When the league table lies: Does outcome bias lead to informationally inefficient markets?

Flepp, Raphael ; Merz, Oliver ; Franck, Egon

DOI: https://doi.org/10.1111/ecin.13163

Originally published at:
Flepp, Raphael; Merz, Oliver; Franck, Egon (2024). When the league table lies: Does outcome bias lead to informationally inefficient markets? Economic Inquiry, 62(1):414-429.
DOI: https://doi.org/10.1111/ecin.13163

Received: 27 April 2022 | Accepted: 5 June 2023

DOI: 10.1111/ecin.13163

ORIGINAL ARTICLE

Economic Inquiry

When the league table lies: Does outcome bias lead to informationally inefficient markets?

Raphael Flepp | **Oliver Merz** | **Egon Franck**

University of Zurich, Zurich, Switzerland

Correspondence
Raphael Flepp, University of Zurich, Plattenstrasse 14, Zurich 8032, Switzerland.
Email: raphael.flepp@business.uzh.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (Swiss National Science Foundation), Grant/Award Number: 162815

Abstract

We study whether outcome bias persists in markets with actors who are financially incentivized to make optimal decisions. We test whether inherently noisy match outcomes from European football are correctly incorporated into prices from a betting exchange market. We find that market prices overestimate (underestimate) the winning probability of teams that previously overperformed (underperformed) in terms of match outcomes compared to their performance based on "expected goals." This pattern is mirrored in negative (positive) betting returns on overperforming (underperforming) teams. These results suggest that even competitive market mechanisms fail to completely erase outcome bias.

KEYWORDS

betting exchange market, market efficiency, outcome bias, soccer

JEL CLASSIFICATION

D90, G14, Z20

1 | INTRODUCTION

People often ignore or underestimate the causal role of external, random or extraneous factors that influence outcomes (Allison et al., 1996). This tendency may lead to outcome bias, which is present whenever individuals assign too much importance to noisy outcomes in evaluating past decisions or task performances (e.g., Baron & Hershey, 1988). For example, assume that a football (soccer) team plays poorly in a match but manages to win the game by scoring one lucky goal. If people overweight the outcome of winning the game compared to the team's task performance on the pitch when evaluating the team's performance, outcome bias is present.

Outcome bias has been shown to be rather persistent, as it has been found in many different experimental settings, such as legal decisions (Alicke et al., 1994), medical decisions (Baron & Hershey, 1988), investment decisions (König-Kersting et al., 2021; Ratner & Herbst, 2005) and ethical judgments (Gino et al., 2009, 2010). Moreover, several studies have provided field evidence consistent with insights from the laboratory (e.g., Gauriot & Page, 2019; Lefgren

Abbreviations: OLT, official league table; xG, expected goals; xGD, expected goal difference; xGT, league table based on expected goals.
Managing Editor: Rob Simmons

et al., 2015; Meier et al., 2022). However, previous research has focused on outcome bias in individual-level decision-making. Thus, it remains unclear whether outcome bias persists in market mechanisms where prices aggregate dispersed information or whether such competitive market mechanisms can overcome the underestimation of randomness in outcomes.

We address this research gap by analyzing the influence of outcome bias in European football on the informational efficiency of market prices from betting exchanges. This setting offers several unique advantages. First, football is a low-scoring sport in which randomness in match outcomes is substantial (Brechot & Flepp, 2020; Wunderlich et al., 2021). This feature makes performance evaluation prone to outcome bias because match outcomes may not reliably reflect team performance on the pitch. Second, market prices of betting exchanges represent forecasts of future events by aggregating dispersed information from numerous, well-informed, and financially incentivized individuals trading with each other (Brown et al., 2019). Moreover, the true value of each betting contract is revealed at the end of each match against which the informational efficiency of market prices can be tested (Vaughan Williams, 1999).

We use data on matches from the top five European football leagues played between the 2013/2014 and 2017/2018 seasons. Football data are obtained from Gracenote, and odds data are obtained from the betting exchange Matchbook via www.oddsportal.com. Following Brechot and Flepp (2020), we employ "expected goals" as a more reliable performance evaluation measure compared to match outcomes. The expected goals metric is based on quantified scoring chances rather than actual goals and has been shown to better predict future sporting results than metrics based on match outcomes (e.g., Anzer & Bauer, 2021). We measure a teams' over- or underperformance in terms of match outcomes compared to their performance based on expected goals by calculating the table rank difference between the rank in the official league table (OLT) and the rank in a league table based on expected goals (xGT).

To analyze whether market prices correctly incorporate past match outcomes, we follow previous research (e.g., Brown et al., 2018) and employ a binary probability model using the actual result of a bet on either the home or the away team as the dependent variable equaling 1 if the bet is won and 0 if the bet is lost. If the market prices are efficient, all relevant information should be reflected in them, and no additional variables should have predictive power regarding the actual result of the betting event. In other words, if the market prices are not influenced by outcome bias, the table rank difference should not have explanatory power beyond the winning probabilities implied by the betting odds. However, if bettors are overweighting the importance of past match outcomes in the aggregate, the information contained in the table rank difference is expected to be incorrectly reflected in the market prices.

We find that the table rank difference has explanatory power regarding the actual result while controlling for implied winning probabilities. Thus, market prices do not adequately incorporate the information contained in the table rank difference. More specifically, our results suggest that teams overperforming (underperforming) by four or more ranks, which corresponds to a one-standard deviation increase from the mean of zero ranks, are associated with an approximately 4.7 (2.9) percentage points lower (higher) actual winning probability not accounted for in the betting prices.

This finding is mirrored in consistently negative betting returns for bets on previously overperforming teams and consistently positive returns for bets on previously underperforming teams. Moreover, we form a simple betting strategy by betting on underperforming teams and betting against overperforming teams. An out-of-sample backtest of this strategy yields a return of 5.0% before commission costs and 2.2% after commissions. Overall, our findings suggest that the aggregated predictions of participants in betting exchange markets overweight outcome-based performance, which is consistent with outcome bias.

This paper makes several important contributions. First, we extend the previous literature by demonstrating that outcome bias is not limited to individuals but also exists in betting exchange market prices reflecting the aggregated beliefs of many individuals. Second, we contribute to the literature on betting exchange market efficiency by showing that betting prices fail to reflect all available historical information. This finding is in line with previous research finding evidence of informational inefficiency in betting exchange markets (e.g., Angelini et al., 2022) but contradicts Croxson and James Reade (2014), who conclude that betting exchange markets are informationally efficient. Finally, our paper contributes to the literature on sports analytics by showing that process-oriented performance metrics such as expected goals are valuable for evaluating team strength over and above market-based predictions.

The remainder of this paper is structured as follows. In Section 2, we summarize the relevant literature and state our research hypothesis. In Section 3, we describe the data, variables and empirical methods used. In Section 4, we present our results, several robustness tests and an out-of-sample betting strategy. In Section 5, we conclude the paper with a discussion.

2 | RELATED LITERATURE AND HYPOTHESIS

2.1 | Outcome bias

Research from psychology has shown that people often struggle to make decisions under uncertainty and exhibit various cognitive biases that distort not only decision-making but also the evaluation of past decisions (see, e.g., Earl, 1990; Rabin, 1998; Tversky & Kahneman, 1974). One well-established cognitive bias is outcome bias, which refers to the tendency of people to overweight the importance of the outcome when evaluating a past decision or task performance (Baron & Hershey, 1988; Kausel et al., 2019).

When judging the quality of a past decision, an objective evaluator should consider all the information known to the decision maker at the time of the decision to assess whether the decision was optimal. However, the evaluation of decision quality should not depend on randomly or exogenously determined outcomes (Bazerman & Moore, 2012; Hastie & Dawes, 2009). Outcomes should be considered in the evaluation process only if they provide additional information about intentionality, culpability, or characteristics of the actor's personality (Hershey & Baron, 1992; Mazzocco et al., 2004). However, when the outcome information lacks any additional information about the actor, the quality of the decision should not be judged differently depending on the outcome (Gino et al., 2010).

Baron and Hershey (1988) discovered that students are prone to be outcome biased when evaluating medical procedures. The students assessed decisions as more appropriate when the outcome was successful than when it was unsuccessful despite all other information being equal. After the seminal study of Baron and Hershey (1988), the existence of outcome bias was demonstrated by many subsequent studies in various laboratory settings (e.g., Brownback & Kuhn, 2019; Cushman et al., 2009; Gurdal et al., 2013; König-Kersting et al., 2021; Marshall & Mowen, 1993; Mazzocco et al., 2004; Mowen & Stone, 1992; Rubin & Sheremeta, 2016). For instance, Gino et al. (2009, 2010) found that ethically questionable behavior is perceived as more unethical when it produces negative outcomes than when it produces positive outcomes. Consistent results regarding outcome bias have also been found in legal contexts (e.g., Alicke et al., 1994; Mazzocco et al., 2004), salespeople's performance evaluations (Marshall & Mowen, 1993), and investment decisions (König-Kersting et al., 2021; Ratner & Herbst, 2005).

Consistent with insights from the laboratory, several studies have provided field evidence of outcome bias (e.g., Emerson et al., 2010; Meier et al., 2022; Tinsley et al., 2012). Most importantly to our study, several studies document outcome-biased behavior using sports data. Lefgren et al. (2015) found that coaches in the National Basketball Association change their strategy more often after losing a game than after winning a game, even when comparing strategy changes after uninformative narrow wins and losses. Furthermore, Gauriot and Page (2019) investigated whether outcome bias is present in performance evaluations of football (soccer) players. They found that players' shots on the post that resulted in a goal overly influenced the players' evaluations compared to shots that hit the post and missed. Players who scored in a match after hitting the post were not only rewarded by their manager with more play time in upcoming matches but were also rated higher by journalists and sports fans.

Similarly, Kausel et al. (2019) showed that journalists rated players significantly better than their opponents when their team won the penalty shootout despite comparable in-game performance. These findings indicate that desirable outcomes are overly rewarded in subjective performance evaluations. Finally, Flepp and Franck (2021) showed that coach dismissals in football (soccer) lead to a boost in subsequent team performance only when previous match outcomes were the result of actual disappointing performance on the pitch rather than "bad luck." This finding emphasizes the importance of unbiased performance and decision-making evaluations, as coach dismissals are very costly at the top level of football.

Overall, the literature shows that decision-makers have difficulty evaluating the quality of past decisions or performance when random events also influence the final outcome. A favorable outcome often justifies a decision, strategy, or performance, even if there is other evidence indicating otherwise. These studies have demonstrated that individual decision-makers are outcome biased in a variety of different settings, both in the laboratory and in the field. In this paper, we extend the previous literature by analyzing whether these individual-level biases are also present on a market level. To do so, we analyze price efficiency in a betting exchange market environment where prices reflect the aggregated beliefs of numerous individuals.

2.2 | Research hypothesis

As participants in betting exchange markets trade with each other the outcomes of future events, that is, results of football matches in our setting, the market prices reflect the aggregated beliefs of the participants (Brown et al., 2019).

Several studies have demonstrated that these kinds of markets are informationally efficient (e.g., Croxson & James Reade, 2014; Gauriot & Page, 2020). Furthermore, prices from betting exchanges generally forecast results more accurately than prices from bookmakers (Franck et al., 2010) or predictions from opinion polls (Vaughan Williams & Reade, 2016).

The market-level setting of betting exchanges offers much more room to eliminate outcome bias and forms a stronger test for the bias to persist compared to individual-level settings, because even a small fraction of informed bettors could theoretically eliminate any mispricing by generating positive betting returns. If the competitive market mechanism of betting exchanges can overcome the underestimation of randomness in football match outcomes, we expect the following null hypothesis to hold:

H0 *Market prices accurately incorporate the influence of randomness on previous match outcomes.*

However, there is also evidence from the previous literature that prices from betting exchange markets fail to reflect all relevant information (e.g., Abinzano et al., 2016; Angelini et al., 2022; Brown et al., 2018). Thus, if the bettors are outcome biased in the aggregate and underestimate the role played by randomness in football match outcomes, we expect that market prices will reflect this tendency. In particular, we expect that market prices overestimate the winning probability of previously overperforming teams, that is, teams that performed better in terms of match outcomes compared to their task performance on the pitch. Analogously, we expect that market prices underestimate the winning probability of previously underperforming teams, that is, teams that performed worse in terms of match outcomes compared to their task performance on the pitch. Thus, hypothesis *H1* is stated as follows:

H1 *Market prices underestimate the influence of randomness on previous match outcomes in that market prices overestimate the winning probability of overperforming teams and underestimate the winning probability of underperforming teams.*

3 | METHODS

3.1 | Data

We obtained football data from Gracenote, a subsidiary of Nielsen Holdings Plc. Among other things, Gracenote provides sports metadata for a variety of different sports. Our main dataset contains shot information for all 9130 matches from the top five European football leagues for the 2013/2014 through 2017/2018 seasons. More specifically, we have data on the goals and shots of 1530 matches from the German Bundesliga and 1900 matches each from the French Ligue 1, the Italian Serie A, the English Premier League and the Spanish La Liga. For each shot, we know the exact location, the rule setting and the part of the body that was used. Additionally, we obtained data on all 1826 matches from the same leagues for the 2018/2019 season to test whether our results hold in an out-of-sample betting strategy.

The odds data stem from Matchbook, one of the most popular betting exchanges, and were collected from www.oddsportal.com. We use the closing back odds, that is, the latest odds available to bet on a team before the start of the match. The closing back odds thus reflect the latest market offers at the betting exchange.[1] Due to missing odds data on www.oddsportal.com for 156 matches, our sample of matches with odds data reduces to 8974.

3.2 | League table based on expected goals

The expected goals metric is based on quantified scoring chances rather than actual goals and has been shown to better predict future sporting results than metrics based on match outcomes (Anzer & Bauer, 2021; Brechot & Flepp, 2020). We follow Brechot and Flepp (2020) in estimating the scoring probabilities of shots based on the distance, angle, rule setting of the shot (i.e., open play, free kick or penalty kick), and body part used. Additionally, we include team and opposing team fixed effects in the logistic regression to account for unobserved team quality characteristics, such as defensive or goal scoring skills. In total, we estimate the scoring probability of 214,194 shots from all 9130 matches in our sample. We outline the logit regression model used to estimate the scoring probability of each shot and briefly describe the variables employed in the appendix.

We then aggregate the estimated scoring probabilities for each team within each match to derive the number of expected goals per match. For instance, if Manchester United had 5 shots against Leicester City with expected scoring probabilities of .80, .70, .50, .25, and .15, the expected goals for Manchester United would be 2.40. For each match, we calculate the expected goals for both teams as well as the expected goal difference (xGD). In the example, if Manchester United's expected goals were 2.40 and Leicester City's were 0.80, Manchester United's expected goal difference is equal to 1.60 (2.40–0.80), and Leicester City's xGD is equal to −1.60. Thus, the expected goal metric allows us to evaluate a team's task performance on the pitch in terms of creating valuable scoring chances independent of actual match outcomes.[2]

Finally, we rank the teams according to their cumulative xGD and construct a league table based on expected goals (xGT). The rank in the xGT should reflect a team's playing quality on the pitch more accurately than the rank in the OLT because the OLT is based solely on actual match outcomes, in which randomness in match outcomes fully translates into the ranking. For example, Manchester United could play well on the pitch in terms of expected goals because it created scoring chances of higher total value than its opponent. However, Manchester United might actually lose the match in terms of outcomes because the scoring chances did not translate into actual goals. In such situations, the rank in the xGT should be higher than the rank in the OLT. Conversely, in situations when the team played poorly on the pitch but won the match, the rank in the xGT should be worse than the rank in the OLT.

3.3 | Variables of interest

We use the difference between the rank in the xGT and the rank in the OLT to test whether market prices correctly include overperformance and underperformance in terms of match outcomes compared to performance based on expected goals. Specifically, we define our main variable of interest as follows:

$$\textit{rank difference}_{i,j,k} = \textit{rank xGT}_{i,j-1,k} - \textit{rank OLT}_{i,j-1,k} \tag{1}$$

where i denotes a team, j denotes the match week and k denotes the season. Thus, a positive value for *rank difference* indicates that a team has been overperforming, as its rank in the OLT is better than its rank in the xGT. Conversely, a negative value for *rank difference* indicates that the team has been underperforming, as its rank in the OLT is worse than in the xGT.

Based on the rank difference variable, we form the binary subvariables *overperformance of xG*, *underperformance of xG* and *similar performance* to further distinguish the impact of previously over- and underperforming teams on informational efficiency. In our preferred specification, we derive the threshold for over- and underperformance based on the standard deviation of *rank difference,* which is approximately 3.8. The rationale behind this choice is that values within one standard deviation of the mean (i.e., zero rank difference) are fairly common, and both table ranks indicate a similar performance. Thus, we define the variable *overperformance of xG* equal to one if the rank difference is larger or equal to 4 ranks and zero otherwise. Analogously, we define the variable *underperformance of xG* equal to one if the rank difference is smaller or equal to −4 ranks and zero otherwise. Formally, the variables are defined as follows:

$$\textit{overperformance of xG}_{i,j,k} = \begin{cases} 1, \textit{if rank difference}_{i,j,k} \geq 4 \\ 0, \text{otherwise} \end{cases} \tag{2}$$

$$\textit{similar performance}_{i,j,k} = \begin{cases} 1, \textit{if} -4 < \textit{rank difference}_{i,j,k} < 4 \\ 0, \text{otherwise} \end{cases} \tag{3}$$

$$\textit{underperformance of xG}_{i,j,k} = \begin{cases} 1, \textit{if rank difference}_{i,j,k} \leq -4 \\ 0, \text{otherwise} \end{cases} \tag{4}$$

3.4 | Statistical methods

As in previous research, for example, Brown et al. (2018), Bizzozero et al. (2016), and Choi and Hui (2014), we calculate prices as the reciprocal of the odds. The price is the amount of money one must bet to collect 1 unit if the bet wins.

Thus, the price can also be seen as an implied winning probability. For instance, if the betting odds are 2.0 for a team to win the match, then the price would be $\frac{1}{2.00} = 0.5$, which also denotes the implied winning probability of the team. We define the implied winning probability for each team in each match as $impliedprob = \frac{1}{odds}$.

Betting exchange markets are efficient if the market prices reflect all historical information and are the best forecasts of the outcome of a match (Angelini & De Angelis, 2019). Consequently, no other variable should have explanatory power regarding the match outcome after the implied winning probabilities are controlled for. Thus, following previous research, for example, Brown et al. (2018), Forrest and Simmons (2008), and Franck et al. (2011), we estimate two logit regression models as follows:

$$Ln\left[\frac{P(win_{i,j,k}=1)}{P(win_{i,j,k}=0)}\right] = \beta_0 + \beta_1 impliedprob_{i,j,k} + \beta_2 rank\ difference_{i,j,k} + \beta_3 home_{i,j,k} + e_{i,j,k} \tag{5}$$

$$\begin{aligned} Ln\left[\frac{P(win_{i,j,k}=1)}{P(win_{i,j,k}=0)}\right] &= \beta_0 + \beta_1 impliedprob_{i,j,k} + \beta_2 overperformance\ of\ xG_{i,j,k} \\ &+ \beta_3 underperformance\ of\ xG_{i,j,k} + \beta_4 home_{i,j,k} + e_{i,j,k} \end{aligned} \tag{6}$$

where the dependent variable *win* indicates whether the bet was successful, that is, 1 for a winning bet and 0 for a losing bet.[3] If the bet is on the home team to win the match, the dependent variable denotes whether the home team has won the match, and if the bet is on the away team, the dependent variable denotes whether the away team has won the match. The independent variable *impliedprob* denotes the winning probability implied by the betting odds, *home* is a dummy variable controlling for potential home team bias (see, e.g., Forrest & Simmons, 2008), and *rank difference*, *overperformance of xG* and *underperformance of xG* are the variables of interest. Because we simultaneously include bets on the home team and on the away team in the main model, we follow previous research (see, e.g., Brown et al., 2018; Forrest & Simmons, 2008) and compute clustered heteroscedasticity-robust standard errors at the match level.

Under the null hypothesis *H0*, we expect efficient betting prices and the noisy match outcome information to be correctly incorporated into *impliedprob*. In other words, we expect only the coefficient β_1 to be statistically significant, while the coefficients of *home* and *rank difference* in regression (5) and *home*, *overperformance of xG* and *underperformance of xG* in regression (6) should be equal to zero. If bettors are outcome biased at the aggregate level, we expect betting prices to overestimate the winning probabilities of previously overperforming teams and underestimate the winning probabilities of previously underperforming teams. Thus, under *H1*, we expect a negative and significant coefficient for *rank difference* in regression model (5) and a significantly negative (positive) coefficient for *overperformance of xG* (*underperformance of xG*) in regression model (6).

4 | RESULTS

4.1 | Main results

Table 1 displays summary statistics for the variables *win*, *impliedprob*, *home*, *rank difference*, *overperformance of xG* and *underperformance of xG*. Table 1 shows that the implied winning probabilities accurately mirror the actual winning probabilities on average.[4] The *rank difference* variable has a mean of zero, as for each team ranked better in the expected goals table, at least one other team must be ranked worse.

The results of the logit regressions are depicted in Table 2. The results are shown in the form of marginal effects measured at a point where the variables are set to their means. In Column (1), we depict the results of regression model (5) using the *rank difference* variable. As expected, the sign of the *impliedprob* variable is positive and significant at the 1% significance level. Furthermore, the coefficient of the *home* variable is nonsignificant, indicating that no home-team bias is present. Interestingly, however, the sign of the *rank difference* coefficient is negative and significant at the 1% significance level. This implies that market prices fail to correctly incorporate the information contained in the difference between outcome-based performance and performance based on expected goals. More specifically, the coefficient of *rank difference* indicates that a one-rank increase in *rank difference* leads to a 0.4% point decrease in the probability of *win* taking the value of 1. As positive values of *rank difference* indicate that a team was overperforming in terms of match outcomes compared to their performance based on xG, the results of Column (1) show that teams

TABLE 1 Summary statistics.

Variable	*N*	Mean	SD	Min	Max
win (0/1)	18,260	0.377	0.485	0.000	1.000
impliedprob	17,948	0.378	0.202	0.007	0.990
home	18,260	0.500	0.500	0.000	1.000
rank difference	17,770	0.000	3.797	−16.000	18.000
overperformance of xG	17,770	0.147	0.354	0.000	1.000
underperformance of xG	17,770	0.153	0.360	0.000	1.000

Note: The table displays the summary statistics. The variables *rank difference*, *overperformance of xG* and *underperformance of xG* are lagged and the first match week of each season is set to missing.

TABLE 2 Main results.

	win (0/1)	
	(1)	(2)
impliedprob	1.124*** (0.028)	1.122*** (0.028)
rank difference	−0.004*** (0.001)	
overperformance of xG		−0.047*** (0.012)
underperformance of xG		0.029*** (0.011)
home	0.008 (0.011)	0.009 (0.011)
Number of observations	17,484	17,484
Number of clusters	8742	8742
Pseudo R^2	.145	.146

Note: The table reports the marginal effects from a logit regression. The dependent variable *win* denotes the actual outcome of a bet on a team to win the match (1 for a winning bet and 0 for a losing bet). Column (1) shows the results of the *rank difference* variable, and Column (2) shows the results of the variables *overperformance of xG* and *underperformance of xG*. *Home* controls for bets on teams that play at home. The heteroscedasticity-robust and clustered standard errors at the match level are reported in parentheses.

***, **, and * denote significance at the 1-, 5-, and 10-percent levels, respectively.

overperforming in the past are associated with a lower probability of winning than that implied by the prices. In other words, bets on teams that were overperforming in the past are overvalued by the bettors.

In Column (2) of Table 2, we depict the results of regression model (6), where we analyze the impact of considerable over- and underperformance of expected goals separately. As before, the coefficient of *impliedprob* is positive and significant at the 1% significance level, while *home* has no effect. More importantly, the coefficients of *overperformance of xG* and *underperformance of xG* are both statistically significant after the implied winning probabilities are controlled for. As expected, the sign of *overperformance of xG* is negative, indicating that prices of bets on teams that were overperforming by 4 or more table ranks in the past overestimate the true winning probabilities. Conversely, the sign of *underperformance of xG* is positive, indicating that the prices of bets on teams that were underperforming in the past underestimate the true winning probabilities. The marginal effects at the means show that teams that were previously overperforming (underperforming) by 4 or more ranks are associated with a 4.7 (2.9) percentage point decrease (increase) in the probability of *win* taking the value of 1 not accounted for in the implied winning probabilities of prices.

TABLE 3 Robustness tests of the main model.

	win (0/1)				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
impliedprob	1.110*** (0.027)	1.010*** (0.020)		1.124*** (0.033)	1.146*** (0.035)
impliedprob adj.			1.139*** (0.028)		
rank difference	−0.004*** (0.001)	−0.004*** (0.001)	−0.004*** (0.001)	−0.003** (0.001)	−0.005*** (0.001)
home	0.008 (0.010)	0.006 (0.009)	0.008 (0.011)	0.007 (0.012)	0.006 (0.011)
Team FE	-	-	-	-	Yes
Number of observations	17,484	17,484	17,484	8742	17,484
Number of clusters	8742	8742	8742		8742
Pseudo R^2/R^2	.146	.183	.146	.144	.152
Model	Probit	OLS	Logit	Logit	Logit

Note: The table reports the marginal effects from logit, probit and OLS regressions. The dependent variable *win* denotes the actual outcome of a bet on a team to win the match (1 for a winning bet and 0 for a losing bet). Column (1) displays the results obtained using a probit model, and Column (2) displays the results using an OLS model. Column (3) displays the results using adjusted implied probabilities, and Column (4) shows the results of randomly choosing a bet on one team per match. Column (5) shows the results including team fixed effects. *Home* controls for bets on teams that play at home. The heteroscedasticity-robust and clustered standard errors at the match level (except for Column [4]) are reported in parentheses.

Abbreviations: FE, fixed effects; OLS, ordinary least squares.

***, **, and * denote significance at the 1-, 5-, and 10-percent levels, respectively.

In conclusion, the null hypothesis of market efficiency is rejected, as relevant information regarding the difference in outcome-based performance and performance based on expected goals is not fully incorporated into market prices. Rather, our results support hypothesis *H1* and indicate that teams that were overperforming in the past are systematically overvalued, while teams that were underperforming in the past are systematically undervalued by bettors, which is consistent with outcome bias.[5]

4.2 | Robustness tests

To test the robustness of the results, we conducted several variations of our main model. The results of our first series of tests are shown in Table 3. We start by estimating probit and standard ordinary least squares regressions instead of logit regressions. Columns (1) and (2) in Table 3 show that the coefficients remain unchanged both in terms of magnitude and statistical significance. Second, following Forrest and Simmons (2008), we adjusted the implied probabilities of all bets on a match, that is, home win, away win and draw, so that they sum to one. Column (3) shows that our results are again insensitive to this alteration. Third, we randomly selected a bet on one team per match instead of using clustered standard errors and again obtained consistent results (Column (4) in Table 3).[6]

Finally, to address the concern that our results might be driven by systematic over- and underperformance due to limitations of the expected goals model employed, we additionally included team fixed effects in our main regression models. If systematic differences in team strengths were the main driver of our results, we would expect that the inclusion of team fixed effects would reduce the magnitude of the *rank difference* coefficient. However, as shown in Column (5) of Table 3, including team fixed effects increases the magnitude of *rank difference* marginally and thus does not change our main conclusions.[7]

In a second series of robustness tests, we employed alternative measures for the rank difference and show the results in Table 4. First, in our main analysis, we do not account for the over- or underperformance of the opposing team. However, the participants in the betting exchange market must evaluate the expected performance of the focal team relative to the opponent. Thus, we construct the new variable *difference in rank difference* defined as the difference between the rank difference of the focal team and the rank difference of the opposing team. Therefore, higher values of

TABLE 4 Alternative measures for rank difference.

	***win* (0/1)**		
	(1)	**(2)**	**(3)**
impliedprob	1.120*** (0.028)	1.126*** (0.028)	1.124*** (0.029)
difference in rank difference	−0.004*** (0.001)		
rank difference alternative		−0.004*** (0.001)	
last three matches			−0.004** (0.002)
home	0.009 (0.011)	0.008 (0.011)	0.011 (0.011)
Number of observations	17,484	17,484	16,568
Number of clusters	8742	8742	8286
Pseudo R^2	.146	.145	.146

Note: The table reports the marginal effects from a logit regression. The dependent variable *win* denotes the actual outcome of a bet on a team to win the match (1 for a winning bet and 0 for a losing bet). Column (1) displays the results for the difference in rank difference between the focal team and the opposing team. Column (2) shows the results obtained using an alternative approach to calculate the difference in table ranks (*rank difference alternative*), and Column (3) shows the results obtained using only the last three matches to measure discrepancies between outcome-based performance and performance based on xG. *Home* controls for bets on teams that play at home. The heteroscedasticity-robust and clustered standard errors at the match level reported in parentheses.

***, **, and * denote significance at the 1-, 5-, and 10-percent levels, respectively.

difference in rank difference indicate more relative overperformance, whereas lower values indicate more relative underperformance. The results are displayed in Column (1) of Table 4 and remain similar.

Second, we calculated an alternative rank difference measure (*rank difference alternative)* by constructing a variation of the league table based on expected goals. Instead of taking the cumulative expected goal differences to form the table ranking, we formed thresholds for expected goal differences to determine whether the match outcome was considered a win, draw or loss. Following Flepp and Franck (2021), we considered an expected goal difference >0.5 to be a win, an expected goal difference <0.5 to be a loss and everything in between a draw. We then allocated points as in the OLT, that is, 3 points for a win, 1 point for a draw and 0 points for a loss. One potential upside of this approach is that it uses the same values for points as the OLT. Thus, a ranking of teams by points assigned based on expected goals might be more closely comparable to the OLT ranking. However, as displayed in Column (2) of Table 4, the results are again very similar in sign, magnitude, and significance to those of the main specification.

Finally, because one might be concerned that the table positions are ossified toward the end of the season, we measured over- and underperformance of expected goals based on the last three matches instead of using the table rank differences. Specifically, we subtracted the number of points won based on expected goals over the past three matches from the actual number of points won over the past three matches and labeled this difference *last three matches*. A positive value for the *last three matches* variable indicates that a team has won more actual points than points based on expected goals in the past three matches and vice versa for negative values. Column (3) of Table 4 shows, consistent with the results of the main specification, that a larger difference between recent outcome-based performance and performance based on xG is associated with a lower winning probability not accounted for in the implied winning probability of prices.

Instead of using a difference of 4 ranks to determine over- and underperformance, we further tested several alternative threshold values, that is, 2/−2, 3/−3, 5/−5 and 6/−6, to define the variables *overperformance of xG* and *underperformance of xG* variables. The results are depicted in Table A1 in the appendix. Overall, the results remain qualitatively similar to those of the main specification and the magnitudes of the coefficients of interest tend to increase with larger absolute rank difference thresholds.

TABLE 5 Betting returns comparisons.

rank difference threshold	*N*	Mean	SE	*t*-statistic
Panel A: Returns for bets on overperforming teams				
2	5219	−0.061	0.026	−2.340***
3	3752	−0.094	0.030	−3.139***
4	2576	−0.133	0.034	−3.914***
5	1787	−0.123	0.042	−2.927***
6	1226	−0.121	0.048	−2.526***
Panel B: Returns for bets on underperforming teams				
−2	5129	0.046	0.025	1.869**
−3	3785	0.073	0.029	2.506***
−4	2684	0.089	0.035	2.550***
−5	1879	0.035	0.039	0.879
−6	1270	0.110	0.051	2.161**

Note: The table displays the returns on bets with a stake equaling 1. Panel A shows the returns for bets on overperforming teams, and Panel B shows the returns for bets on underperforming teams. The *t*-statistics refer to a *t*-test of zero mean returns. Statistical significance refers to a one-sided test of negative returns in Panel A and positive returns in Panel B.

***, **, and * denote significance at the 1-, 5-, and 10-percent levels, respectively.

4.3 | Betting returns comparisons

To investigate whether the inefficiencies found are large enough to translate into economic profits, we compare the betting returns depending on the degree of over- and underperformance. Because we have shown that the implied probabilities for previously underperforming teams are too low, betting on these teams should yield more positive returns. Conversely, the implied probabilities for previously overperforming teams are too high, indicating that betting on those teams should result in more negative returns. We calculated betting returns on one-unit bets as follows:

$$return_i = \begin{cases} odds_i - 1 \quad if \;\; bet\ is\ successful \\ -1 \quad if \;\; bet\ is\ unsuccessful \end{cases} \tag{7}$$

Table 5 displays the betting returns on teams for various thresholds of *rank difference*. The results show that the returns on bets on previously overperforming teams are strictly negative, while the returns on bets on previously underperforming teams are strictly positive.

The returns on bets on previously overperforming teams show a slight downward trend for larger rank difference thresholds, ranging from −6.1% for a rank difference threshold of 2% to −12.1% for a threshold of 6. In contrast, the returns on bets on previously underperforming teams tend to increase with a larger absolute rank difference, ranging from 4.6% for a threshold of 2 ranks to 11.0% for a threshold of 6 ranks. This finding indicates that the more a team overperformed in the past, the more it is overvalued by bettors, and conversely, the more a team underperformed in the past, the more it is undervalued by bettors. Single-sample *t*-tests show that the returns on overperforming teams are significantly negative using a one-sided test. In contrast, returns on underperforming teams are significantly positive for all but one threshold.[8]

4.4 | Out-of-sample betting strategy

We used the 2018/2019 season to externally test a betting strategy built upon the previous insights. Employing the same methods as in the main analysis, we estimated the expected goals for each team using data from the preceding 2017/

TABLE 6 Out-of-sample betting returns.

***rank difference* threshold**	***N***	**Mean**	**SD**
Panel A: Returns for bets on overperforming teams			
2	1184	0.010	1.888
3	864	−0.036	1.742
4	652	−0.074	1.804
5	470	−0.138	1.677
6	342	−0.128	1.770
Panel B: Returns for bets on underperforming teams			
−2	1120	0.041	2.096
−3	872	0.079	2.216
−4	653	0.047	2.093
−5	470	0.057	2.037
−6	328	0.089	2.205

Note: The table displays the out-of-sample returns on bets with a stake equaling 1. Panel A shows the returns for bets on overperforming teams, and panel B shows the returns for bets on underperforming teams.

TABLE 7 Betting strategy returns.

	N	**Mean**	**SD**	**Avg. limit order volume**
Panel A: Returns before commission				
Backing underperforming teams (*rank difference* ≤−4)	653	0.047	2.093	1884
Laying overperforming teams (*rank difference* ≥4)	652	0.054	1.856	1879
Combined	1305	0.050		1882
Panel B: Returns after commission				
Backing underperforming teams (*rank difference* ≤−4)	653	0.019	2.023	1884
Laying overperforming teams (*rank difference* ≥4)	652	0.026	1.842	1879
Combined	1305	0.022		1882

Note: The table displays the returns on back and lay bets with a stake equaling 1. Panel A shows the returns before commission, and panel B shows the returns after a commission rate of 4% on all winning has been deducted. The average limit order volume is in Euros.

2018 season and derived the variable *rank difference*. Table 6 shows the out-of-sample betting returns for various rank difference thresholds. As in the previous analysis, the returns are mainly negative for teams characterized as overperforming and positive for teams characterized as underperforming. Moreover, the absolute values of the returns seem to increase depending on the degree of over- and underperformance.

On that basis, we derive a betting strategy to exploit this finding. The simplest way to profit from this finding is to bet on, that is, to "back," all teams characterized as underperforming and bet against, that is, to "lay," all teams characterized as overperforming. Table 7 summarizes the returns on this simple strategy if we use the threshold difference of 4/-4 ranks for the variables *overperformance of xG* and *underperformance of xG* as in our main specification.

By strictly backing all teams that were considerably underperforming in the past and laying all teams that were considerably overperforming in the past, a positive return of 5.0% could be achieved before commission costs, and a return of 2.2% could be achieved after the standard commission fee of 4% for winning bets.[9] Importantly, the average available limit order per betting opportunity, that is, the maximal amount that could have been wagered on each bet for the observed odds, was approximately 1800 Euros. Overall, the size of the achieved return seems plausible and striking, especially considering the simplicity of the deployed betting strategy.

5 | DISCUSSION AND CONCLUSION

We show that betting prices from overperforming teams overstate the winning probability in subsequent matches, while prices from underperforming teams understate the winning probability in subsequent matches. This pattern is reflected in higher returns on bets on previously underperforming teams and lower returns on bets on previously overperforming teams, and a simple betting strategy based on this finding leads to a net return of 2.2% in an out-of-sample backtest. Overall, these findings suggest that bettors overweight the information contained in previous outcomes of football matches, leading to informationally inefficient markets.

Our findings further suggest that the inefficiency found in betting prices persists within our sample period. This contradicts other studies investigating betting market inefficiency and finding that market inefficiency is short-lived. For example, Deutscher et al. (2018) find that mispricing of recently promoted teams disappears after mid-season. Furthermore, Meier et al. (2021) show that the effect of ghost games on home advantage was incorrectly incorporated into betting prices. However, this inefficiency disappeared after approximately 5 weeks.

We suspect that the continual inefficiency found in our paper could be due to the following two reasons. First, outcome bias has been shown to be persistent in previous studies investigating individual decision-making (e.g., König-Kersting et al., 2021; Lefgren et al., 2015). Even in laboratory environments with perfect information, individual judgments are outcome biased and cannot easily be eliminated by, for example, using independent third parties (Brownback & Kuhn, 2019). Second, even if a team's task performance on the pitch is observable on TV broadcasts, the availability of detailed data of within-match events such as shots is restricted. Thus, without a systematic analysis of such data, betting exchange market participants might be unable, even over time, to correctly judge the random component in match outcomes. However, as aggregated match data on expected goals becomes more widespread and accessible, outcome bias might decrease in the future.

Several limitations of our paper should be considered when interpreting the results. First, our expected goals model is subject to estimation error. Most importantly, our employed model does not account for defensive pressure or the goalkeeper's position at the time of the shot. Furthermore, we only account for time-constant differences in quality between teams but not for differences in quality between players and goalkeepers. Thus, some teams might systematically overperform in terms of match outcomes, while other teams systematically underperform. However, as including team fixed effects in our main regression model leads to more pronounced estimates for the rank difference, we are confident that systematic errors in the expected goal metric are not the main source of our findings.

Second, teams may become more confident after lucky wins, which enhances their playing strength in subsequent games. A similar pattern has, for example, been documented by Rosenqvist and Skans (2015) in professional golf. Thus, the effect of match outcomes on a team's confidence may act as a countervailing force to outcome bias in betting exchange prices. If, however, betting exchange market participants overestimate the reaction of teams to lucky match outcomes, our estimates of outcome bias might be partly attributable to this error.

Finally, as research has shown that coaches' strategic decisions are also affected by outcome bias (e.g., Gauriot & Page, 2019; Lefgren et al., 2015), biases in betting prices might arise due to unanticipated changes in strategic decisions. However, the betting odds in our data are “closing odds” taken just seconds before the match starts. Thus, all observable information regarding important strategic decisions of coaches (e.g., starting line-up or team formation) is also available to bettors. Moreover, one would expect that outcome bias in coaching decisions should result in inferior playing strength following both overperformance and underperformance, which is difficult to explain with our finding that teams win more than expected by the betting exchange market after underperforming in terms of match outcomes.

Our findings have several important implications for practice. Outcome bias seems to be far-reaching, as even a competitive market mechanism fails to completely eliminate it. This might be due to a large fraction of casual or fan bettors that are subject to outcome bias, which prevents informed bettors from trading sufficiently against them and eliminating mispricing. Consequently, outcome bias might impact not only the efficiency of market prices in the betting exchange market but also in other market settings. For example, Heuer et al. (2017) experimentally show that even sophisticated retail investors seem to misperceive luck for skill when investing in mutual funds. If such misperceptions are widespread across investors, they might also aggregate into mispricing. Finally, our findings suggest that performance evaluation in football is fundamentally outcome biased. This could impair not only the predictions of future team performance but also player valuation or head coach evaluation. Thus, fostering the development of more process-oriented performance evaluation measures seems crucial to improve decision-making in the football industry.

ACKNOWLEDGMENTS

We thank Alex Krumer, Robert Simmons, the participants at the ESEA Conference on Sport Economics, the participants at the research seminar of the Department of Business Administration at the University of Zurich, and three anonymous referees for their helpful comments.

Open access funding provided by Universitat Zurich.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from Gracenote. Restrictions apply to the availability of these data, which were used under license for this study. Details for what data was used and how the data was analyzed can be found in openICPSR at https://doi.org/10.3886/E187181V2, reference number openicpsr-187181 (Flepp et al., 2023).

ORCID

Raphael Flepp https://orcid.org/0000-0002-4183-3557
Oliver Merz https://orcid.org/0000-0001-5266-1909
Egon Franck https://orcid.org/0000-0003-1280-6864

ENDNOTES

[1] The amount of money that could have been wagered using the closing odds depends on the limit order volumes offered by other bettors. However, betting exchange odds on top European teams do not seem to be subject to a thin market problem. For example, the amount of money that could have been wagered on the closing back odds for the home team to win is approximately EUR 3500 on average.

[2] One might be concerned that the xGD of a match depends on the timing of the actual goals that occurred during a particular match due to tactical changes of the teams. To address this concern, we checked whether the xGD measure is systematically different for matches in which the first goal was scored early or late in the game. We find that the time of the first goal has no statistically significant impact on the xGD of a match.

[3] We opted for the logit model because the estimated probabilities are bounded between zero and one and several previous studies investigating betting market efficiency similarly employed logit regression models (e.g., Deutscher et al., 2018; Merz et al., 2021).

[4] Censoring our sample at the 1% (5%) quantiles to address the concern that our findings might be driven by extreme values of *impliedprob* does not change our results.

[5] To substantiate our main results based on the odds from a betting exchange, we followed Deutscher et al. (2018) and collected the betting odds from the bookmaker bet365. Rerunning our main analyses using bookmaker odds leads to similar results.

[6] Deutscher et al. (2018) show that bookmakers underestimate the winning probabilities of newly promoted teams in the German Bundesliga at the start of the season. To test whether our results are driven by this inefficiency in the bookmaker market, we excluded all games of promoted teams played in the first half of each season from our analysis in an additional robustness check. The results remain virtually unchanged. Furthermore, as changes in head coaches might affect bettor expectations and perceptions about team performance and thus the accuracy of betting odds, we excluded the first match after each within-season head coach change in an auxiliary robustness test. Our results remain robust to this exclusion.

[7] To control for potential differences across football leagues affecting market efficiency, we also included league dummy variables in our estimation models. None of the league dummies were statistically significant and our main results remained unaffected. Furthermore, controlling for the match day within a season does not alter our results. Finally, we excluded all games of the teams at the top and bottom of the official league table because teams at the top will never be classified as underperforming and teams at the bottom will never be classified as overperforming. Our results remain robust after this exclusion.

[8] The positive betting returns are not concentrated on a few teams with similar characteristics. For example, 25 out of the potential 28 English Premier League teams fulfill the rank difference threshold of minus 4 or less at least once in our sample.

[9] In betting exchange markets, there is a spread between back and lay odds, commonly referred to as transaction costs. Because we are using data on back and lay odds, the transaction costs are already considered.

REFERENCES

Abinzano, I., Muga, L. & Santamaria, R. (2016) Game, set and match: the favourite-long shot bias in tennis betting exchanges. *Applied Economics Letters*, 23(8), 605–608. Available from: https://doi.org/10.1080/13504851.2015.1093074

Alicke, M.D., Davis, T.L. & Pezzo, M.V. (1994) A posteriori adjustment of a priori decision criteria. *Social Cognition*, 12(4), 281–308. Available from: https://doi.org/10.1521/soco.1994.12.4.281

Allison, S.T., Mackie, D.M. & Messick, D.M. (1996) Outcome biases in social perception: implications for dispositional inference, attitude change, stereotyping, and social behavior. In: Zanna, M.P. (Ed.) *Advances in experimental social psychology*, 28. San Diego, CA: Academic Press, pp. 53–93.

Angelini, G. & De Angelis, L. (2019) Efficiency of online football betting markets. *International Journal of Forecasting*, 35(2), 712–721. Available from: https://doi.org/10.1016/j.ijforecast.2018.07.008

Angelini, G., De Angelis, L. & Singleton, C. (2022) Informational efficiency and behaviour within in-play prediction markets. *International Journal of Forecasting*, 38(1), 282–299. Available from: https://doi.org/10.1016/j.ijforecast.2021.05.012

Anzer, G. & Bauer, P. (2021) A goal scoring probability model for shots based on synchronized positional and event data in football (soccer). *Frontiers in Sports and Active Living*, 3, 624475. Available from: https://doi.org/10.3389/fspor.2021.624475

Baron, J. & Hershey, J.C. (1988) Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 54(4), 569–579. Available from: https://doi.org/10.1037/0022-3514.54.4.569

Bazerman, M.H. & Moore, D.A. (2012) *Judgment in managerial decision making*. Hoboken, NJ: John Wiley & Sons.

Bizzozero, P., Flepp, R. & Franck, E. (2016) The importance of suspense and surprise in entertainment demand: evidence from Wimbledon. *Journal of Economic Behavior & Organization*, 130, 47–63. Available from: https://doi.org/10.1016/j.jebo.2016.07.006

Brechot, M. & Flepp, R. (2020) Dealing with randomness in match outcomes: how to rethink performance evaluation in European club football using expected goals. *Journal of Sports Economics*, 21(4), 335–362. Available from: https://doi.org/10.1177/1527002519897962

Brown, A., Rambaccussing, D., Reade, J.J. & Rossi, G. (2018) Forecasting with social media: evidence from tweets on soccer matches. *Economic Inquiry*, 56(3), 1748–1763. Available from: https://doi.org/10.1111/ecin.12506

Brown, A., Reade, J.J. & Vaughan Williams, L. (2019) When are prediction market prices most informative? *International Journal of Forecasting*, 35(1), 420–428. Available from: https://doi.org/10.1016/j.ijforecast.2018.05.005

Brownback, A. & Kuhn, M.A. (2019) Understanding outcome bias. *Games and Economic Behavior*, 117, 342–360. Available from: https://doi.org/10.1016/j.geb.2019.07.003

Choi, D. & Hui, S.K. (2014) The role of surprise: understanding overreaction and underreaction to unanticipated events using in-play soccer betting market. *Journal of Economic Behavior & Organization*, 107, 614–629. Available from: https://doi.org/10.1016/j.jebo.2014.02.009

Croxson, K. & James Reade, J. (2014) Information and efficiency: goal arrival in soccer betting. *The Economic Journal*, 124(575), 62–91. Available from: https://doi.org/10.1111/ecoj.12033

Cushman, F., Dreber, A., Wang, Y. & Costa, J. (2009) Accidental outcomes guide punishment in a "trembling hand" game. *PLoS One*, 4(8), e6699. Available from: https://doi.org/10.1371/journal.pone.0006699

Deutscher, C., Frick, B. & Ötting, M. (2018) Betting market inefficiencies are short-lived in German professional football. *Applied Economics*, 50(30), 3240–3246. Available from: https://doi.org/10.1080/00036846.2017.1418082

Earl, P.E. (1990) Economics and psychology: a survey. *The Economic Journal*, 100(402), 718–755. Available from: https://doi.org/10.2307/2233656

Emerson, G.B., Warme, W.J., Wolf, F.M., Heckman, J.D., Brand, R.A. & Leopold, S.S. (2010) Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. *Archives of Internal Medicine*, 170(21), 1934–1939. Available from: https://doi.org/10.1001/archinternmed.2010.406

Flepp, R. & Franck, E. (2021) The performance effects of wise and unwise managerial dismissals. *Economic Inquiry*, 59(1), 186–198. Available from: https://doi.org/10.1111/ecin.12924

Flepp, R., Merz, O., & Franck, E. (2023). ECIN replication package for "When the league table lies: does outcome bias lead to informationally inefficient markets?" Available from: https://doi.org/10.3886/E187181V2

Forrest, D. & Simmons, R. (2008) Sentiment in the betting market on Spanish football. *Applied Economics*, 40(1), 119–126. Available from: https://doi.org/10.1080/00036840701522895

Franck, E., Verbeek, E. & Nüesch, S. (2010) Prediction accuracy of different market structures—bookmakers versus a betting exchange. *International Journal of Forecasting*, 26(3), 448–459. Available from: https://doi.org/10.1016/j.ijforecast.2010.01.004

Franck, E., Verbeek, E. & Nüesch, S. (2011) Sentimental preferences and the organizational regime of betting markets. *Southern Economic Journal*, 78(2), 502–518. Available from: https://doi.org/10.4284/0038-4038-78.2.502

Gauriot, R. & Page, L. (2019) Fooled by performance randomness: overrewarding luck. *The Review of Economics and Statistics*, 101(4), 658–666. Available from: https://doi.org/10.1162/rest_a_00783

Gauriot, R. & Page, L. (2020) How market prices react to information: evidence from a natural experiment. Working Paper, New York University Abu Dhabi.

Gino, F., Moore, D.A. & Bazerman, M.H. (2009) No harm, no foul: the outcome bias in ethical judgments. Working paper, Harvard University.

Gino, F., Shu, L.L. & Bazerman, M.H. (2010) Nameless+ harmless= blameless: when seemingly irrelevant factors influence judgment of (un) ethical behavior. *Organizational Behavior and Human Decision Processes*, 111(2), 93–101. Available from: https://doi.org/10.1016/j.obhdp.2009.11.001

Gurdal, M.Y., Miller, J.B. & Rustichini, A. (2013) Why blame? *Journal of Political Economy*, 121(6), 1205–1247. Available from: https://doi.org/10.1086/674409

Hastie, R. & Dawes, R.M. (2009) *Rational choice in an uncertain world: the psychology of judgment and decision making*. Los Angeles, CA: Sage Publications.

Hershey, J.C. & Baron, J. (1992) Judgment by outcomes: when is it justified? *Organizational Behavior and Human Decision Processes*, 53(1), 89–93. Available from: https://doi.org/10.1016/0749-5978(92)90056-d

Heuer, J., Merkle, C. & Weber, M. (2017) Fooled by randomness: investor perception of fund manager skill. *Review of Finance*, 21(2), 605–635. Available from: https://doi.org/10.1093/rof/rfw011

Kausel, E.E., Ventura, S. & Rodríguez, A. (2019) Outcome bias in subjective ratings of performance: evidence from the (football) field. *Journal of Economic Psychology*, 75, Article 102132. Available from: https://doi.org/10.1016/j.joep.2018.12.006

König-Kersting, C., Pollmann, M., Potters, J. & Trautmann, S.T. (2021) Good decision vs. good results: outcome bias in the evaluation of financial agents. *Theory and Decision*, 90(1), 31–61. Available from: https://doi.org/10.1007/s11238-020-09773-1

Lefgren, L., Platt, B. & Price, J. (2015) Sticking with what (barely) worked: a test of outcome bias. *Management Science*, 61(5), 1121–1136. Available from: https://doi.org/10.1287/mnsc.2014.1966

Marshall, G.W. & Mowen, J.C. (1993) An experimental investigation of the outcome bias in salesperson performance evaluations. *Journal of Personal Selling and Sales Management*, 13(3), 31–47.

Mazzocco, P.J., Alicke, M.D. & Davis, T.L. (2004) On the robustness of outcome bias: no constraint by prior culpability. *Basic and Applied Social Psychology*, 26(2-3), 131–146. Available from: https://doi.org/10.1207/s15324834basp2602&3_3

Meier, P.F., Flepp, R., Meier, P. & Franck, E. (2022) Outcome bias in self-evaluations: Quasi-experimental field evidence from Swiss driving license exams. *Journal of Economic Behavior & Organization*, 201, 293–309. Available from: https://doi.org/10.1016/j.jebo.2022.07.013

Meier, P.F., Flepp, R. & Franck, E. (2021) Are sports betting markets semistrong-form efficient? Evidence from the COVID-19 pandemic. *International Journal of Sport Finance*, 16(3), 111–126.

Merz, O., Flepp, R. & Franck, E. (2021) Sonic Thunder vs. Brian the Snail: are people affected by uninformative racehorse names? *Journal of Behavioral and Experimental Economics*, 93, 1–10. Available from: https://doi.org/10.1016/j.socec.2021.101724

Mowen, J.C. & Stone, T.H. (1992) An empirical analysis of outcome biases in constituent evaluations of public policy decision makers. *Journal of Public Policy and Marketing*, 11(1), 24–32. Available from: https://doi.org/10.1177/074391569201100103

Rabin, M. (1998) Psychology and economics. *Journal of Economic Literature*, 36(1), 11–46.

Ratner, R.K. & Herbst, K.C. (2005) When good decisions have bad outcomes: the impact of affect on switching behavior. *Organizational Behavior and Human Decision Processes*, 96(1), 23–37. Available from: https://doi.org/10.1016/j.obhdp.2004.09.003

Rosenqvist, O. & Skans, O.N. (2015) Confidence enhanced performance?–The causal effects of success on future performance in professional golf tournaments. *Journal of Economic Behavior & Organization*, 117, 281–295. Available from: https://doi.org/10.1016/j.jebo.2015.06.020

Rubin, J. & Sheremeta, R. (2016) Principal–agent settings with random shocks. *Management Science*, 62(4), 985–999. Available from: https://doi.org/10.1287/mnsc.2015.2177

Tinsley, C.H., Dillon, R.L. & Cronin, M.A. (2012) How near-miss events amplify or attenuate risky decision making. *Management Science*, 58(9), 1596–1613. Available from: https://doi.org/10.1287/mnsc.1120.1517

Tversky, A. & Kahneman, D. (1974) Judgment under uncertainty: heuristics and biases. *Science*, 185(4157), 1124–1131. Available from: https://doi.org/10.1126/science.185.4157.1124

Vaughan Williams, L. (1999) Information efficiency in betting markets: a survey. *Bulletin of Economic Research*, 51(1), 1–39. Available from: https://doi.org/10.1111/1467-8586.00069

Vaughan Williams, L. & Reade, J.J. (2016) Forecasting elections. *Journal of Forecasting*, 35(4), 308–328. Available from: https://doi.org/10.1002/for.2377

Wunderlich, F., Seck, A. & Memmert, D. (2021) The influence of randomness on goals in football decreases over time. An empirical analysis of randomness involved in goal scoring in the English Premier League. *Journal of Sports Sciences*, 39(20), 2322–2337. Available from: https://doi.org/10.1080/02640414.2021.1930685

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Flepp, R., Merz, O. & Franck, E. (2023) When the league table lies: does outcome bias lead to informationally inefficient markets? *Economic Inquiry*, 1–16. Available from: https://doi.org/10.1111/ecin.13163

APPENDIX

Expected goals model

Using 214,194 shots, we estimate the following logistic regression model:

$$Ln\left[\frac{P(goal_l = 1)}{P(goal_l = 0)}\right] = \beta_1 distance_l + \beta_2 angle_l + \beta_3 freekick_l + \beta_4 penaltykick_l + \beta_5 header_l + \theta + \tau + e_l \quad \text{(A1)}$$

where l denotes a shot. The dependent variable *goal* denotes whether the shot results in a goal (0/1). Following Brechot and Flepp (2020), *distance* measures the distance between the goal and the shot location, and *angle* is defined as the angle between the two goalposts and the shot location. Furthermore, *freekick*, *penaltykick*, and *header* denote whether the shot was a free kick, a penalty kick, and a header, respectively. Finally, θ and τ refer to team fixed effects and opposing team fixed effects, respectively.

After estimating the parameters in Equation A1, we predict the scoring probability of each individual shot. For example, in the match between Tottenham Hotspur and Chelsea FC on August 20, 2017, the open play shot of Harry Kane (Tottenham Hotspur) in minute 30 taken 17.2 m away from the goal with an angle of 17.6° has an estimated scoring probability of 8%.

TABLE A1 Alternative threshold values for over- and underperformance of xG.

	win (0/1)			
	(1) 2/-2	**(2) 3/-3**	**(3) 5/-5**	**(4) 6/-6**
impliedprob	1.126***	1.122***	1.122***	1.126***
	(0.028)	(0.028)	(0.028)	(0.028)
overperformance of xG	−0.016*	−0.035***	−0.049***	−0.039**
	(0.010)	(0.010)	(0.014)	(0.016)
underperformance of xG	0.017*	0.021**	0.018	0.036**
	(0.010)	(0.010)	(0.013)	(0.015)
home	0.007	0.008	0.008	0.008
	(0.011)	(0.011)	(0.011)	(0.011)
Number of observations	17,484	17,484	17,484	17,484
Number of clusters	8742	8742	8742	8742
Pseudo-R^2	.145	.146	.145	.145

Note: The table reports the marginal effects from a logit regression. The dependent variable *win* denotes the actual outcome of a bet on a team to win the match (1 for a winning bet and 0 for a losing bet). In Column (1), *overperformance of xG* and *underperformance of xG* are equal to one if the absolute rank difference is equal or larger than 2 ranks and zero otherwise. In Column (2), Column (3), and Column (4) the absolute rank difference used is 3, 5, and 6, respectively. *Home* controls for bets on teams that play at home. The heteroscedasticity-robust and clustered standard errors at the match level are reported in parentheses.

***, **, and * denote significance at the 1-, 5-, and 10-percent levels, respectively.