Efficiency of Online Football Betting Markets

Article *in* SSRN Electronic Journal · November 2017

DOI: 10.2139/ssrn.3070329

2 authors:

Giovanni Angelini
University of Bologna

Luca De Angelis
University of Bologna

Efficiency of online football betting markets

Giovanni Angelini* Luca De Angelis*,†

November 13, 2017

Abstract

This paper evaluates the efficiency in online betting markets for European (association) football championships. The existing literature shows mixed empirical evidence regarding the degree to which betting markets are efficient. We propose a forecast-based approach to formally test the efficiency of online betting markets. By considering the odds proposed by 41 bookmakers on 11 European championships over the last 11 years, we find evidence of different degree of efficiency among markets. We show that, if best odds are selected across bookmakers, seven markets are efficient while four markets show inefficiencies which imply profit opportunities for bettors. In particular, our approach allows the estimation of the odd thresholds that could be used to set a profitable betting strategy both ex post and ex ante.

Keywords: Market efficiency, Sports forecasting, Probability forecasting, Favourite-longshot bias, Betting markets

*School of Economics, Management and Statistics, University of Bologna, Italy
†Corresponding author: l.deangelis@unibo.it

1 Introduction

The issue of the degree of efficiency is crucial in the analysis of markets as market inefficiencies, if correctly predicted and measured, may create significant opportunities to profit. For financial markets, Fama (1970) introduces the renowned Efficient-Market Hypothesis which, in its weak form, postulates that markets are efficient in the sense that current prices reflect all information contained in historical prices, thus ruling out the possibility to achieve excess returns using technical analysis techniques. In general, informational efficiency requires that prices represent the best forecasts on the outcome of future events. Therefore, investors cannot achieve a risk-adjusted return in excess of the market by trading on new information.

Market efficiency naturally applies to many kinds of markets, including betting markets. After the flourishing of the online betting industry in the last decade, a number of scholars have focused on betting markets especially because they represent sort of 'real world laboratories' where efficiency can be investigated straightforwardly (see e.g. the seminal paper by Thaler and Ziemba, 1988, and the comprehensive review for both financial and betting markets by Vaughan Williams, 2005). In fact, unlike financial markets, betting market participants are in general well informed, motivated and experienced and sports breaking news are usually reported cleanly and are easily shared and processed by the agents. In other words, there is very little room for information leakage which, conversely, affects financial market efficiency.[1] Moreover, bets are characterized by a precise deadline after which their value becomes certain, thus making testing for market efficiency much simpler.

Betting market efficiency implies that market prices, i.e. bookmaker odds, reflect all relevant historical information and represent the best forecasts on the match outcome probabilities. Therefore, after considering bookmaker commissions neither bettors nor bookmakers can pursue profit opportunities as all information available is already reflected in the odds quoted. Nevertheless, for (association) football matches, among others, Dixon and Pope (2004), Goddard and Asimakopoulos (2004), Koopman and Lit (2015), Boshnakov et al. (2017), and Angelini and De Angelis (2017) show out-of-sample abnormal positive returns from betting strategies based on econometric approaches, specifically, Poisson, ordered probit, dynamic state-space, bivariate Weibull count and Poisson autogressive models, respectively. As these methods use information on past match results, their results imply betting market inefficiencies. Moreover, forecasting models which produce abnormal positive returns are proposed for other sports, including American football (Glickman and Stern,

[1] Forrest and Simmons (2000) show that private or semi-public information which might be possessed by professional English newspaper tipsters enhances the forecast of the match outcomes only slightly and that there is no overwhelming evidence that the predictions of match results from regression models are inferior to those made by professional experts which claim to possess private (insider) information.

1998; Boulier and Stekler, 2003), tennis (McHale and Morton, 2011), horserace (Lessman et al., 2010) and Australian Rules football (Grant and Johnstone, 2010; Rydall and Bedford, 2010).

The topic of betting market efficiency is rather developed in the literature but the main findings are mixed empirical evidences on the degree to which betting markets are efficient. In particular, the efficiency of win-draw-lose match outcomes in football betting markets is still an open issue. To the best of our knowledge, the (recent) literature lacks of contributions which formally test the efficiency of betting markets for online football wagers. The only studies that we know of that rigorously test for market efficiency for win-draw-lose match outcomes in football are Pope and Peel (1989) and Kuypers (2000). In particular, Pope and Peel (1989) propose an approach based on linear probability and logit models to test for betting market efficiency for the 1981/1982 football season in the UK. Their findings suggest that there is no bias in the bookmaker's odds-setting processes for home and away wins and that no profitable betting strategy can be identified. Using OLS regression between the outcome probability and the probability implied by the odds, Kuypers (2000) concludes that there is no systematic bias in the odds and that the market is weakly efficient for the 1993/1994 and 1994/1995 seasons of the four divisions of the English football league. Kuypers (2000) finds however rare occurrence of both inefficiency and profitable betting opportunities if an ordered logit model with publicly available information variables is employed. With respect to these studies, we propose an innovative approach where we model the bookmaker's forecast errors in order to formally test for market efficiency, after considering bookmaker commissions. Moreover, our analysis considers a larger sample size in terms of time span and number of football leagues, as well as in terms of betting market coverage as Kuypers (2000) and Pope and Peel (1989) consider only one country and one and four bookmakers, respectively.

In this paper we investigate the degree of efficiency in European football online betting markets by testing the predictability of football match outcomes through the information contained in the odds offered in the market. In particular, we test for efficiency in the football online betting market related to single European championships in order to investigate possible differences in the degree of market (in)efficiency among leagues in Europe. To achieve this goal, we consider a data set comprising odds proposed by 41 international bookmakers on 11 championships over the last 11 years (2006-2017), for a total of 33,060 football matches.

One well-known deviation from efficiency is the favourite-longshot bias which claims that odds on favourites are more profitable than odds on longshots (see Sorensen and Ottaviani, 2008, for a review). This bias is well-documented in racetrack betting markets (Griffith, 1949; Thaler and Ziemba, 1988; Hurley and McDonough, 1995) but not in other sports betting markets; e.g. Woodland and Woodland (1994, 2001) show that betting markets for Major League Baseball and National Hockey League matches reveal the opposite bias. By only exploiting information contained in odds and without relying on any econometric model, Direr (2013) shows that systematically picking out betting odds of overwhelmingly favourites (whose probability of winning exceeds 90%) leads to abnormal positive returns. His evidence appears to be against the market efficiency hypothesis and consistent with the literature that documents the presence of a favourite-longshot bias in betting markets.

A further strand of the literature focuses on the accuracy of betting odds-based probability forecasts *per sé* (Štrumbelj and Šikonja, 2010) and in comparison with model-based probability forecasts (Forrest et al., 2005; Štrumbelj and Vračar, 2012). The empirical evidence suggests that betting odds are the most accurate source of sports forecasts. Therefore, in line with this literature, we consider the implied probabilities provided by the online market odds as the 'best' available forecasts of the match outcomes and, under the null hypothesis of market efficiency, we analyze the forecast errors to test for market-specific efficiency.

Our main findings show that all the markets either are efficient or allow extra profits for bookmakers only, if the mean market odds are used. Conversely, if the maximum odds offered by the market are considered we find evidence of four European leagues where the favourite-longshot bias is sufficient to create profitable opportunities for bettors.

This paper is organized as follows. Section 2 outlines the testing approach for market efficiency. In Section 3 we analyze the degree of market efficiency for 11 European championships. Specifically, Section 3.1 describes the data, Section 3.2 presents the results for the market efficiency tests and Section 3.3 investigates the implications of market inefficiencies by presenting a simple but profitable betting strategy. Section 4 concludes.

2 Testing for efficiency of online football betting markets

Let y_i be a dichotomous variable which assumes value 1 if the match i is ended with the outcome under consideration, i.e. home team win, draw, or away team win. The variable y_i is then distributed as a Bernoulli with (true) probability π_i, i.e. $y_i|\Omega_i \sim Bin(1, \pi_i)$, where Ω_i denotes the hypothetical information set containing all the information in the universe. The literature on sports forecasting agrees upon the empirical evidence that betting odds are the most accurate source for predicting the probability of the match outcomes (Štrumbelj and Šikonja, 2010; Štrumbelj, 2014). Therefore, the odds quoted on the online (fixed-odds) betting market represent the 'best' available (*ex ante*) forecasts of the likelihood of the outcome of match i. Let o_i be the bookmaker odd for a particular outcome of match i (e.g. home win) and $p_i = 1/o_i$ be the corresponding implied probability forecast. Hence, the bookmaker's probability forecast should be $p_i = E(y_i|I_i)$ where $I_i \subset \Omega_i$ is the (actual) information set available to the bookmakers on match i. However, bookmakers do not offer fair odds because these also incorporate the bookmaker commission or margin also known as the 'vig'. Therefore, the bookmaker's probability forecast which is *de facto* employed to set the odds offered in the market is $p_i = E(y_i|I_i) + \kappa_i$, where $\kappa_i > 0$ is the bookmaker commission. As a consequence, p_i's are not actual probabilities as their sum over all the possible outcomes exceeds one.[2] Since the bookmaker commission κ_i is in general not fixed and may change among matches, bookmakers and over time, one popular way to circumvent this issue is odds normalization, i.e. dividing the inverse odds by the sum of the inverse odds. However, this approach implies assuming that bookmakers add their margin proportionately across all possible outcomes. Moreover, despite more sophisticated methods to determine probability forecasts from betting odds exist (see e.g. Štrumbelj, 2014), Levitt (2004) shows that bookmakers set odds in order to exploit bettors' biases and thus implied probabilities will be different from those expected even after normalization.[3]

Let $\varepsilon_i = y_i - p_i$ be the bookmaker's forecast error for the outcome of match i. Under the null hypothesis of market efficiency, we have that, in general, p_i overestimates π_i, i.e. $p_i > E(y_i|\Omega_i)$, and, as a consequence, the conditional expectation of ε_i is not null but equals (minus) the bookmaker commission and possible price distortions resulting from bettor's bias exploitation. More specifically, betting market efficiency implies that there is no bias in bookmaker's forecasts apart from commissions (and additional margins), so that the

[2]Occasional cases where the sum of the inverse odds is smaller than one may occur when considering the best odds offered by the market. These cases provide arbitrage opportunities for bettors. Arbitrage opportunities are very rare if only online bookmarkers are considered. Vlastakis et al. (2009) find that less than one match every 1000 matches allows for arbitrage opportunities on online betting markets. However, they also show that this ratio increases to a non-negligible 0.5% if both online and fixed-odds bookmakers are considered. Arbitrage positions can be also achieved by combining bets at exchange and online bookmaker markets (Franck et al., 2013).

[3]For NFL American football, Levitt (2004) provides evidence of the bookmaker's ability to set odds and concludes that they are better than gamblers at predicting match outcomes. Nevertheless, the odds offered by the bookmakers deviate systematically from those expected as they are set also to exploit bettors' biases and thus earning extra profits. A similar argument is made by Kuypers (2000).

expectation of ε_i conditional on the information set I_i available on match i is[4]

$$\begin{aligned} E(\varepsilon_i|I_i) &= E(y_i - p_i|I_i) = E(y_i|I_i) - p_i \quad (1) \\ &= E(y_i|I_i) - (E(y_i|I_i) + \kappa_i) = -\kappa_i. \quad (2) \end{aligned}$$

Market efficiency for championship $j = 1, \ldots, J$ can thus be evaluated by estimating the following model:

$$\varepsilon_{i,j} = \alpha_{1,j} + \sum_{t=2}^{T} \alpha_{t,j} d_t + \beta_j p_{i,j} + \nu_{i,j}, \quad \nu_{i,j} \sim i.i.d.(0, \sigma^2_{i,j}), \quad i = 1, \ldots, N_j \quad (3)$$

where N_j is the number of matches considered for championship j and d_t is a dummy variable which assumes value 1 for season t and 0 otherwise, for $t = 2, \ldots, T$, so that $\alpha_{1,j}$ captures the average bookmaker commission for the j-th championship in season 1 (as reference) and $\alpha_{t,j}$, for $t = 2, \ldots, T$, captures the possible development over time of the bookmaker margins.[5] Since the regression coefficient β_j in (3) captures the possible effect of the probability $p_{i,j}$ on the forecast error $\varepsilon_{i,j}$, the market efficiency of championship j can be evaluated by investigating its statistical significance. More specifically, once we accounted for the bookmaker commissions which are measured by the α coefficients in (3), market efficiency would imply that the conditional expectation $E(\varepsilon_i|I_i)$ is zero so that a rejection of the null hypothesis $H_0 : \beta_j = 0$ would imply that market j is inefficient.[6] Moreover, the inclusion of the intercept dummy variables for each season in the model specification in (3) allows to test whether the bookmaker margin is time-invariant and to capture its possible evolution over time.

Ioannidis and Peel (2005) show that forecast errors can exhibit heteroskedasticity under the null of market efficiency. To account for this issue, the estimation of (3) is obtained through Weighted Least Squares (WLS) where the $N_j \times N_j$ weighting matrix is diagonal with elements $\sigma^2_{1,j}, \ldots, \sigma^2_{N_J,j}$. In our setup, $\sigma^2_{i,j}$ can be approximated by $p_{i,j}(1 - p_{i,j})$.

In the next section we investigate the extent of efficiency in European online betting markets.

[4]Note that, by the law of iterated expectations, $E(y_i|I_i) = E\left[E(y_i|\Omega_i)|I_i\right] = E(\pi_i|I_i)$.

[5]In model (3) we therefore imply that the bookmaker commission may vary over time and across championships. In Section 3.2 we test whether these commissions are time-invariant.

[6] Note that, under the efficient market hypothesis and in case of no bookmaker commission ($\kappa_i = 0$), $E(\varepsilon_i|I_i) = 0$ regardless of the regressor belonging to the information set I_i we might include in the model specification.

3 Empirical results

3.1 Data

The data used in this paper are taken from *www.football-data.co.uk*, a big database on European football match results and fixed-odds, where odds are recorded on Friday afternoons for weekend matches and on Tuesday afternoons for midweek matches. These data comprise the odds offered by the 41 international online bookmakers considered by the BetBrain portal (*www.betbrain.com*), for the football matches played in 11 major European leagues in the period that spans from August 2006 to February 2017, for a total of 33,060 matches. The championships considered in the analysis are the following: English Premier League, Scottish Premier League, German Bundesliga, Italian Serie A, Turkish Super Lig, Portuguese Primeira Liga, French Ligue 1, Spanish Liga, Greek Super League, Dutch Eredivisie, and Belgian Jupiler League. For each match, we consider both the mean and the maximum odd offered by the market. The sample sizes (N_j) which are reported in the last row of Tables 1 and 2 for each championship are rather large so that the theoretical convergence to the normal distribution (of sums) of Bernoulli variables should be attained.

Pope and Peel (1989) show that the variability of draw probabilities is very low and that draw odds have no significant predictive content for the draw outcome. Therefore, in what follows, we do not consider draws.[7]

In our analysis, we also focus on the deviation from market efficiency provided by the favourite-longshot bias, which is an empirical regularity documented in many sports betting markets, as discussed in Section 1.

3.2 Efficiency of online European football betting markets

In this section we test for the efficiency of the online betting markets for the 11 major European football leagues listed in Section 3.1.

If betting markets are efficient, then the conditional expectation of the forecast errors should be equal to minus the bookmaker commissions. Therefore, from the estimation of model (3) for the j-th championship we would expect (i) to find that the estimate for $\alpha_{1,j}$ may be (significantly) negative as this parameter captures the bookmaker margin and (ii) not to reject the null hypothesis $H_0 : \beta_j = 0$. The results for the estimation of the models in (3) are reported in Tables 1 and 2 for the mean and the maximum odds, respectively.

The results in Table 1 show that, considering the mean of the odds offered by the 41 online bookmakers analyzed, we do not reject the null hypothesis of market efficiency for all the

[7] The results including the draw outcome are available upon request.

championships, except for German Bundesliga (only at a level of significance of 0.1), Italian Serie A at 5% significance level, and Portuguese Primeira Liga and Greek Super League even at a 1% level of significance. Quite surprisingly, we find evidence of a negative slope in Germany (as well as in Dutch Eredivisie, albeit not significant). All the other regression slopes (including the non-significant ones) are positive thus implying that, on average, the bookmaker's forecast error tends to increase as the forecast probability increases. This is consistent with the renowned favourite-longshot bias which we investigate below.

The results for the estimates of $\alpha_{1,j}$ reported in Table 1 show that these are all negative (as expected from (2)). Focusing on $\hat{\alpha}_{t,j}$, for $t = 2, ..., 10$, we find evidence that the bookmaker commission has not significantly changed over time. Therefore, in order to improve the power of the test, we simplify the model in (3) by imposing the restriction of time-invariant intercept and we re-estimate the following model:

$$\varepsilon_{i,j} = \alpha_j + \tilde{\beta}_j p_{i,j} + \tilde{\nu}_{i,j}. \tag{4}$$

We report the results for model (4) in the lower panel of Table 1. We find that, on average, the bookmaker commission is significantly different from zero at least at the 5% significance level for all the championships, except for Germany, and ranges from 2.19% (Spain) to 5.24% (Portugal). The results for the efficiency tests are not affected by the restricted model in (4) in terms of significance of the regression coefficients and the implications of evidence of market inefficiencies in Italy, Portugal, Greece, and Germany.

When wagering, bettors tend to select the best price that the market has to offer. Therefore, it is interesting to evaluate the degree of market efficiency when maximum odds are considered (instead of mean odds). Table 2 reports the results for the maximum odds. As in the case of mean odds (cf. Table 1), we do not find evidence of time-varying intercepts and hence we consider the restricted model in (4). Compared to the case of mean odds, the results in Table 2 suggest that, when considering the maximum odds offered by the market, bettors can substantially reduce the bookmaker commissions as only three championships reveal significant (negative) estimates of α, i.e. Italy, Portugal and Greece. For the same three championships we also find evidence of significant estimates of $\tilde{\beta}_j$ at 1% significance level, which confirms that these markets are inefficient, as found in the case of mean odds. Conversely, Germany is not found significant in the case of maximum odds but still displays a negative estimate for the regression slope, thus suggesting a sort of reversed favourite-longshot bias (albeit not significant). Moreover, the results in Table 2 show that we also reject the null hypothesis of market efficiency for Spanish Liga at the 10% significance level.

To better understand the differences in the degree of market efficiency among the European leagues, we derive the efficiency curve for the j-th championship as

$$\hat{G}_j(p_G) = \hat{\alpha}_j + \hat{\hat{\beta}}_j p_G, \quad p_G \in (0,1) \tag{5}$$

where $\hat{\alpha}_j$ and $\hat{\hat{\beta}}_j$ are the estimates of the parameters in (4), and the related confidence bands are computed as

$$CI_j \;=\; \left[\hat{G}_j(p_G) - z_{\alpha/2}\; s.e.\left(\hat{G}_j(p_G)\right),\; \hat{G}_j(p_G) + z_{\alpha/2}\; s.e.\left(\hat{G}_j(p_G)\right)\right] \tag{6}$$

where $s.e.\left(\hat{G}_j(p_G)\right) = \left[\nabla\hat{G}_j(p_G)'\; V_{WLS}\; \nabla\hat{G}_j(p_G)\right]^{1/2}$, $z_{\alpha/2}$ is the $100(1-\alpha/2)$-th percentile of the standard normal distribution, $\nabla\hat{G}_j(p_G) = (1,\; p_G)'$ is the gradient and V_{WLS} is the variance of the WLS estimator.

In Figures 1 and 2, the efficiency curves $\hat{G}_j$ in (5) for each championship are plotted against $p_G \in (0,1)$ for the mean and the maximum odds, respectively. For a fixed value of p_G, say $p_G^0 \in (0,1)$, $\hat{G}_j(p_G^0) = 0$ implies market efficiency. Conversely, when $\hat{G}_j(p_G^0) \neq 0$, there is evidence of inefficiency and the sign of $\hat{G}_j(p_G^0)$ suggests us which side might profit from this inefficiency. In particular, $\hat{G}_j(p_G^0) > 0$ would imply positive returns for bettors, whereas $\hat{G}_j(p_G^0) < 0$ would entail profits for bookmakers.

The efficiency curves depicted in Figure 1 show that, for all the championships except for German Bundesliga and Dutch Eredivisie, since $\hat{\hat{\beta}}_j > 0$, $\hat{G}_j$ tends to increase as the outcome probability p_G increases. We find therefore evidence that market probabilities of underdogs (favourites) overpredict (underpredict) on average their empirical probabilities. This implies that longshots are underpriced and that wagering on favourites is more profitable for bettors. We can interpret these results as evidence supporting the favourite-longshot bias. However, as can be seen from Figure 1, all the efficiency curves are below the zero line, except in the case of the largest values of p_G for Italy, Portugal and Greece, and the related confidence bands, which are depicted in Figure 1 for the 95% confidence level, show that no significant positive values of $\hat{G}_j(p_G)$ can be achieved. This empirical evidence implies that bettors cannot systematically achieve positive returns and that bookmakers profit in the long run, especially from longshots.[8]

Figure 2 shows the efficiency curves for the case of maximum odds. When the best odds offered by the market are considered, $\hat{G}_j$ suggests lower profitability for bookmakers as the values for $\hat{\alpha}_j$ are now close to zero (cf. Table 2). In particular, according to the 95% confidence bands depicted in Figure 2 neither bookmakers nor bettors can achieve significant returns

[8]Conversely, in German Bundesliga bookmakers appear to profit from favourites and not from longshots. Interestingly, for Dutch Eredivisie bookmakers seem to achieve significant profits for any value of p_G.

in the following championships: English Premier League, Scottish Premier League, German Bundesliga, Turkish Super Lig, French Ligue 1, Dutch Eredivisie, and Belgian Jupiler League. Hence, for these seven European leagues the online betting markets are found to be efficient if the best odds are considered. Conversely, Italian Serie A, Portuguese Primeira Liga and Greek Super League allow significant returns for bettors when p_G is large (close to one) and significant profits for bookmakers for smaller values of p_G (close to zero). Therefore, in the case of maximum odds, we find empirical evidences in favor of the favourite-longshot bias and there is room for profit opportunities for both bettors and bookmakers. Finally, an interesting case is given by the Spanish Liga where picking the best odds on the market appear to deliver significant positive returns for bettors but not for bookmakers.

A way to exploit these empirical findings using a simple but profitable betting strategy is described in the next section in terms of both *ex ante* and *ex post* forecasting performances.

3.3 Implications of market inefficiency: A simple and profitable betting strategy

In this section we propose a betting strategy which aims to exploit the market inefficiencies found in the previous section. Figure 2 shows that the efficiency curves in (5) are significantly positive for four European online betting markets when picking the maximum odds offered by the market. Indeed, as discussed in Section 3.2, the Italian, Portuguese, Spanish and Greek leagues show positive values of $\hat{G}_j(p_G)$ associated with the largest probabilities p_G (the smallest odds). Our betting strategy for championship j can be summarized in the following steps:

1. Estimate the model in (4) by WLS as described in Section 2, considering the observations until season T^* as information set I_i:

$$\varepsilon_{i,j} = \alpha_j^* + \tilde{\beta}_j^* p_{i,j} + \tilde{\nu}_{i,j}, \quad i = 1, \dots, N_j^{T^*}, \tag{7}$$

   where $N_j^{T^*}$ is the number of matches played in the j-th championship in the seasons $t = 1, \dots, T^*$.

2. Using the results obtained in step 1, compute the efficiency curve for championship j up to season T^* as

$$\hat{G}_j^*(p_G) = \hat{\alpha}_j^* + \hat{\tilde{\beta}}_j^* p_G, \quad p_G \in (0, 1) \tag{8}$$

and the related confidence bands

$$CI_j^* = \left[\hat{G}_j^*(p_G) - z_{\alpha/2} s.e. \left(\hat{G}_j^*(p_G) \right), \ \hat{G}_j^*(p_G) + z_{\alpha/2} s.e. \left(\hat{G}_j^*(p_G) \right) \right]. \qquad (9)$$

3. Define the 'profitable probability range' as $P_j^* = \{p_G \in (0,1) : \underline{CI}_j^* > 0\}$, where $\underline{CI}_j^*$ denotes the lower bound of the confidence interval in (9), the threshold probabilities as the infimum and supremum of the profitable probability range, i.e. $p_{G,j}^{*(L)} = \inf_{p_G \in (0,1)} P_j^*$ and $p_{G,j}^{*(U)} = \sup_{p_G \in (0,1)} P_j^*$, and the related threshold odds $o_j^{*(L)} = \left(p_{G,j}^{*(L)}\right)^{-1}$ and $o_j^{*(U)} = \left(p_{G,j}^{*(U)}\right)^{-1}$.
4. Systematically wager all the matches of championship j whose odds are in the 'profitable odds range', $O_j^* = \left[o_j^{*(L)}, \ o_j^{*(U)}\right]$, for either all the seasons after T^* (out-of-sample forecast) or all the seasons in the sample (in-sample forecast).[9]

We adopt the above betting strategy to evaluate both the in-sample (*ex post*) and the out-of-sample (*ex ante*) forecasting performances. The results are reported in Table 3.

First we focus on the in-sample forecast which considers the whole information set available, i.e. $T^* = 2016/17$. The results in the upper panel of Table 3 show that the betting strategy delivers positive mean returns for all the four European leagues which are found to be inefficient in Section 3.2. In particular, by systematically betting odds inferior to 1.78, 2.50, 1.23, and 1.92, we achieve mean returns of 2.71%, 2.55%, 5.11%, and 3.64% for Italian Serie A, Spanish Liga, Portuguese Primeira Liga, and Greek Super League, respectively.

Next we investigate whether abnormal out-of-sample returns can be obtained from the proposed betting strategy. To evaluate the out-of-sample forecasting performance, we fix $T^* = 2013/14$ as information set and we use the 2014/15, 2015/16, and 2016/17 seasons as out-of-sample period.[10] The figures reported in the bottom panel of Table 3 depict the efficiency curves $\hat{G}_j^*(p_G)$ for the four inefficient championships (Italy, Spain, Portugal and Greece) and the related 95% confidence bands CI_j^* computed as in (8) and (9), respectively. From these figures, we note that no profitable probability range is found for Portugal when considering $T^* = 2013/14$ as information set, i.e. the estimated lower bound of the confidence interval $\underline{CI}_j^*$ is below the zero line for all values of p_G. Conversely, for Italy, Greece and Spain there are values of p_G where the condition $\underline{CI}_j^* > 0$ is satisfied. Hence, following steps 3 and 4 of our betting strategy we compute the profitable probability range, P_j^*, and the

[9] Note that in our approach the profitable odds range is estimated and not arbitrarily chosen as in previous analyses (see e.g. Direr, 2013).

[10] In this out-of-sample forecasting exercise, we have extended the sample period which now incorporates also the end of season 2016/17, i.e. the matches played from March to June 2017.

corresponding profitable odds range, O_j^* (cf. middle panel of Table 3). In particular, we observe that the profit opportunities for bettors can be pursued in Italian Serie A and Greek Super League when wagering on match outcomes with odds inferior to 1.41 and 2.17 or, equivalently, implied probability larger than 0.7092 and 0.4608, respectively. For Spanish Liga the profitable range is limited to $O^* = [1.41, 2.56]$, i.e. $P^* = [0.3906, 0.7092]$. The results reported in the middle panel of Table 3 show that the mean returns are positive for all the three championships. In particular, in the out-of-sample period from 2014/15 to 2016/17 seasons the betting strategy delivers mean returns of 4.68%, 1.24% and 1.58% for Italy, Spain and Greece, respectively.

4 Conclusions

Online betting markets have developed, evolved and thrived over the last decades and scholars have increased their interest in investigating the characteristics of these markets. In this paper, we focus on the degree of efficiency of the online betting market for European football using a large data set. Considering the mean market odds, we provide evidence that online betting markets are moderately (weak-form) efficient and that, when inefficiencies are detected, these provide extra profits for bookmakers. However, thanks to a highly competitive market, bettors can choose between many bookmakers and pick the best odds offered by the market. Repeating the analysis using maximum odds, we find that the majority of the online betting markets are efficient but we also find evidence of inefficiencies that can be exploited to define profitable betting strategies. In particular, our analysis shows that one of the most popular inefficiency in betting markets, the favourite-longshot bias, is indeed present in four European football markets. We show that a simple betting strategy which exploits this bias leads to abnormal positive returns for bettors, after considering bookmaker commissions. Moreover, our results show that the commissions of online bookmakers have not significantly changed over time in the period from 2006 to 2017 but appear to be different across championships.

References

Angelini, G., & De Angelis, L. (2017). PARX model for football matches predictions. *Journal of Forecasting*, forthcoming.

Boshnakov, G., Kharrat, T., & McHale, I.G. (2017). A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, 33(2), 458–466.

Boulier, B.L., & Stekler, H.O. (2003). Predicting the outcomes of National Football League games. *International Journal of Forecasting*, 19(2), 257–270.

Direr, A. (2013). Are betting markets efficient? Evidence from European Football Championships. *Applied Economics*, 45, 343–356.

Dixon, M.J., & Pope, P.F. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting*, 20, 686–700.

Fama, E.F. (1970). Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, 25, 383–417.

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

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

Franck, E., Verbeek, E., & Nüesch, S. (2013). Inter-market Arbitrage in Betting. *Economica*, 80, 300–325.

Glickman, M. E., & Stern, H. S. (1998). A state-space model for National Football League scores. *Journal of the American Statistical Association*, 93, 25–35.

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

Grant, A., & Johnstone D. (2010). Finding profitable forecast combinations using probability scoring rules. *International Journal of Forecasting*, 26, 498–510.

Griffith, R.M. (1949). Odds Adjustment by American Horse-Race Bettors. *American Journal of Psychology*, 62, 290–294.

Hurley, W., & McDonough, L. (1995). Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting. *American Economic Review*, 85, 949–955.

Koopman, S.J., & Lit, R. (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *Journal of the Royal Statistical Society: Series A*, 178(1), 167–186.

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

Ioannidis, C., & Peel, D.A. (2005). Testing for market efficiency in gambling markets when the errors are non-normal and heteroskedastic an application of the wild bootstrap." *Economics Letters*, 87, 221–226.

Lessmann, S., Sungb, M.-C., & Johnson, J.E.V. (2010). Alternative methods of predicting competitive events: An application in horserace betting markets. *International Journal of Forecasting*, 26, 518–536.

Levitt, S. (2004). Why are gambling markets organised so differently?. *The Economic Journal*, 114, 223–46.

McHale, I., & Morton, A. (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, 27(2), 619–630.

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

Rydall, R., & Bedford, A. (2010). An optimized ratings-based model for forecasting Australian Rules football. *International Journal of Forecasting*, 26, 511–517.

Sorensen, P.N., & Ottaviani, M. (2008). The favorite-longshot bias: an overview of the main explanations. in Handbook of Sports and Lottery Markets (Eds) D. B. Hausch and W. T. Ziemba, North-Holland, Amsterdam, 83–101.

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

Štrumbelj, E., & Vračar, P. (2012). Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. *International Journal of Forecasting*, 28(2), 532–542.

Štrumbelj, E., & Šikonja, R. (2010). Online bookmakers odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting*, 26(3), 482–488.

Thaler, R., & Ziemba, W. (1988). Parimutual betting markets: racetracks and lotteries. *Journal of Economic Perspectives*, 2, 161–174.

Vaughan Williams, L. (2005). *Information Efficiency in Financial and Betting Markets*. Cambridge: Cambridge University Press.

Vlastakis, N., Dotsis, G., & Markellos, R.N. (2009). How Efficient is the European Football Betting Market? Evidence from Arbitrage and Trading Strategies. *Journal of Forecasting*, 28, 426–444.

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

Woodland, L.M., & Woodland, B.M. (2001). Market efficiency and profitable wagering in the National Hockey League: Can bettors' score on longshots?. *Southern Economic Journal*, 67, 983–995.

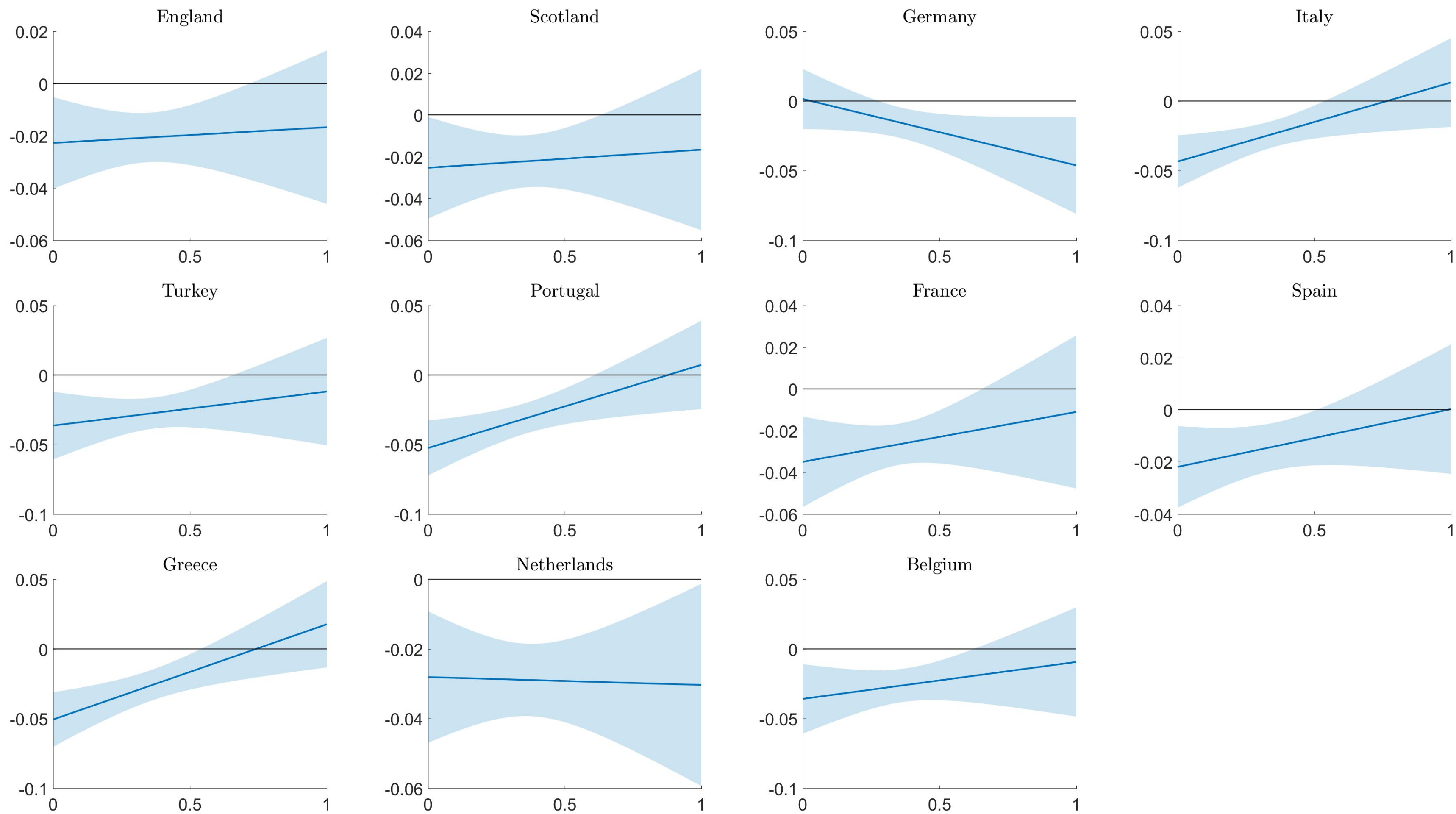

Figure 1: Efficiency curves $\hat{G}_j(p_g)$ in (5) and related 95% confidence bands in (6) computed considering the mean of the odds offered by the betting market.

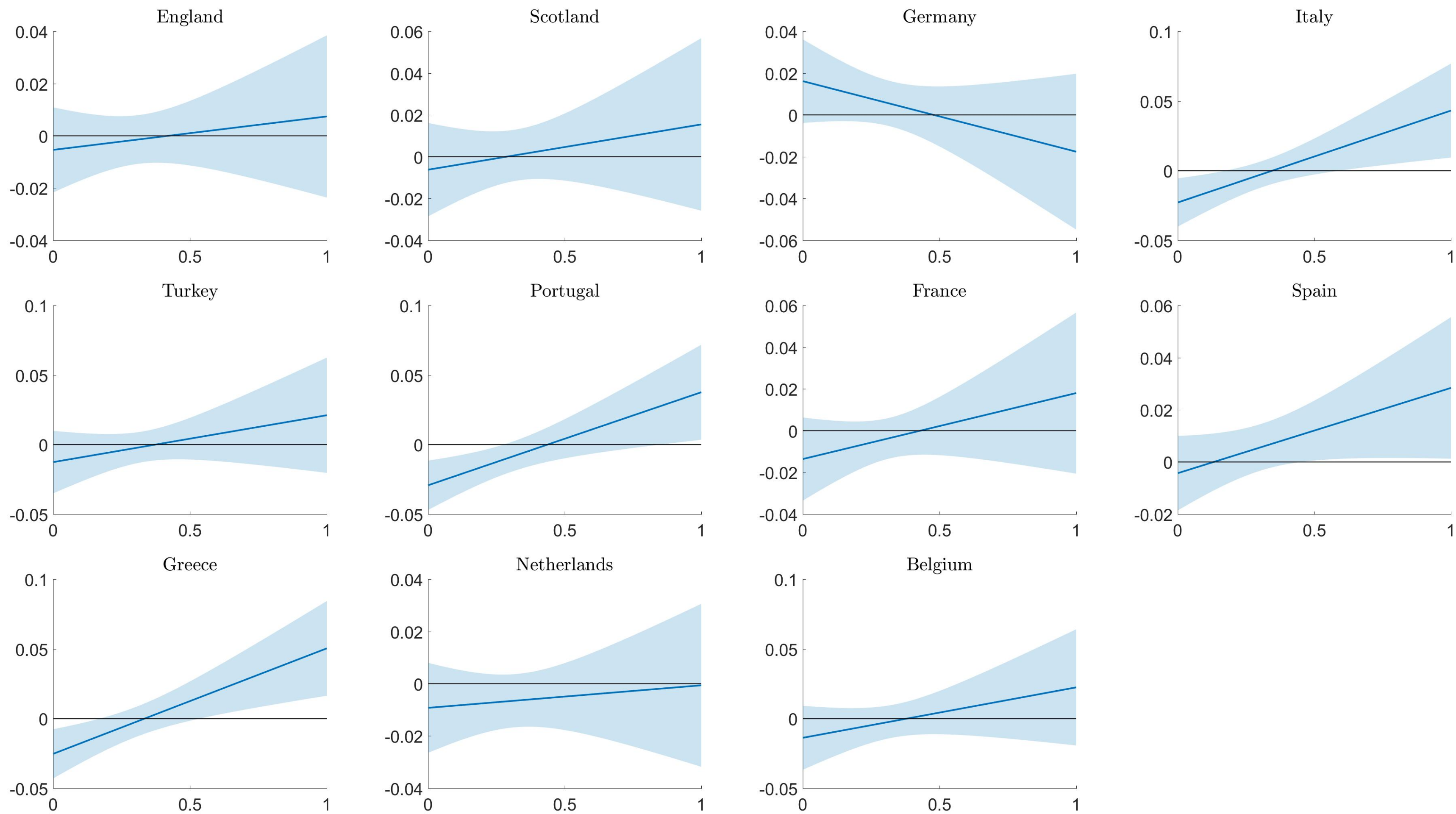

Figure 2: Efficiency curves $\hat{G}_j(p_g)$ in (5) and related 95% confidence bands in (6) computed considering the maximum odds offered by the betting market.

						Mean odds					
	England	Scotland	Germany	Italy	Turkey	Portugal	France	Spain	Greece	Holland	Belgium
$\hat{\alpha}_1$	-0.0417^{**}	−0.0226	−0.0066	-0.0571^{***}	-0.0351^{*}	-0.0799^{***}	-0.0453^{**}	−0.0134	-0.0489^{**}	−0.0183	-0.0488^{**}
	(−2.4792)	(−1.0276)	(−0.3242)	(−3.3231)	(−1.6960)	(−3.8502)	(−2.3891)	(−0.7841)	(−2.4537)	(−0.9776)	(−2.3438)
$\hat{\alpha}_2$	0.0116	−0.0085	0.0062	0.0173	0.0081	0.0220	0.0024	−0.0004	−0.0212	−0.0114	0.0218
	(0.5437)	(−0.3064)	(0.2448)	(0.8071)	(0.3324)	(0.8246)	(0.1058)	(−0.0175)	(−0.8255)	(−0.4799)	(0.8902)
$\hat{\alpha}_3$	0.0155	−0.0173	−0.0081	0.0118	−0.0195	0.0144	0.0235	−0.0257	0.0001	0.0000	0.0181
	(0.746)	(−0.6313)	(−0.3221)	(0.5527)	(−0.7713)	(0.5491)	(1.0504)	(−1.1941)	(−0.0011)	(−0.0019)	(0.6699)
$\hat{\alpha}_4$	0.0102	0.0052	0.0326	0.0214	0.0020	0.0250	−0.0180	−0.0004	0.0010	−0.0075	0.0025
	(0.4831)	(0.1921)	(1.2828)	(0.9963)	(0.0827)	(0.9460)	(−0.8000)	(−0.0191)	(0.0399)	(−0.3222)	(0.0965)
$\hat{\alpha}_5$	0.0277	0.0019	0.0031	0.0058	0.0038	0.0409	0.0071	−0.0170	0.0080	0.0009	−0.0026
	(1.315)	(0.0677)	(0.1219)	(0.2696)	(0.1543)	(1.5840)	(0.3135)	(−0.8071)	(0.3125)	(0.0410)	(−0.1011)
$\hat{\alpha}_6$	0.0067	−0.0195	0.0037	0.0223	−0.0035	0.0276	0.0071	−0.0013	0.0031	−0.0130	0.0210
	(0.3150)	(−0.6967)	(0.1468)	(1.0404)	(−0.1395)	(1.0550)	(0.3160)	(−0.0602)	(0.1200)	(−0.5556)	(0.7967)
$\hat{\alpha}_7$	0.0395^{*}	0.0068	0.0179	0.0167	−0.0006	0.0338	0.0074	−0.0035	0.0081	−0.0269	0.0218
	(1.8793)	(0.2454)	(0.7283)	(0.7878)	(−0.0235)	(1.2999)	(0.3327)	(−0.1687)	(0.3410)	(−1.1349)	(0.8394)
$\hat{\alpha}_8$	0.0295	0.0188	0.0008	−0.0034	−0.0086	0.0246	0.0318	−0.0047	−0.0062	−0.0047	0.0123
	(1.3871)	(0.6817)	(0.0327)	(−0.1583)	(−0.3470)	(1.0099)	(1.4252)	(−0.2247)	(−0.2558)	(−0.1987)	(0.4694)
$\hat{\alpha}_9$	0.0205	0.0040	0.0150	0.0180	0.0025	0.0398	0.0149	−0.0133	0.0001	−0.0148	0.0188
	(0.9541)	(0.1455)	(0.6138)	(0.8505)	(0.1004)	(1.6388)	(0.6657)	(−0.6377)	(0.0046)	(−0.6211)	(0.7131)
$\hat{\alpha}_{10}$	0.0294	−0.0262	0.0092	0.0317	0.0038	0.0357	0.0279	−0.0202	−0.0177	−0.0284	0.0172
	(1.2326)	(−0.8654)	(0.3305)	(1.3529)	(0.1334)	(1.3148)	(1.1277)	(−0.8526)	(−0.6455)	(−1.0835)	(0.6266)
$\hat{\beta}$	0.0067	0.0087	-0.0476^{*}	0.0568^{**}	0.0244	0.0610^{***}	0.0249	0.0220	0.0685^{***}	−0.0023	0.0267
	(0.3066)	(0.2963)	(−1.8107)	(2.3869)	(0.8156)	(2.5823)	(0.8931)	(1.2040)	(2.9817)	(−0.1033)	(0.8844)
$\hat{\alpha}$	-0.0227^{**}	-0.0253^{**}	0.0014	-0.0434^{***}	-0.0363^{***}	-0.0524^{***}	-0.0349^{***}	-0.0219^{***}	-0.0506^{***}	-0.0281^{***}	-0.0359^{***}
	(−2.5386)	(−2.0410)	(0.1306)	(−4.5256)	(−2.9317)	(−5.2108)	(−3.1620)	(−2.7464)	(−5.0843)	(−2.9173)	(−2.8172)
$\hat{\hat{\beta}}$	0.0060	0.0087	-0.0476^{*}	0.0567^{**}	0.0244	0.0597^{**}	0.0239	0.0221	0.0683^{***}	−0.0023	0.0265
	(0.2772)	(0.2963)	(−1.8117)	(2.3837)	(0.8188)	(2.5325)	(0.8585)	(1.2126)	(2.9738)	(−0.1031)	(0.8764)
N_j	7320	4394	5886	7338	5810	4976	7356	7294	4868	5922	4954

Table 1: Estimates of the models in (3) and (4) when we consider the mean of the odds offered on the betting market. t-stats are reported in brackets. Asterisks denote significance at 1%(∗∗∗), 5%(∗∗), and 10%(∗) levels. The last row reports the number of matches N_j played in each championship.

	England	Scotland	Germany	Italy	Turkey	Portugal	France	Spain	Greece	Holland	Belgium
						Max odds					
$\hat{\alpha}_1$	−0.0224	−0.0008	0.0146	−0.0301*	−0.0075	−0.0499**	−0.0206	0.0106	−0.0212	0.0052	−0.0218
	(−1.3858)	(−0.0395)	(0.7436)	(−1.8319)	(−0.3762)	(−2.4822)	(−1.1230)	(0.6326)	(−1.1027)	(0.2907)	(−1.0890)
$\hat{\alpha}_2$	0.0132	−0.0071	0.0042	0.0143	0.0112	0.0209	0.0013	−0.0031	−0.0253	−0.0122	0.0199
	(0.6346)	(−0.2609)	(0.1679)	(0.6821)	(0.4619)	(0.7953)	(0.0593)	(−0.1409)	(−1.0026)	(−0.5222)	(0.8274)
$\hat{\alpha}_3$	0.0124	−0.0188	−0.0124	0.0053	−0.0209	0.0077	0.0209	−0.0341	−0.0029	−0.0087	0.0145
	(0.6117)	(−0.6953)	(−0.5003)	(0.2518)	(−0.8358)	(0.2971)	(0.9446)	(−1.6021)	(−0.1143)	(−0.3798)	(0.5404)
$\hat{\alpha}_4$	0.0095	0.0001	0.0266	0.0142	−0.0016	0.0193	−0.0204	−0.0053	0.0013	−0.0136	−0.0032
	(0.4541)	(0.0012)	(1.0588)	(0.6724)	(−0.0664)	(0.7451)	(−0.9119)	(−0.2518)	(0.0499)	(−0.5937)	(−0.1248)
$\hat{\alpha}_5$	0.0273	−0.0018	−0.0048	−0.0004	−0.0019	0.0299	0.0035	−0.0241	0.0089	−0.0043	−0.0086
	(1.3221)	(−0.0664)	(−0.1955)	(−0.0209)	(−0.0775)	(1.1781)	(0.1552)	(−1.1682)	(0.3580)	(−0.1898)	(−0.3347)
$\hat{\alpha}_6$	0.0033	−0.0178	−0.0037	0.0158	−0.0105	0.0183	0.0036	−0.0107	0.0005	−0.0163	0.0158
	(0.1578)	(−0.6431)	(−0.1509)	(0.7528)	(−0.4215)	(0.7166)	(0.1621)	(−0.5067)	(0.0213)	(−0.7107)	(0.6062)
$\hat{\alpha}_7$	0.0370*	0.0002	0.0075	0.0079	−0.0090	0.0276	0.0036	−0.0075	0.0012	−0.0316	0.0161
	(1.7942)	(0.0089)	(0.3120)	(0.3790)	(−0.3636)	(1.0837)	(0.1621)	(−0.3642)	(0.0504)	(−1.3557)	(0.6260)
$\hat{\alpha}_8$	0.0270	0.0148	−0.0055	−0.0076	−0.0108	0.0151	0.0272	−0.0128	−0.0073	−0.0090	0.0062
	(1.2864)	(0.5470)	(−0.2250)	(−0.3649)	(−0.4403)	(0.6361)	(1.2323)	(−0.6301)	(−0.3109)	(−0.3860)	(0.2407)
$\hat{\alpha}_9$	0.0177	0.0028	0.0054	0.0078	−0.0076	0.0340	0.0083	−0.0239	−0.0055	−0.0227	0.0093
	(0.8299)	(0.1023)	(0.2262)	(0.3763)	(−0.3065)	(1.4264)	(0.3761)	(−1.1590)	(−0.2244)	(−0.9627)	(0.3577)
$\hat{\alpha}_{10}$	0.0242	−0.0352	0.0002	0.0196	−0.0034	0.0246	0.0212	−0.0239	−0.0178	−0.0365	0.0087
	(1.0282)	(−1.1787)	(0.0064)	(0.8461)	(−0.1212)	(0.9176)	(0.8623)	(−1.0215)	(−0.6628)	(−1.4124)	(0.3200)
$\hat{\beta}$	0.0130	0.0214	−0.0342	0.0658***	0.0341	0.0681***	0.0328	0.0319*	0.0760***	0.0090	0.0364
	(0.5918)	(0.7227)	(−1.2758)	(2.7476)	(1.1219)	(2.8608)	(1.1731)	(1.6971)	(3.2209)	(0.4055)	(1.1924)
$\hat{\alpha}$	−0.0054	−0.0062	0.0162	−0.0228***	−0.0126	−0.0292***	−0.0136	−0.0043	−0.0253***	−0.0092	−0.0138
	(−0.6460)	(−0.5427)	(1.5829)	(−2.5835)	(−1.0994)	(−3.2321)	(−1.3383)	(−0.5900)	(−2.8087)	(−1.0501)	(−1.1779)
$\hat{\hat{\beta}}$	0.0128	0.0217	−0.0338	0.0660***	0.0337	0.0669***	0.0316	0.0327*	0.0757***	0.0086	0.0362
	(0.5838)	(0.7306)	(−1.2598)	(2.7541)	(1.1101)	(2.8141)	(1.1322)	(1.7421)	(3.2104)	(0.3897)	(1.1884)
N_j	7320	4394	5886	7338	5810	4976	7356	7294	4868	5922	4954

Table 2: Estimates of the models in (3) and (4) when we consider the maximum of the odds offered on the market. t-stats are reported in brackets. Asterisks denote significance at 1%(∗∗∗), 5%(∗∗), and 10%(∗) levels. The last row reports the number of matches N_j played in each championship.

	Italy	Spain	Portugal	Greece
In-sample forecasting performance				
Percentage mean return	2.71	2.55	5.11	3.64
Number of bets	1190	3078	150	1231
Percentage of correct bets	68.32	58.71	88.67	68.80
Profitable probability range	$P^* = [.5618,\ 1)$	$[.4000,\ 1)$	$[.8130,\ 1)$	$[.5208,\ 1)$
Profitable odds range	$O^* = (1,\ 1.78]$	$(1,\ 2.50]$	$(1,\ 1.23]$	$(1, 1.92]$
Out-of-sample forecasting performance (up to season $T^* = 2013/14$)				
Percentage mean return	4.68	1.24	-	1.58
Number of bets	139	773	-	521
Percentage of correct bets	81.16	52.72	-	64.61
Profitable probability range	$P^* = [.7092,\ 1)$	$[.3906,\ .7092]$	-	$[.4608,\ 1)$
Profitable odds range	$O^* = (1,\ 1.41]$	$[1.41,\ 2.56]$	-	$(1, 2.17]$

Efficiency curve

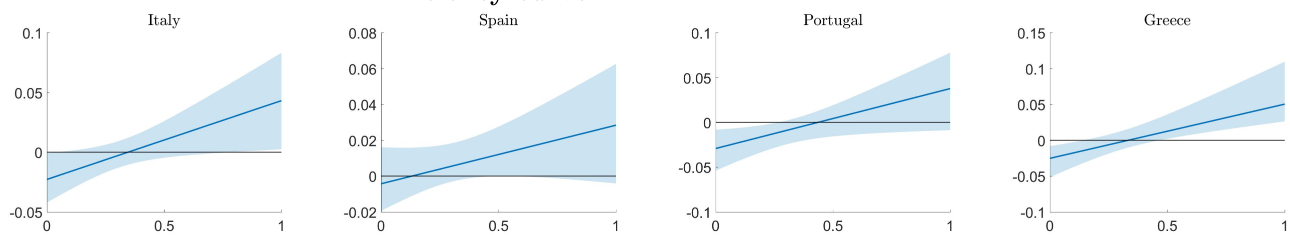

Table 3: Upper panel: in-sample forecasting performance of the betting strategy described in Section 3.3. Middle panel: out-of-sample forecasting performance of the betting strategy described in Section 3.3. Bottom panel: efficiency curves $\hat{G}^*_j(p_g)$ in (8) (blue lines) and related 95% confidence bands computed as in (9) used in the out-of-sample forecast.