\theta})}{\partial \alpha_1} \quad \frac{\partial l_i(\boldsymbol{\theta})}{\partial \alpha_{-1}} \quad \frac{\partial l_i(\boldsymbol{\theta})}{\partial \beta_1} \quad \frac{\partial l_i(\boldsymbol{\theta})}{\partial \beta_{-1}} \Big]', \quad i = 1, \ldots, N.$$

To derive the LM test statistic, we use the log-odds ratios (4) to rewrite the log-likelihood function

$$
\begin{aligned}
l(\boldsymbol{\theta}) \;\; &= \;\; \sum_{i=1}^{N} l_i(\boldsymbol{\theta}) = \log\Big[\prod_{i=1}^{N} P_{i,1}^{y_{i,1}} P_{i,-1}^{y_{i,-1}} P_{i,0}^{1-y_{i,1}-y_{i,-1}} \Big] \\
&= \;\; \log\Big[\prod_{i=1}^{N} \Big(\frac{P_{i,1}}{P_{i,0}}\Big)^{y_{i,1}} \Big(\frac{P_{i,-1}}{P_{i,0}}\Big)^{y_{i,-1}} P_{i,0} \Big] \\
&= \;\; \sum_{i=1}^{N} y_{i,1}\pi_{i,1} + y_{i,-1}\pi_{i,-1} - \log\big(1 + \exp(\pi_{i,1}) + \exp(\pi_{i,-1})\big),
\end{aligned}
$$

where $\pi_{i,j}, j = -1, 1$ are given in (6). The score with respect to α_j is

$$
\begin{aligned}
\frac{\partial l_i(\boldsymbol{\theta})}{\partial \alpha_j} &= y_{i,j}\left(\frac{\partial}{\partial \alpha_j}\pi_{i,j}\right) - \frac{\partial}{\partial \alpha_j}\log\left(1 + \exp(\pi_{i,1}) + \exp(\pi_{i,-1})\right) \\
&= y_{i,j}\left(\frac{\partial}{\partial \alpha_j}\pi_{i,j}\right) - \frac{1}{1 + \exp(\pi_{i,1}) + \exp(\pi_{i,-1})}\exp(\pi_{i,1})\left(\frac{\partial}{\partial \alpha_j}\pi_{i,j}\right) \\
&= \left(y_{i,j} - P_{i,j}\right)\left(\frac{\partial}{\partial \alpha_j}\pi_{i,j}\right) \\
&= y_{i,j} - P_{i,j}, \quad j = -1, 1.
\end{aligned}
$$

Similarly, we get (under the null hypothesis (9))

$$
\begin{aligned}
\frac{\partial l_i(\boldsymbol{\theta})}{\partial \beta_j} &= \left(y_{i,j} - P_{i,j}\right)\left(\frac{\partial}{\partial \beta_j}\pi_{i,j}\right) \\
&= \left(y_{i,j} - P_{i,j}\right)\bar{\pi}_{i,j}^{ip}, \quad j = -1, 1,
\end{aligned}
$$

where $\pi_{i,j}^{ip,j}$ and $P_i^{ip,j}$ are given in (6) and (7), respectively. The LM test statistic can thus be written (see, e.g., Davidson and MacKinnon, 1984)

$$
\begin{aligned}
LM &= \boldsymbol{\iota}'S(\widetilde{\boldsymbol{\theta}})\left(S(\widetilde{\boldsymbol{\theta}})'S(\widetilde{\boldsymbol{\theta}})\right)^{-1}S(\widetilde{\boldsymbol{\theta}})'\boldsymbol{\iota} \\
&= \sum_{i=1}^{N} s_i(\widetilde{\boldsymbol{\theta}})'\left(S(\widetilde{\boldsymbol{\theta}})'S(\widetilde{\boldsymbol{\theta}})\right)^{-1}\sum_{i=1}^{N} s_i(\widetilde{\boldsymbol{\theta}}),
\end{aligned}
$$

where $\widetilde{\boldsymbol{\theta}}$ is the restricted ML estimate, $\boldsymbol{\iota}$ is $(N \times 1)$ vector of ones and the matrix $S(\widetilde{\boldsymbol{\theta}})$ is given by

$$
S(\widetilde{\boldsymbol{\theta}}) = \begin{bmatrix} s_1'(\widetilde{\boldsymbol{\theta}}) & s_2'(\widetilde{\boldsymbol{\theta}}) & \dots & s_N'(\widetilde{\boldsymbol{\theta}}) \end{bmatrix}'.
$$

Under the null hypothesis (9) (see also (3),

$$
s_i(\widetilde{\boldsymbol{\theta}}) = \begin{bmatrix} y_{i,1} - \bar{P}_i^{ip,1} & y_{i,-1} - \bar{P}_i^{ip,-1} & \left(y_{i,1} - \bar{P}_i^{ip,1}\right)\bar{\pi}_{i,1}^{ip} & \left(y_{i,-1} - \bar{P}_i^{ip,-1}\right)\bar{\pi}_{i,-1}^{ip} \end{bmatrix}'.
$$

Appendix B: Likelihood Ratio Efficiency Tests in Independent Logit Models

The log-likelihood function of the binary logit model (10) for the outcome j is

$$
l_j(\boldsymbol{\varphi}_j) = \sum_{i=1}^{N} y_{i,j}\log\left(P(y_{i,j} = 1|\bar{\mathbf{P}}_i^{ip})\right) + (1 - y_{i,j})\log\left(1 - P(y_{i,j} = 1|\bar{\mathbf{P}}_i^{ip})\right),
$$

where the outcome-specific parameters are included in the vector $\boldsymbol{\varphi}_j, j = -1, 0, 1$. The conditional probability of $y_{i,j} = 1$, conditional on the implicit probabilities (3), is denoted by $P(y_{i,j} = 1 | \bar{\mathbf{P}}_i^{ip})$. The ML estimate $\widehat{\boldsymbol{\varphi}}_j$ is obtained by maximizing the above log-likelihood function using numerical methods. To test the null hypothesis (11) for each j, the Likelihood ratio (LR) test for the component j is

$$LR_j^{IL} = 2\big(l_j(\widehat{\boldsymbol{\varphi}}_j) - l_j(\widetilde{\boldsymbol{\varphi}}_j)\big), \quad j = -1, 0, 1,$$

where $\widehat{\boldsymbol{\varphi}}_j$ and $\widetilde{\boldsymbol{\varphi}}_j$ denote the unrestricted and restricted (under the null hypothesis (11)) ML estimates and IL denotes an independent logit model. Under the null hypothesis (11), LR_j^{IL} follows an asymptotic χ_2^2-distribution.

As in this outcome-specific testing procedure the logit models (10) are treated independent, a joint test for the market efficiency can be constructed using the LR test

$$LR_{sum}^{IL} = 2\Big(\sum_{j=-1}^{1} l_j(\widehat{\boldsymbol{\varphi}}_j) - \sum_{j=-1}^{1} l_j(\widetilde{\boldsymbol{\varphi}}_j)\Big).$$

In other words, LR_{sum}^{IL} is based on the sum of the log-likelihood functions evaluated at the restricted and unrestricted ML estimates obtained for the independent logit models. Compared with the single component tests ($LR_j^{IL}, j = -1, 0, 1$), in this case we have six restrictions to be tested and hence LR_{sum}^{IL} follows an asymptotic χ_6^2-distribution under the null hypothesis.

References

[1] Akhtar, S. and Scarf, P. (2012) Forecasting test cricket match outcomes in play. *International Journal of Forecasting*, **28**, 632–643.

[2] Bastos, L.S., and da Rosa, J.M.C (2013) Predicting probabilities for the 2010 FIFA World Cup games using a Poisson-Gamma model. *Journal of Applied Statistics*, **40**, 1533–1544.

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

[4] Crowder, M., Dixon, M., Ledford, A. and Robinson, M. (2002) Dynamic modelling and prediction of English football league matches for betting. *Journal of the Royal Statistical Society, Series D (Statistician)*, **51**, 157–168.

[5] Davidson, R. and MacKinnon, J. (1984) Convenient specification tests for logit and probit models. *Journal of Econometrics*, **25**, 241–262.

[6] Direr, A. (2013) Are betting markets efficient? Evidence from European football championships. *Applied Economics*, **45**, 343–356.

[7] Dixon, M.J, and Coles, S.G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **46**, 265–280.

[8] Dixon, M. J. and Pope, P.F. (2004) The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting*, **20**, 697–711.

[9] Estrella, A. (1998) A new measure of fit for equations with dichotomous dependent variables. *Journal of Business and Economic Statistics*, **16**, 198–205.

[10] Figlewski, S. (1979) Subjective information and market efficiency in a betting market. *Journal of Political Economy*, **87**, 75–88.

[11] Forrest, D. and Simmons, R. (2000) Forecasting sport: the behaviour and performance of football tipsters. *International Journal of Forecasting*, **16**, 317–331.

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

[13] Franck, E., Verbeek, E. and Nuesch, S. (2010) Prediction accuracy of different market structures – bookmakers versus a betting exchange. *International Journal of Forecasting*, **26**, 448–459.

[14] Gandar, J., Zuber, R., O'Brien, T. and Russo, B. (1988) Testing rationality in the point spread betting market. *Journal of Finance*, **43**, 995–1008.

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

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

[17] Golec, J. and Tamarkin, M. (1991) The degree of inefficiency in the football betting market: Statistical tests. *Journal of Financial Economics*, **30**, 311–323.

[18] Gray, P.K. and Gray, S. F. (1997) Testing market efficiency: Evidence from the NFL sports betting market. *Journal of Finance*, **52**, 1725–1737.

[19] Greene, W. (2012) *Econometric Analysis.* 7th International Edition. Pearson Education.

[20] Karlis, D. and Ntzoufras, I. (2003) Analysis of sports data by using bivariate Poisson models.*Journal of the Royal Statistical Society, Series D (Statistician)*, **52**, 381-393.

[21] Koning, R.H. (2000) Balance in competition in Dutch soccer. *Journal of the Royal Statistical Society, Series D (The Statistican)*, **49**, 419–431.

[22] Koning, R.H. (2012) Regression tests and the efficiency of fixed odds betting markets. *International Journal of Sport Finance*, **7**, 262–274.

[23] Koopman, S.J. and Lit, R. (2014) A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *Journal of the Royal Statistical Society, Series A*, forthcoming.

[24] Levitt, S. (2004) Why are gambling markets organized so differently from financial markets? *Economic Journal*, **114**, 223–246.

[25] Lim, K-P. and Brooks, R. (2011) The evolution of stock market efficiency over time: A survey of the empirical literature. *Journal of Economic Surveys*, **25**, 69–108.

[26] Maher, M. J. (1982) Modeling association football scores. *Statistica Neerlandica*, **36**, 109–118.

[27] Malkiel, B.G. (2003) The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, **17**, 59–82.

[28] Pankoff, L.D. (1968) Market efficiency and football betting. *Journal of Business*, **41**, 203–214.

[29] Pope, P.F. and Peel, D.A. (1989) Information, prices and efficiency in a fixed-odds betting market. *Economica*, **56**, 323–341.

[30] Rue, H. and Salvesen, O. (2000) Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society, Series D (The Statistican)*, **49**, 399–418.

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

[32] Skinner, G.K., and Freeman, G.H. (2009) Soccer matches as experiments: how often does the 'best' team win? *Journal of Applied Statistics*, **36**, 1087–1095.

[33] Summers, L.H. (1986) Does the stock market rationally reflect fundamental values? *Journal of Finance*, **41**, 591–601.

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

[35] Vlastakis, N., Dotsis, G. and Markellos, R.N. (2009) How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting*, **28**, 426–444.

[36] Zuber, R.A., Gandar, J.M. and Bowers, B.D. (1985) Beating the spread: Testing the efficiency of the gambling market for NFL games. *Journal of Political Economy*, **92**, 800–806.

Table 1: The descriptive statistics of the dataset. The data is obtained from the Football-Data.co.uk (www.football-data-co.uk) and it covers the matches played between the seasons 2000–2001 and 2012–2013. The full sample in the first row contains all the leagues whereas PL, CH, L1 and L2 denote the Premier League, The Championship, League One and League Two, respectively.

	Observations (N)	Match outcomes, relative frequencies		
		Home wins	Draws	Away wins
Full sample	26463	0.446	0.272	0.282
PL	4940	0.467	0.261	0.272
CH	7172	0.444	0.274	0.282
L1	7175	0.442	0.273	0.285
L2	7176	0.436	0.277	0.287

Table 2: Estimation and market efficiency testing results based on the multinomial logit model. The numbers in the parentheses are the standard errors of the estimated coefficients. The p-values of the efficiency test statistics (see Appendix A) are based on the asymptotic χ_4^2-distribution.

	Full sample	PL	CH	L1	L2
α_1	-0.052	-0.029	0.015	-0.066	-0.094
	(0.029)	(0.054)	(0.062)	(0.062)	(0.064)
β_1	1.107	1.128	0.971	1.128	1.152
	(0.051)	(0.079)	(0.117)	(0.114)	(0.124)
α_{-1}	-0.033	-0.110	-0.014	-0.021	-0.015
	(0.017)	(0.042)	(0.032)	(0.033)	(0.032)
β_{-1}	1.119	1.205	1.156	1.170	0.930
	(0.055)	(0.090)	(0.121)	(0.121)	(0.127)
Observations (N)	26463	4940	7172	7175	7176
log-L	-27372.16	-4801.59	-7511.74	-7484.81	-7566.30
pseudo$-R^2$	0.073	0.169	0.049	0.060	0.043
Brier$_1$	0.233	0.216	0.238	0.236	0.237
Brier$_0$	0.198	0.191	0.199	0.198	0.200
Brier$_{-1}$	0.191	0.172	0.194	0.194	0.198
LR	20.326	24.874	2.030	6.676	2.169
(p-value)	(0.004)	(0.000)	(0.730)	(0.154)	(0.704)
W	19.929	23.620	2.010	6.572	2.169
(p-value)	(0.005)	(0.000)	(0.734)	(0.160)	(0.705)
LM	21.320	27.665	2.057	6.940	2.186
(p-value)	(0.003)	(0.000)	(0.725)	(0.139)	(0.702)

Table 3: Empirical size of the test statistics at the 10%, 5% and 1% nominal significance levels (10 000 simulated replications). The bootstrap sample size is denoted by N_B whereas LR_{sum}^{IL} denotes the joint Likelihood ratio test based on the independent binary logit models (10) (see details at Appendix B).

	N_B=100			N_B=500		
	10%	5%	1%	10%	5%	1%
LR	12.0	5.2	0.5	10.3	5.4	0.9
W	8.3	3.2	0.4	9.9	4.7	0.8
LM	12.2	6.8	1.4	11.3	5.3	1.0
LR_{sum}^{IL}	14.1	6.7	2.4	13.4	8.2	2.2
	N_B=2 000			N_B=10 000		
	10%	5%	1%	10%	5%	1%
LR	9.7	5.3	1.4	9.7	3.9	0.7
W	9.7	5.0	1.6	9.5	3.8	0.6
LM	10.0	5.1	1.4	9.9	3.8	0.8
LR_{sum}^{IL}	12.7	8.2	3.3	11.8	6.7	1.8

Table 4: Size-adjusted empirical power when the DGP is based on the Premier league (PL) games and the parameter values presented in Table 2. See also the notes to Table 3.

	N_B=100			N_B=500		
	10%	5%	1%	10%	5%	1%
LR	12.1	9.4	2.8	34.7	23.6	9.5
W	11.0	6.2	1.5	32.2	20.4	7.9
LM	17.6	10.8	5.2	39.2	29.9	13.5
LR_{sum}^{IL}	14.9	8.6	2.6	35.7	25.1	10.9
	N_B=2 000			N_B=10 000		
	10%	5%	1%	10%	5%	1%
LR	88.0	80.3	53.8	100	100	100
W	87.1	79.7	52.4	100	100	100
LM	89.3	82.8	60.7	100	100	100
LR_{sum}^{IL}	88.3	80.5	59.1	100	100	100

Table 5: Size-adjusted empirical power based on The Championship (CH) parameters (4th column in Table 2). See also the notes to Tables 3 and 4.

	N_B=100			N_B=500		
	10%	5%	1%	10%	5%	1%
LR	9.2	5.6	1.5	11.2	6.0	1.3
W	10.2	5.3	0.9	11.2	5.9	1.3
LM	10.3	5.6	1.7	11.5	7.1	1.4
LR_{sum}^{IL}	11.5	5.2	0.9	10.3	6.2	1.1
	N_B=2 000			N_B=10 000		
	10%	5%	1%	10%	5%	1%
LR	17.6	9.0	1.1	35.8	25.4	12.1
W	16.3	9.2	1.2	35.1	24.8	11.6
LM	17.7	9.9	1.4	35.6	28.1	12.4
LR_{sum}^{IL}	15.4	8.1	1.1	31.1	21.1	6.5

Table 6: Estimation and efficiency testing results of the independent logit models (see Appendix B). The standard errors of the estimated coefficients are given in the parentheses. The outcome-specific Likelihood ratio test statistics (LR_1^{IL}, LR_0^{IL} and LR_{-1}^{IL}) and the joint test (LR_{sum}^{IL}) and their p-values (given in the parentheses) are also reported.

Dependent variable		Full sample	PL	CH	L1	L2
Home win	a_1	0.036	0.108	0.015	0.032	-0.000
($y_i = 1$)		(0.014)	(0.032)	(0.028)	(0.028)	(0.029)
	b_1	-1.102	-1.130	-1.036	-1.133	-1.063
		(0.030)	(0.048)	(0.065)	(0.064)	(0.069)
	LR_1^{IL}	12.820	15.064	0.394	4.299	1.256
	(p-value)	(0.002)	(0.001)	(0.821)	(0.117)	(0.534)
Draw	a_0	0.305	0.412	0.124	0.014	0.438
($y_i = 0$)		(0.141)	(0.197)	(0.380)	(0.300)	(0.427)
	b_0	-1.304	-1.405	-1.127	-1.013	-1.429
		(0.143)	(0.190)	(0.391)	(0.305)	(0.436)
	LR_0^{IL}	4.829	4.711	0.106	0.011	1.555
	(p-value)	(0.089)	(0.095)	(0.948)	(0.995)	(0.460)
Away win	a_{-1}	0.076	0.038	0.063	0.133	0.008
($y_i = -1$)		(0.030)	(0.052)	(0.064)	(0.064)	(0.067)
	b_{-1}	-1.114	-1.179	-1.082	-1.158	-1.011
		(0.031)	(0.051)	(0.067)	(0.067)	(0.069)
	LR_{-1}^{IL}	16.314	21.376	1.639	5.795	0.027
	(p-value)	(0.000)	(0.000)	(0.441)	(0.055)	(0.986)
	LR_{sum}^{IL}	33.963	41.151	2.140	10.105	2.836
	(p-value)	(0.000)	(0.000)	(0.906)	(0.120)	(0.829)

Table 7: Results of the multinomial logit models at different stages of the season. The estimated coefficients α_j and β_j, their standard errors (given in the parentheses) as well as the efficiency tests based on the Lagrange Multiplier test statistic (p-values are in the parentheses) are presented.

League		Aug-Oct	Nov-Dec	Jan-Feb	Mar-May
PL	α_1	-0.119 (0.103)	0.003 (0.104)	0.008 (0.125)	-0.002 (0.104)
	β_1	1.206 (0.153)	1.061 (0.152)	1.023 (0.178)	1.196 (0.153)
	α_{-1}	-0.104 (0.079)	-0.022 (0.081)	-0.299 (0.100)	-0.084 (0.080)
	β_{-1}	1.141 (0.177)	0.888 (0.173)	1.512 (0.212)	1.367 (0.173)
	LM	5.765 (0.217)	1.182 (0.881)	22.451 (0.000)	18.980 (0.001)
CH	α_1	-0.163 (0.120)	0.042 (0.132)	0.127 (0.145)	0.050 (0.111)
	β_1	1.321 (0.233)	1.051 (0.252)	0.912 (0.269)	0.705 (0.197)
	α_{-1}	-0.013 (0.057)	0.051 (0.067)	0.050 (0.076)	-0.119 (0.063)
	β_{-1}	0.870 (0.236)	0.769 (0.259)	1.233 (0.277)	1.617 (0.217)
	LM	2.114 (0.714)	1.994 (0.737)	2.571 (0.632)	11.192 (0.024)
L1	α_1	-0.024 (0.123)	-0.029 (0.154)	-0.214 (0.125)	-0.013 (0.107)
	β_1	1.075 (0.240)	1.173 (0.283)	1.298 (0.230)	1.013 (0.189)
	α_{-1}	0.041 (0.056)	0.043 (0.078)	-0.123 (0.070)	-0.059 (0.063)
	β_{-1}	1.143 (0.241)	1.412 (0.299)	0.932 (0.245)	1.272 (0.209)
	LM	1.809 (0.771)	5.370 (0.251)	4.853 (0.303)	2.937 (0.568)
L2	α_1	-0.195 (0.125)	0.010 (0.164)	0.033 (0.138)	-0.111 (0.110)
	β_1	1.205 (0.248)	1.123 (0.315)	0.932 (0.261)	1.226 (0.204)
	α_{-1}	-0.045 (0.055)	0.118 (0.079)	0.068 (0.069)	-0.127 (0.063)
	β_{-1}	0.855 (0.246)	1.297 (0.314)	1.102 (0.265)	0.805 (0.220)
	LM	4.820 (0.306)	5.334 (0.254)	1.456 (0.835)	8.596 (0.072)

Figure 1: Predicted probabilities from the multinomial logit model (7) against the implicit probabilities (3) for the home win (left panel) and away win (right panel). The solid line is the robust loess curve between the probabilities while the dashed line is the 45-degree line.
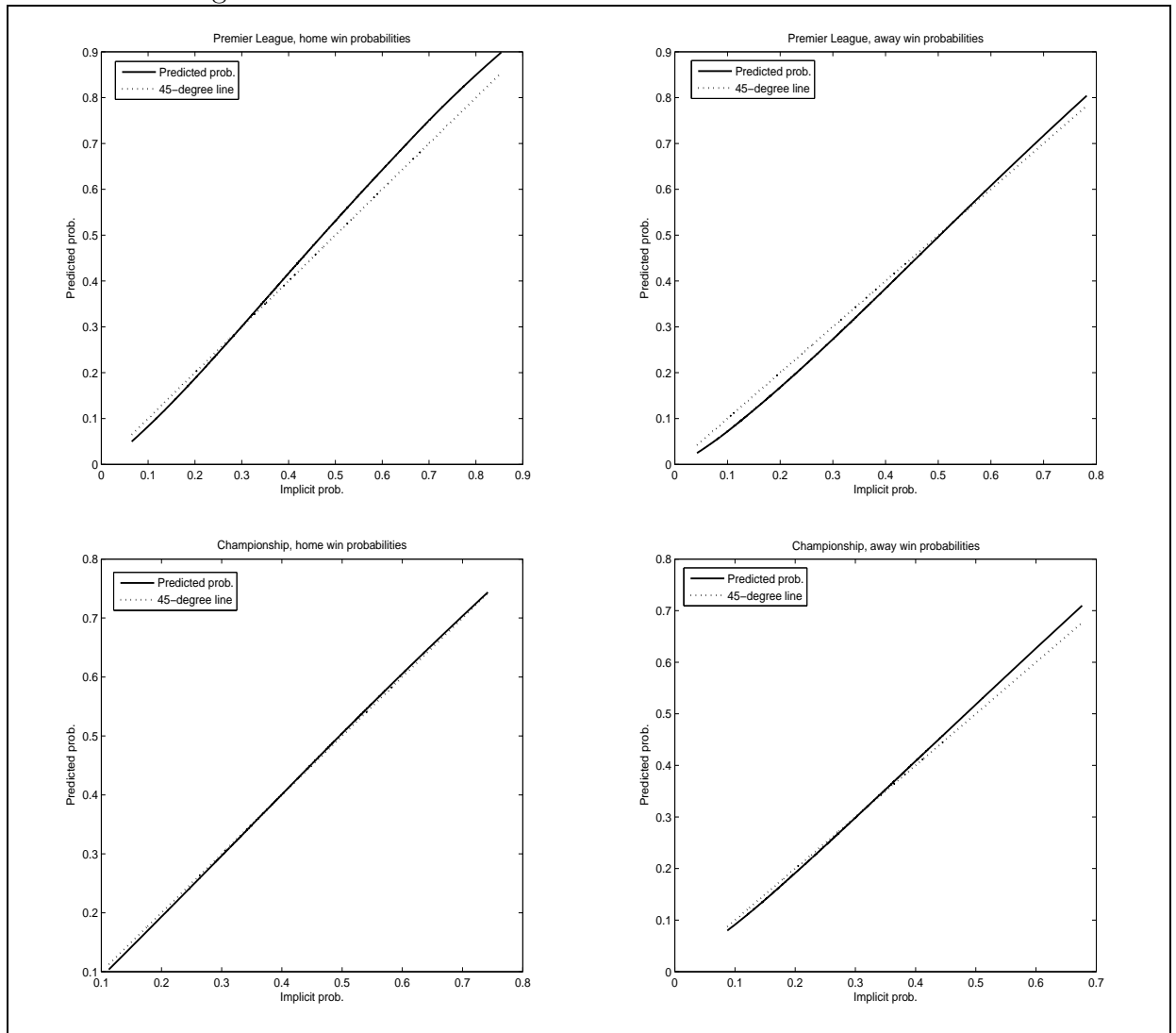
Figure 2: Predicted and implicit probabilities for the League One and League Two. See also the notes to Figure 1.