Proceedings of ICFTBA 2026 Symposium: Strategic Management & Business Analytics: Data-Driven Performance Models
DOI: 10.54254/2754-1169/2026.LD38233

How Accurately Do Bookmaker Odds Predict English Premier League Match Outcomes? - Evidence from the 2021/22–2023/24 Seasons

Kening Lai

Shanghai Pudong Experimental School, Shanghai, China
laikkonen421@gmail.com

Abstract. Football results are uncertain, but bookmakers express their expectations through betting odds. These odds can be studied as probability forecasts. This paper examines market-average closing odds for 1,140 Premier League matches from 2021/22 to 2023/24. Home, draw and away odds were converted into implied probabilities and proportionally normalized. The analysis was quantitative and reproducible, using publicly available closing odds and standard forecast evaluation measures. Forecasts were assessed using favourite accuracy, the multiclass Brier score, and five-bin calibration. The highest-probability outcome occurred in 58.25% of matches, and the Brier score was 0.551, compared with 0.667 for equal probabilities. Calibration was close overall, with an expected calibration error of 1.40 percentage points, although higher-probability outcomes occurred slightly more often than predicted. The odds provided useful forecasts and were well calibrated overall. However, assuming the normalized probabilities are correct, betting at these odds would still imply a theoretical loss due to the average overround of 3.91%.

Keywords: Premier League, betting odds, implied probability, Brier score, calibration

1. Introduction

A football match has three standard full-time outcomes: home win, draw and away win. Team strength, injuries and chance make these outcomes difficult to predict. Bookmakers consider such information when setting odds. Previous research found that bookmaker odds were a useful source of forecasts for European football [1]. Research on football tipsters also found limited additional value from expert predictions [2].

An efficient betting market should reflect available information, making its forecasts difficult to improve on. However, studies have reached different conclusions. Angelini and De Angelis found that market-average odds were generally efficient, while results using the best available odds were more mixed [3]. Vlastakis, Dotsis and Markellos found some arbitrage and profitable trading opportunities across bookmakers [4]. A few successful bets alone therefore cannot establish whether a market is efficient.

This paper asks how accurately normalized closing odds predicted Premier League outcomes across three seasons. It compares seasonal performance and the calibration of high- and low-

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.LD38233

probability outcomes. It does not develop a betting strategy. The aim is to evaluate forecasts, which may be statistically useful even when the offered prices are unattractive to bettors.

2. Theoretical background

Decimal odds show the total return, including the stake, from a successful one-unit bet. Their reciprocals resemble probabilities, but the three reciprocals usually sum to more than one. The excess is called the overround, which reflects a built-in margin rather than the bookmaker's actual profit. Probabilities must be adjusted before evaluating them as a complete forecast. Several adjustment methods exist, including Shin's method, and they can give different results [5]. This paper uses proportional normalization because it is simple and reproducible. It assumes that the margin is distributed proportionally across the three outcomes.

Forecast accuracy involves more than counting correct favourites. A 60% prediction is not necessarily poor when the event fails to occur. If similar events occur about 60% of the time, the predictions are well calibrated. The Brier score evaluates probability forecasts using squared differences between predicted probabilities and observed outcomes [6]. It penalizes confident errors more heavily than cautious ones. Calibration compares mean predicted probabilities with observed frequencies. Together, these measures assess both individual forecasts and their agreement with results across groups of matches. Neither measure requires a forecast to identify the winner of every match.

3. Methodology

3.1. Data and sample

The dataset covers all 380 matches in each of the 2021/22, 2022/23 and 2023/24 seasons. Full-time results and odds came from Football-Data [7]. The fields were FTR for the result and AvgCH, AvgCD and AvgCA for market-average closing home, draw and away odds. Football-Data identifies closing odds with "C" and market averages with "Avg" [7].

Data checks identified no duplicate matches, missing-odds entries, or invalid result codes, so all 1,140 matches were retained. Results cover regulation time, including stoppage time. Closing odds were chosen to reflect information available shortly before kick-off. Market averages reduce dependence on one bookmaker's prices.

3.2. Quantitative methods

For match j and outcome i, where i denotes home win (H), draw (D) or away win (A), Equation (1) converts decimal odds into raw implied probabilities.

$$q_{ij} = \frac{1}{o_{ij}} \tag{1}$$

Here, o denotes market-average closing odds and q their reciprocal. Equation (2) calculates the overround for each match.

$$m_j = q_{Hj} + q_{Dj} + q_{Aj} - 1 \tag{2}$$

An overround of 0.04 means the raw probabilities sum to 1.04. Equation (3) normalizes them by dividing each probability by this sum.

DOI: 10.54254/2754-1169/2026.LD38233

$$p_{ij} = \frac{q_{ij}}{q_{Hj} + q_{Dj} + q_{Aj}} \tag{3}$$

Four quantitative measures were used. Descriptive statistics summarized outcome frequencies. Favourite accuracy measured how often the highest-probability outcome occurred. The multiclass Brier score assessed the full forecast using Equation (4).

$$BS = \frac{1}{N} \sum_{j=1}^{N} \sum_{i \in \{H,D,A\}} (p_{ij} - y_{ij})^2 \tag{4}$$

N is the number of matches; y equals one for the observed outcome and zero otherwise. Lower scores are better. Two benchmarks were used: equal probabilities of one-third and constant probabilities based on overall outcome frequencies. The latter is an in-sample benchmark, not a forecast based only on pre-match information.

For calibration assessment, all 3,420 outcome probabilities were pooled into five bins: 0-20%, 20-40%, 40-60%, 60-80% and 80-100%. Each bin includes its lower boundary but excludes its upper boundary, except the last, which includes 100%. Equation (5) calculates expected calibration error (ECE).

$$ECE = \sum_{b=1}^{B} \frac{n_b}{3N} |\bar{p}_b - f_b| \tag{5}$$

Here, B is five, n is the number of predictions in bin b, and the barred p and f denote its mean predicted probability and observed frequency. ECE is reported in percentage points. The calibration plot shows the direction of the differences.

4. Results and analysis

4.1. Match outcomes and bookmaker margin

Table 1 summarizes outcomes and overround by season. Home wins were most common, rising from 42.89% to 48.42% before falling to 46.05%. Draws declined from 23.16% to 21.58%. Overall, home wins accounted for 45.79% of matches, draws for 22.54% and away wins for 31.67%.

Table 1. Match outcomes and mean overround by season

Season	Matches	Home wins (%)	Draws (%)	Away wins (%)	Mean overround (%)
2021/22	380	42.89	23.16	33.95	3.87
2022/23	380	48.42	22.89	28.68	3.90
2023/24	380	46.05	21.58	32.37	3.97
Overall	1,140	45.79	22.54	31.67	3.91

Source: Author's calculations using Football-Data closing odds.

The mean overround rose slightly from 3.87% to 3.97%, averaging 3.91%. This is not the bettor's expected loss. Assuming the normalized probabilities are correct, Equation (6) gives the expected net return per unit staked on an outcome.

$$E(R_{ij}) = p_{ij} o_{ij} - 1 = \frac{1}{1 + m_j} - 1 \tag{6}$$

DOI: 10.54254/2754-1169/2026.LD38233

Averaging this calculation across matches gives a theoretical return of −3.77%. Under the stated assumption, a bettor accepting average market odds would expect to lose about 3.77 units per 100 units staked. This is not a measured betting return.

4.2. Forecast accuracy and Brier scores

Table 2 reports forecast performance. Favourite accuracy was 58.25% overall, ranging from 55.53% in 2022/23 to 60.00% in 2023/24. No draw was the highest-probability outcome: the forecasts selected 710 home wins and 430 away wins. Draw probabilities were still included in the Brier score.

Table 2. Forecast performance by season

Season	Favourite accuracy (%)	Brier score	Five-bin ECE (p.p.)
2021/22	59.21	0.554	1.94
2022/23	55.53	0.571	0.10
2023/24	60.00	0.527	2.32
Overall	58.25	0.551	1.40

Source: Author's calculations using Football-Data closing odds.

The overall Brier score was 0.551, below the equal-probability benchmark of 0.667 and the frequency benchmark of 0.639. The mean squared forecast error was about 17.4% lower than that under equal probabilities. Scores ranged from 0.527 in 2023/24 to 0.571 in 2022/23. The latter season had the lowest ECE but the highest Brier score. A low grouped calibration error does not necessarily mean better individual forecasts.

Mean predicted home-win probability was 43.95%, against 45.79% observed. Draws were predicted at 23.75% and occurred 22.54% of the time. Away wins were predicted at 32.30% and occurred 31.67% of the time. The market slightly underestimated home wins and overestimated draws.

4.3. Calibration

Figure 1 compares mean predictions with observed frequencies in the five bins. Perfect calibration lies on the 45-degree dashed line. The points were close to this line, with an overall ECE of 1.40 percentage points.

In the 0–20% bin, mean predicted probability was 13.95% and observed frequency was 13.35%. The corresponding figures were 28.03% and 26.89% in the 20–40% bin. In the three higher bins, observed frequencies exceeded the corresponding predicted values: 52.30%, 70.67% and 86.67%, against 49.50%, 68.99% and 83.69%. Low-probability outcomes occurred slightly less often than predicted; higher-probability outcomes occurred slightly more often.

This resembles the favourite–longshot bias discussed in British football betting research [8]. However, the highest bin contained only 60 predictions. Normalization may also affect the pattern. These descriptive results do not establish the existence of a stable bias or identify a profitable betting opportunity.

DOI: 10.54254/2754-1169/2026.LD38233

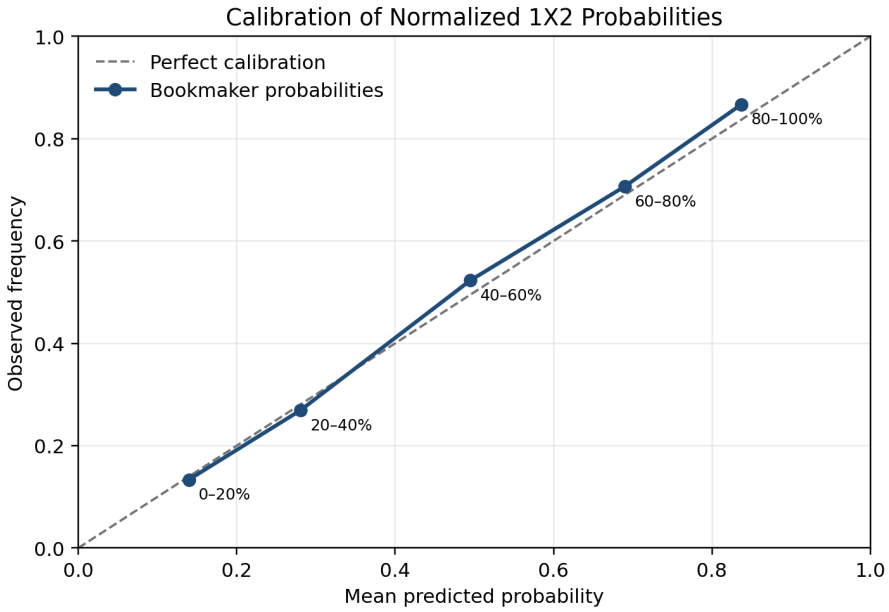

Figure 1. Calibration of normalized 1X2 probabilities

Source: Author's calculations using Football-Data closing odds.

5. Discussion

Market-average closing odds provided useful probability forecasts. Their Brier score was lower than both benchmarks, and their calibration was close overall. This finding agrees with earlier research on bookmaker forecasts [1]. Closing odds can reflect team news and betting activity, although this study cannot separate the effects of these inputs.

Performance varied by season. In 2022/23, home wins occurred about 4.85 percentage points more often than the model predicted, while away wins occurred about 3.42 percentage points less often. Club strength, newly promoted teams, injuries and random variation may contribute to seasonal differences. The present analysis does not test these explanations.

The calibration pattern is consistent with a possible favourite–longshot bias [8], but the largest bin-level gap was only 2.98 percentage points. Without uncertainty estimates, these gaps cannot establish market inefficiency. Conclusions may also depend on whether average or best available odds are used [3]. This paper assesses the average forecast, not the best price a bettor could find.

The margin remains important even when forecasts are useful. On average, raw implied probabilities summed to 103.91%, giving a theoretical expected loss of 3.77% under proportional probability normalization. Actual betting returns could differ substantially. Wunderlich and Memmert distinguish betting profitability from forecast accuracy [9]. This supports evaluating the probabilities directly rather than judging them by short-term betting profits.

Using complete seasons avoids selecting only memorable matches. The listed fields and formulas also make the calculations reproducible. However, proportional normalization is only one adjustment method. Štrumbelj found that Shin's method can produce more accurate probabilities in some settings [5]. Comparing methods would help show whether the small calibration pattern depends on the chosen adjustment.

DOI: 10.54254/2754-1169/2026.LD38233

6. Conclusion

This study examined market-average closing odds for 1,140 Premier League matches across the 2021/22, 2022/23 and 2023/24 seasons. After proportional normalization, the highest-probability outcome occurred in 58.25% of matches. The overall multiclass Brier score was 0.551, below 0.667 for equal probabilities and 0.639 for the constant frequency benchmark. The odds therefore contained useful information about match outcomes, although many individual results remained uncertain.

Overall calibration was close, with a five-bin ECE of 1.40 percentage points. Outcomes in the higher-probability bins occurred slightly more often than predicted, while those in the lower bins occurred slightly less often. This suggests a possible favourite–longshot pattern, but the descriptive analysis does not show that it is stable. Seasonal results also differed. The season with the lowest ECE did not have the lowest Brier score, so one measure alone would give an incomplete assessment.

The average overround was 3.91%. If the normalized market probabilities were correct, this would imply a mean theoretical return of about −3.77% per unit staked at average odds. This value does not measure the return from any real-world betting strategy. The clearest conclusion is that Premier League closing odds were useful but imperfect probability forecasts in this sample. Their predictive value should be assessed separately from whether the offered prices could make betting profitable.

This study covers three seasons and one data provider. Results may differ with other periods, bookmakers or opening odds. Proportional normalization may miss unequal margin allocation. Five broad bins can hide smaller calibration differences, and no confidence intervals or significance tests were calculated. The pooled predictions also share matches, so they are not independent observations. The analysis cannot explain the causes of forecast errors.

References

[1] Štrumbelj, E., & Robnik-Šikonja, M. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting, 26*(3), 482–488. https://doi.org/10.1016/j.ijforecast.2009.10.005

[2] Forrest, D., & Simmons, R. (2000). Forecasting sport: The behaviour and performance of football tipsters. *International Journal of Forecasting, 16*(3), 317–331. https://doi.org/10.1016/S0169-2070(00)00050-9

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

[4] Vlastakis, N., Dotsis, G., & Markellos, R. N. (2009). How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting, 28*(5), 426–444. https://doi.org/10.1002/for.1085

[5] Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting, 30*(4), 934–943. https://doi.org/10.1016/j.ijforecast.2014.02.008

[6] Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

[7] Football-Data. (n.d.). Historical football results and betting odds data. Retrieved September 7, 2026, from https://www.football-data.co.uk/data.php

[8] Cain, M., Law, D., & Peel, D. (2000). The favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy, 47*(1), 25–36. https://doi.org/10.1111/1467-9485.00151

[9] Wunderlich, F., & Memmert, D. (2020). Are betting returns a useful measure of accuracy in (sports) forecasting? *International Journal of Forecasting, 36*(2), 713–722. https://doi.org/10.1016/j.ijforecast.2019.08.009