Odds at Play: Testing Efficiency in the Premier League and Serie A

Introduction

The efficiency of football odds has been widely debated over the past years, especially as competition grew in the industry and forced bookmakers to improve their predictive power on the back of lower margins. In this article, we explore some of the biases present in the football betting industry and backtest the past 19 competitive seasons of the Premier League and Serie A to see if they persist through time. After giving an overview of the industry, we implement some basic strategies based on these biases to check for profitability. Lastly, we run Monte Carlo simulations to check if there is also statistical noise in this market aside from biases.

Overview of the Sports Betting Industry and its Biases

Sports betting markets are financial markets, as a bet on a team is very similar to buying a stock in a company. In financial markets, efficiency implies that market participants cannot use strategies to beat the market and profit financially. Transferred to sports betting, market efficiency implies that betting odds (the assets) reflect all available information. Accordingly, there are no systematic strategies that would enable bettors to generate positive returns. In this article we want to investigate whether market inefficiencies are systematic and persist over time, or whether their appearance is of temporary and random nature.

Another motivation for continuing to investigate betting market systems is the development of the market itself. With the spread of online betting platforms, bettors benefit as bookmakers' margins have decreased. It remains to understand whether bookmakers have improved their ability to predict events, compensating for the reduction in margins.

The panel below shows the average margins calculated as $\frac{1}{M}\sum_{m=1}^{M}\left(\sum_{i\in\{h,d,a\}} O_{m,i}^{-1} - 1\right)$ for matches $m = 1, \ldots, M$ from seasons 2005/06 to 2023/24. In the two leagues covered, average margins decreased from more than 10% at the start of our observation period to about 5% in the 2017/18 season. We can then observe how COVID-19 had a positive impact on bookmakers' margins, which persists even today. To remain profitable despite decreasing margin, bookmakers would have to improve their predictive power. We investigate this assumption by considering the Brier score which is given as:
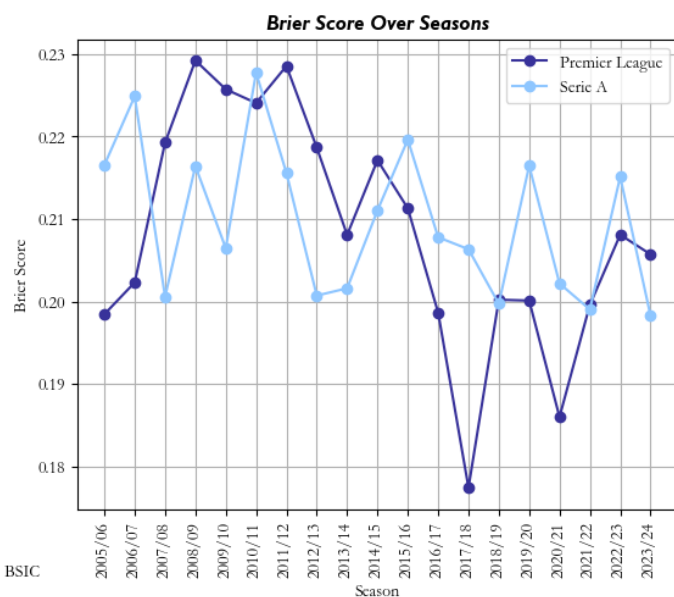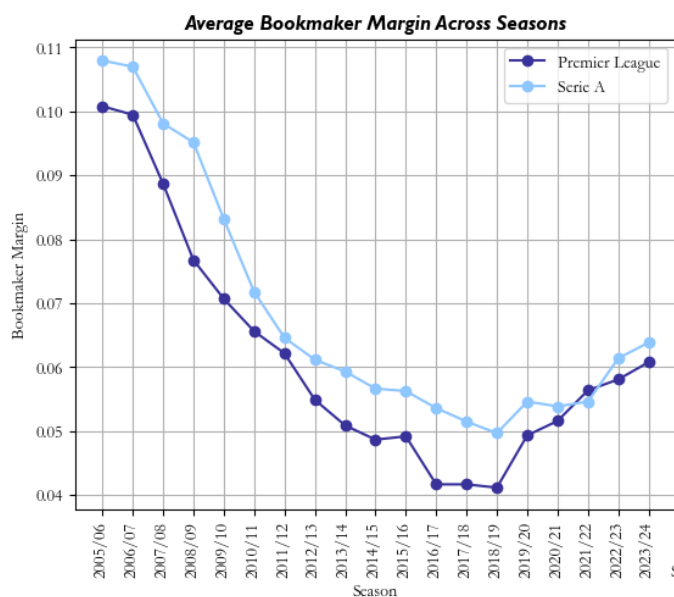
$$\frac{1}{n}\sum_{i=1}^{n}(\pi_i - y_i)^2$$

where π_i denotes the implied probability of bet i according to the bookmakers' odds and y_i indicates whether the bet won ($y_i = 1$) or lost ($y_i = 0$). Perfect predictions would lead to a Brier score of 0, while Brier scores increase in the inaccuracy of predicted game outcomes.

Comparing both panels, we observe that relatively high (low) Brier scores co-occur with high (low) margins. Spikes in the Brier score are observable in the two leagues considered, indicating that the predictive power of bookmakers' odds varies considerably between seasons.

As it stands, we will focus on four main biases that have been mostly researched across academic papers, that will represent the cornerstone for our further analysis and strategies implementation.

The favourite-longshot bias represents the tendency of bettors to overvalue underdogs and undervalue favourites, potentially as a result of risk preference. Bookmakers could in turn deviate the actual betting odds away from the fair odds and offer lower returns on underdogs and higher returns on favourites. If such deviation is large enough, bettors can generate positive returns by simply betting on favourites. Similarly, the reverse favourite-longshot bias inversely suggests undervalued underdogs and positive returns when betting on them. In a later section, we will show that these biases are consistently present in the data.

The home bias refers to increased (lowered) payouts for the home (away) team compared to the fair odds, as bettors very often overestimate the driving force of the spectators. If the bias is large enough, a profitable strategy would suggest to systematically bet on the home team. Biased odds can also result from bettors' sentiment, referred to as sentiment bias in the literature. Here, betting odds are found to be biased towards the more popular teams, quantified based on their average attendance in home games, resulting in positive returns when betting on them. Lastly, we analyse a bias towards newly promoted teams each season. We will later discover that all of these can be found in the data but usually do not persist over consecutive seasons.

Data & Methodology

To start with, we extracted the data for each football match in the Premier League and Serie A starting from 2005/06 to 2023/24, including final scores and odds for each outcome from the 5 most relevant bookmakers, from Football-Data (www.fooxtball-data.co.uk). The attendance data was sourced from https://www.traxnsfermarkt.it/ for the Premier League and from stadiapostcards.com for Serie A. Data was analyzed for each season individually to reflect both specific aspects to that season and potential trends through time.

The first step in our data analysis was to fill the missing odds data with existent odds from other bookmakers and create a column with the average odd of the 5 bookmakers, for each outcome. Subsequently a few binary variables are created – Home, reflecting whether the bet was placed on the home team or not, or Draw, equal to 1 if the

bet was placed on a newly promoted team playing at home; 'OnPromA', equal to 1 if the bet was placed on a newly promoted team playing away; 'VsPromH', representing bets against newly promoted teams playing at home; and 'VsPromA', representing bets against newly promoted teams playing away. A 'DiffAttend' variable was also created for each match, to show the difference in average attendance between the two teams across the respective season. Lastly, a binary variable 'Won' is created to show whether the bet was won or not.

Season 2023-2024 Filtered Premier League Data

	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	B365H	...	Avg Home	Avg Draw	Avg Away	Implied H	Implied A	BetH_H	BetA_H	...	DiffAttend
0	11/08/2023	Burnley	Man City	0	3	A	8	...	8,45	5,3	1,32	0,11116231	0,711607214	1	0	...	-32281
1	12/08/2023	Arsenal	Nott'm	2	1	H	1,18	...	1,162	7,15	14,7	0,80543501	0,06366772	1	0	...	30850
2	12/08/2023	Bournemout	West Ham	1	1	D	2,7	...	2,66	3,35	2,526	0,351237285	0,369869825	1	0	...	-51471
3	12/08/2023	Brighton	Luton	4	1	H	1,33	...	1,308	5,22	9	0,716379566	0,10411383	1	0	...	20299
4	12/08/2023	Everton	Fulham	0	1	A	2,2	...	2,19	3,34	3,21	0,427728508	0,291814776	1	0	...	14757
5	12/08/2023	Sheffield	Crystal	0	1	A	3	...	2,97	3,2	2,372	0,314442377	0,393715792	1	0	...	5157
6	12/08/2023	Newcastle	Aston Villa	5	1	H	1,75	...	1,726	3,74	4,46	0,540981267	0,209357324	1	0	...	10310
7	13/08/2023	Brentford	Tottenham	2	2	D	2,75	...	2,7	3,5	2,4	0,345252774	0,388409371	1	0	...	-44377
8	13/08/2023	Chelsea	Liverpool	1	1	D	2,9	...	2,816	3,48	2,326	0,331141455	0,400900403	1	0	...	-16109
9	14/08/2023	Man United	Wolves	1	0	H	1,33	...	1,294	5,36	9,4	0,725122418	0,099820043	1	0	...	42498

Source: BSIC, Football-Data

The following table represents a sample dataframe for the Premier League 2023-2024 season, showcasing only the first nine matches rather than the complete dataset of 379 matches for that year. Important variables including the home team, away team, the full-time result (FTR) and the full-time home goals (FTHG) are included in this sample. It also includes the betting odds from Bookmaker 365, such as the B365H (home team odds). On the other hand, the B365A (away team odds), as well as the B365D (draw odds), and the other bookmakers' odds are not shown, due to limited space. Furthermore, the sample dataframe does not include the average odds for home, away, and draw across all five bookmakers. However, it does show the binary variables that were created for the BetH_Home (1) and the BetA_Home (0), to differentiate between home and away team bets. Yet, due to space constraints, the sample dataframe does not include several other binary variables: BetH_OnPromH and BetH_OnPromA indicate bets on newly promoted home and away teams, respectively, while BetH_VsPromH and BetH_VsPromA reflect bets against these teams in home and away matches. Similarly, BetA_OnPromH and BetA_OnPromA represents bets on away teams against newly promoted teams, while BetA_VsPromh and BetA_VsPromA track bets against newly promoted away teams, providing insights into market perceptions and biases related to these teams. These variables are later used for the trading strategies involving promoted teams. Lastly, the DiffAttend variable captures the difference in average attendance between the two teams, offering insights into sentiment bias and its impact on betting behaviour.

The Favourite-Longshot Bias

To check whether underdogs and favourites are correctly priced by bookmakers, we will compare the implied probabilities of a home win or away win outcome with the probabilities predicted by a logistic regression model. The model will use the 'Implied Bet Probability', 'Home' and 'DiffAttend' as independent variables and 'Won' as dependent variable, for a simplistic approach where the pricing of odds is analyzed based on the team the bet was placed on (home bias) and on how popular that team is versus its opponent (sentiment bias).

The 'Implied Bet Probability' variable is derived from the bookmakers' odds for each outcome as follows:

$$p_i = \frac{O_i^{-1}}{O_h^{-1} + O_d^{-1} + O_a^{-1}}$$

where p_i is the implied probability of the bet, O_i represents the average bookmakers' odds for outcome i = {home, draw, away} and $x^{-1} = \frac{1}{x}$.

Using the logistic regression, we can predict the implied probability of a specific bet implied by the model and visually compare it to bookmakers' implied probabilities. To do the following, we utilized all available data from the 2005-06 season to the 2023-2024 season, allowing us to assess potential biases over a comprehensive timeframe. The data was divided into two sets: one for home matches and another for away matches. For the away matches, we used the Implied Away Probability and DiffAttend as predictor variables, with the dependent variable indicating whether the away team won. For home matches, we incorporated the Implied Home Probability, a binary indicator for bets placed on the home team, and DiffAttend as a sentiment measure, with the outcome variable reflecting whether the home team secured victory. We then fitted independent logistic regression models to each dataset. This helps identifying any prevailing biases in how bookmakers' price these outcomes over seasons by allowing us to quantify the links between the predictors and the likelihood of winning.

In turn, it also improves our understanding of betting market inefficiencies by gaining important insights into the alignment (or misalignment) of bookmakers' odds with actual winning probability by examining the estimated coefficients and forecasted probabilities.

Source: BSIC
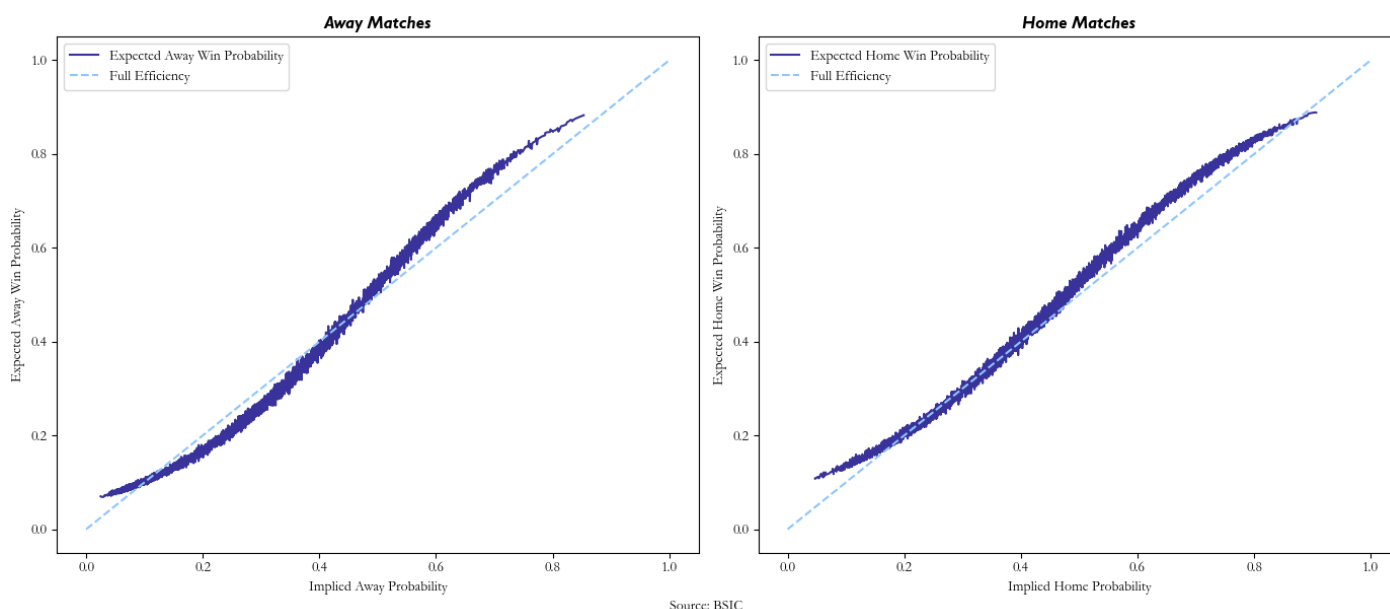
As the charts show, across seasons in the Premier League, bookmakers tend to undervalue favourites with implied probabilities between 0.5 and 0.8 playing at home, as the expected probabilities implied by the model is higher within this interval, whereas they tend to overvalue underdogs with implied probabilities between 0.2 and 0.4 playing away, as the expected probabilities implied by the model are lower for this interval. These findings signal the existence of a persistent favourite-longshot bias, in line with previous academic research done by Direr (2011) and Franke (2020).

Similarly, the outcomes of the logistic regression study for Serie A closely matched those of the Premier League. A consistent favourite-longshot bias was seen, with bookmakers overvaluing underdogs with implied probabilities between 0.1 and 0.4 when they played away and undervaluing favourites with indicated probabilities between 0.5 and 0.8 when they played at home.

Checking for Profitable Strategies

In this section, we are going to check whether the biases presented above persist over time or not and correspondingly see if bettors can exploit them and consistently make a profit through simple strategies. We are going to run three types of strategies, based on home bias, promotion bias and sentiment bias. All of them bet the same amount on each match, as a proportion of the total notional. The first strategy places a bet on the home team for all the matches taking place during a season. The second, third and fourth strategies place money on matches involving only one newly promoted team, in the following sequence: betting on promoted teams playing at home, betting against promoted teams playing at home and betting against promoted teams playing away. We removed the strategy placing bets on promoted teams playing away as we considered that suboptimal in comparison to the others. The last strategy is based on the popularity of teams, and only places bets when the difference in average attendance between the two teams exceeds the 95th quantile of the variable "DiffAttend".

Below are the results of the first four strategies for the Premier League, showing both the returns for each individual season and the weighted average across all the seasons.

Home Betting		**On Promoted Home**		**Vs Promoted Home**		**Vs Promoted Away**	
Season	*Return*	*Season*	*Return*	*Season*	*Return*	*Season*	*Return*
2005-06	**2.82%**	**2005-06**	-29.83%	**2005-06**	**20.36%**	**2005-06**	-9.07%
2006-07	-0.31%	**2006-07**	-5.22%	**2006-07**	-35.93%	**2006-07**	**4.23%**
2007-08	-10.22%	**2007-08**	-29.64%	**2007-08**	-10.46%	**2007-08**	**27.13%**
2008-09	-6.85%	**2008-09**	**9.67%**	**2008-09**	-12.59%	**2008-09**	**1.93%**
2009-10	**7.82%**	**2009-10**	**0.66%**	**2009-10**	-32.76%	**2009-10**	**9.20%**
2010-11	-0.20%	**2010-11**	-18.16%	**2010-11**	-10.48%	**2010-11**	-13.36%
2011-12	-4.49%	**2011-12**	**24.60%**	**2011-12**	-32.25%	**2011-12**	**0.84%**
2012-13	-11.61%	**2012-13**	-11.03%	**2012-13**	-20.22%	**2012-13**	**1.92%**
2013-14	-0.22%	**2013-14**	**16.04%**	**2013-14**	-10.45%	**2013-14**	**6.51%**
2014-15	-2.99%	**2014-15**	-18.63%	**2014-15**	-8.97%	**2014-15**	**12.73%**
2015-16	-10.10%	**2015-16**	-35.98%	**2015-16**	**21.76%**	**2015-16**	-1.59%
2016-17	**5.52%**	**2016-17**	**31.36%**	**2016-17**	-1.89%	**2016-17**	**14.51%**
2017-18	**1.49%**	**2017-18**	**19.81%**	**2017-18**	-13.98%	**2017-18**	-5.43%
2018-19	**3.47%**	**2018-19**	-6.95%	**2018-19**	**20.52%**	**2018-19**	**13.21%**
2019-20	**4.89%**	**2019-20**	**32.04%**	**2019-20**	**4.12%**	**2019-20**	-4.20%
2020-21	-15.86%	**2020-21**	-50.95%	**2020-21**	**17.71%**	**2020-21**	-11.59%
2021-22	-10.49%	**2021-22**	-30.06%	**2021-22**	**20.59%**	**2021-22**	-1.58%
2022-23	**7.60%**	**2022-23**	**47.27%**	**2022-23**	-33.87%	**2022-23**	-4.36%
2023-24	-6.64%	**2023-24**	-45.59%	**2023-24**	**10.37%**	**2023-24**	**5.68%**
All	-2.44%	**All**	-5.29%	**All**	-5.70%	**All**	2.45%

Source: BSIC, Football-Data

As can be seen above, consistently betting on the home team in the Premier League does not yield positive results in the long-term. Although betting on the home team increases the probability of winning a bet according to our

previous model, this strategy delivers positive returns in only 7 out of 19 seasons considered. The only period seeing consecutive seasons associated with positive returns is from 2016/17 to 2019/20; however, using the same strategy going forward in 2020/21 would have erased all the gains. Overall, the negative returns exceed the positive ones in absolute value, making the weighted average over the period considered negative. As for the strategies involving newly promoted teams, betting only on the home games of these teams, as one can expect, would yield worse returns than the previous strategy. Since the number of matches considered per season is only 51 compared to 380 previously and the odds for promoted teams are considerably higher, there are some seasons like 2019/20 or 2022/23 that could deliver outsized returns following favourable outcomes. However, the strategy exhibits a lot of variance and is unattractive in the long run. When it comes to betting on teams playing against the promoted ones, it might be surprising that betting against promoted teams playing at home is less profitable on average than betting on them. Even though the strategy delivered positive results in 5 out of the past 6 seasons, the variance profile is very similar to the previous strategy's one, making it daunting for bettors. Lastly, betting against promoted teams when they play away appears to be profitable, at least for the period considered, delivering about 2.45% per season, on average. The strategy presents positive returns in 11 out of the 19 seasons considered; however, most of these come in the first part of the period considered, potentially showing that bookmakers adjusted their odds through time and that the strategy is not as profitable as before.

Below are the results of the first fours strategies for the Serie A.

Home Betting		*On Promoted Home*		*Vs Promoted Home*		*Vs Promoted Away*	
Season	*Return*	*Season*	*Return*	*Season*	*Return*	*Season*	*Return*
2005-06	-10.95%	2005-06	-14.29%	2005-06	-1.24%	2005-06	-18.33%
2006-07	-12.62%	2006-07	-4.58%	2006-07	-23.03%	2006-07	**1.55%**
2007-08	-6.58%	2007-08	**14.05%**	2007-08	-15.08%	2007-08	-2.95%
2008-09	**1.33%**	2008-09	-54.20%	2008-09	**18.43%**	2008-09	-16.49%
2009-10	-1.88%	2009-10	**9.87%**	2009-10	-10.30%	2009-10	-2.07%
2010-11	-2.17%	2010-11	**3.22%**	2010-11	**5.21%**	2010-11	**0.67%**
2011-12	-4.82%	2011-12	**11.58%**	2011-12	-29.58%	2011-12	-0.44%
2012-13	-5.64%	2012-13	-19.26%	2012-13	**24.97%**	2012-13	-9.63%
2013-14	-4.04%	2013-14	-18.57%	2013-14	**13.70%**	2013-14	-1.09%
2014-15	-13.68%	2014-15	-19.74%	2014-15	-12.40%	2014-15	-15.65%
2015-16	-3.92%	2015-16	-21.41%	2015-16	**14.69%**	2015-16	-12.64%
2016-17	-0.38%	2016-17	-6.69%	2016-17	-7.86%	2016-17	**12.04%**
2017-18	-17.33%	2017-18	-24.22%	2017-18	**5.68%**	2017-18	**7.51%**
2018-19	-8.21%	2018-19	-27.19%	2018-19	-2.62%	2018-19	-5.62%
2019-20	-12.51%	2019-20	-28.50%	2019-20	**0.85%**	2019-20	-17.35%
2020-21	-16.41%	2020-21	-38.43%	2020-21	**1.96%**	2020-21	-8.41%
2021-22	-20.31%	2021-22	-30.60%	2021-22	**6.76%**	2021-22	-17.38%
2022-23	-8.45%	2022-23	-13.03%	2022-23	**3.44%**	2022-23	-9.06%
2023-24	-12.81%	2023-24	-16.54%	2023-24	-23.31%	2023-24	-16.25%
All	-8.49%	**All**	-15.71%	**All**	-1.56%	**All**	-6.92%

Source: BSIC

The conclusion is very similar for Serie A, although the results differ a lot. While 7 seasons of the Premier League generated positive returns in the "home betting" strategy, only 1 of the 19 managed to do so in the Italian league, with a weighted average of -8.49% for the period considered. Although betting on the promoted teams when playing at home yielded positive returns in 4 seasons, the average return Is twice as bad for this strategy. Betting against promoted teams when playing at home appeared to be even better than betting against them away, as the former delivered positive returns in 10 out of 19 seasons (but regardless the strategy's average return is slightly below par), whereas the latter only saw 4 seasons of positive performance, with the past 6 seeing massive drawdowns. Thus, none of the strategies presented delivers positive returns on average over the timeframe considered in Serie A.

Lastly, we have the results of the "sentiment bias" strategy for both the Premier League and Serie A, which excludes the 2019-20 and 2020-21 seasons during which spectators were banned due to COVID-19.

Diff Attend PL		Diff Attend SA	
Season	**Return**	**Season**	**Return**
2005-06	**0.52%**	**2005-06**	**19.62%**
2006-07	**9.82%**	**2006-07**	**18.82%**
2007-08	**7.26%**	**2007-08**	-23.76%
2008-09	**21.68%**	**2008-09**	**16.50%**
2009-10	**16.94%**	**2009-10**	**28.52%**
2010-11	**17.46%**	**2010-11**	**24.08%**
2011-12	**8.50%**	**2011-12**	**36.96%**
2012-13	**26.96%**	**2012-13**	**20.60%**
2013-14	-5.88%	**2013-14**	-19.70%
2014-15	-6.92%	**2014-15**	-23.38%
2015-16	**8.78%**	**2015-16**	-23.62%
2016-17	-29.70%	**2016-17**	-23.16%
2017-18	-1.04%	**2017-18**	-26.70%
2018-19	-4.94%	**2018-19**	-9.66%
2021-22	-9.52%	**2021-22**	-16.82%
2022-23	-1.46%	**2022-23**	-37.14%
2023-24	-10.90%	**2023-24**	**12.26%**
All	**2.79%**	**All**	-1.56%

Source: BSIC

This strategy presents an interesting picture for both leagues. While it delivered astonishing returns until 2012/13, the strategy has only managed to generate positive returns in only one of the past 9 seasons. One of the reasons for this, alongside the decay in returns for other strategies performing well in the first part of the timeframe considered, could be that matches have become more balanced in the recent seasons.

Although betting against promoted teams playing away or on the most popular teams by fanbase in the Premier League brought positive returns on average for the period considered, we cannot consider this an inefficiency of

the market as the return is still a few percentage points below bookmakers' margin. As shown in the tables above, the results of strategies created based on assumed existing biases do not lead to the expected results.

Testing the Efficiency of the Market using Monte Carlo Simulations

To assess whether the biases observed in betting markets are due to chance rather than real inefficiencies, we conducted a simulation-based experiment on 17 seasons of English Premier League data. In a fully functioning market, the probability of an outcome should be entirely determined by the odds published by the bookmakers. To test this hypothesis, we used Monte Carlo simulation to generate 6460 simulated achievements of results (home win, draw, away win) for each of the games between the 2005/06 and 2023/24 seasons, based on the implied probabilities suggested by the odds. We will use $p=0.1$, $p=0.05$ and $p=0.01$ as significance levels and check what percentage of the simulations have a statistically significant bias for at least k seasons, where k=1, 2…,17.

Home Bias Results

#Seasons	p=0.1	p=0.05	p=0.01
0	100	100	99,77
1	100	100	98,17
2	100	99,99	92,93
3	100	99,86	81,24
4	99,99	99,26	62,83
5	99,86	97,52	41,78
6	99,52	92,41	23,58
7	98,31	82,59	10,66
8	94,17	67,59	4,23
9	86,09	49,09	1,20
10	71,81	30,18	0,20
11	51,85	15,64	0,03
12	31,46	6,56	0,01
13	15,14	1,74	0
14	5,35	0,41	0
15	1,33	0,10	0
16	0,11	0,01	0

Source: BSIC

Even when assuming fully efficient markets, in 100% of our simulation runs we find at least one season with significant home bias on the 10% and 5% significance level. This implies that home teams consistently perform better than expected, particularly over short time frames. However, as we extend our analysis to cover more seasons and apply stricter alpha levels, we observe a clear decline in the detection of significant biases. The proportion of runs detecting substantial impacts at the 1% level falls to just 0.01% by the time we get to the 16th season. This trend suggests that although home biases are visible in lone instances, they become much less strong and consistent over time, indicating that those patterns are not as robust as they initially appear.

The outcomes of our Monte Carlo simulation thus underscore an important point: rather than being caused by actual market inefficiencies, statistical noise frequently influences the biases found by the logistic regression. Even

if it is possible to have solid returns in specific season, these outcomes can just be the result of random fluctuations rather than market anomalies. These biases vanish when we examine more seasons or use more stringent significance thresholds, proving that they are not trustworthy markers of long-term inefficiencies. Given this, we draw the conclusion that there are not any long-term viable winning strategies. Although the biases that are frequently mentioned in the literature can be formed randomly and have no consistent statistical relevance, they may occur in some instances. This is particularly crucial for academics and bettors to take into account when analyzing past betting data. The idea that there are enduring market inefficiencies is undermined by the fact that, despite the appearance of exploitable inefficiencies, short-term abnormalities tend to disappear over long periods of time, according to our data. These results also highlight the difference between genuine market biases and random statistical noise.

It is important to distinguish between the potential for short-term financial gain and the presence of a systematic advantage in the market. According to our findings, inefficiencies are frequently the product of transient, random changes that have no lasting relevance. Hence, while biases can appear under specific conditions, the broader betting market remains largely efficient over time, with no reliable long-term strategies for exploiting these temporary deviations. Given the economic significance of these markets and their close ties to the sport itself, research on betting markets will remain interesting.

References

[1] Winkelmann D., Ötting M. and Deutscher C., "Betting Market Inefficiencies in European Football - Bookmakers' Mispricing or Pure Chance?", 2020

[2] Meier P.F., Flepp R. and Franck E., "Are sports betting market semistrong efficient? Evidence from the COVID-19 pandemic", Internation Journal of Sport Finance, 2021

[3] Angelini G. and De Angelis L., "Efficiency of online football betting markets", 2017

[4] Daunhawer I., Schoch D. and Kosub S., "Biases in the football betting market", 2017

[5] Direr A., "Are betting markets efficient? Evidence from European Football Championships", 2011

[6] Franke M., "Do market participants misprice lottery-type assets? Evidence from the European soccer betting market", 2020

TAGS: Sports betting, Serie A, Premier League, Brier score, Logistic regression, Monte Carlo simulations, Efficient markets.