# Evaluating the efficiency of bookmaker odds in the English Premier League (EPL) between the 2000/01 to 2021/22 seasons so far

Authors: Uyi Erhabor


## Abstract

This research paper evaluates the forecast efficiency of 13 bookmakers setting odds for EPL matches between the 2000/01 to the 2021/22 seasons so far[1]. I used the Mincer Zarnowitz regression test (Mincer and Zarnowitz, 1969) to test weak forecast efficiency and forecast encompassing (strong forecast efficiency). My results revealed that there has been a degree of inefficiency demonstrated by bookmakers when setting odds for all match result outcomes in the EPL between the 2000/01 to 2021/22 seasons so far.


**Key Terms:** Forecasting Efficiency, Weak Forecast Efficiency; Forecast Encompassing; Implied Probability Forecasts; Forecast Errors

[1] This study was made possible due to the availability of betting odds data for the EPL from the 2000/01 to the 2021/22 seasons by https://www.football-data.co.uk/

# Introduction

Informational efficiency in betting markets is implied by the plethora of reliable and widely available information that is accessible to participants and therefore betting odds should accurately reflect the probabilities of future outcomes occurring (Elaad, G, Reade, J. J and Singleton, C, 2020). Furthermore, forecast errors should not exist in theory and my research aims to investigate whether this is the case in reality.

My research question is: Can the forecast errors of bookmaker odds for match result outcomes be explained by implied probability forecasts? I have divided this research question into two sub research questions:

- Can the forecast errors of bookmaker odds for match result outcomes be explained by their implied probability forecasts?
- Can the forecast errors of bookmaker odds for match result outcomes be explained by the implied probability forecasts of other bookmakers?

To answer my research questions, I will be using the Mincer-Zarnowitz regression test (Mincer and Zarnowitz, 1969) to test weak forecast efficiency as well as forecast encompassing. I will be evaluating the forecast efficiency of each bookmakers' by assessing the accuracy of their predictions and the accuracy of their predictions considering the corresponding predictions of the other bookmakers.

My research contributes to the existing research that investigates the efficiency of betting markets by focusing on determining the efficiency of the odds provided by 13 bookmakers for more than 8,000 EPL matches since the 2000/01 season. Additionally, there is a short supply of research that investigates the efficiency of prices for home-win and away-win result outcomes (Elaad, G, Reade, J. J and Singleton, C, 2020). My research complements the research investigating the efficiency of odds for match result outcomes as I am evaluating the efficiency of bookmaker odds for all three match result outcomes (home-win, away-win and draws).

The main findings of my research are there has been some inefficiency in the odds for home wins, away wins and draws for the EPL between the 2000/01 to the 2021/22 seasons so far according to forecast encompassing tests I have carried out. There is evidence that markets of individual bookmakers are not efficient as they do not consider the information of their competitors (Elaad, G, Reade, J. J and Singleton, C, 2020). This is a finding that the forecast encompassing tests I have carried out support.

However, only the draw and away win odds have been found to have displayed some inefficiency by the weak forecasting efficiency tests I have carried out. Additionally, the forecast encompassing tests reveals that there has been inefficiency in the draw odds at the 1% level of significance, making this finding most notable. There is evidence that posted draw odds do not fully consider available information (Pope and Peel, 1989). This evidence for the notable inefficiency in the draw outcome odds for bookmakers in England is supported by the weak forecasting efficiency tests I have carried out.

## Data

The data set I used is a panel data set from https://www.football-data.co.uk/. The data set comprises match betting odds, total goals betting odds, Asian handicap betting odds and closing odds as well as match statistics for 8268 matches from the 2000/01 to the 2021/22 EPL seasons so far. There is a mixture of online and high street bookmakers (13 bookmakers in total) that are included in the data set. These bookmakers have betting odds for match result outcomes for matches across the 2000/01 to the 2021/22 EPL seasons so far. These bookmakers are bet365, Blue Square, Bet&Win, Gamebookers Ltd, Interwetten, Ladbrokes, Pinnacle, Sporting, Sportingbet, StanJames, Stanleybet, VC Bet and William Hill.

The EPL 2021/22 season is still currently running and therefore the betting odds and match statistics for this season are not complete. Additionally, the number of betting odds for match result outcomes across the 2000/01 to the 2021/22 EPL seasons so far varies by bookmaker. Therefore, my conclusions regarding the efficiency of bookmaker odds across the three match result outcomes will be more reliable for bookmakers who have provided a greater number of betting odds for match result outcomes.

## Methodology

Firstly, I combined the betting odds and match statistics data for the EPL from the 2000/01 season to the 2021/22 season so far. Next I converted the bookmaker decimal odds for home wins, away wins and draws into probability forecasts. Then I computed the forecast errors for the three match result outcomes for the 13 bookmakers in the dataset. Next, I ran the following Mincer Zarnowitz regression test (Mincer and Zarnowitz, 1969) for weak forecast efficiency and forecast encompassing for the three match result outcomes:

- $\hat{e}_i = \alpha + \beta\hat{y}_i + u_i \qquad u_{ij} \sim (0, \sigma^2)$.

Where $\hat{e}_i$ represents the forecast error of a given bookmaker for one a home win, an away win or a draw, $\hat{y}_i$ represents the implied probability forecast of a given bookmaker for a home win, an away win or a draw. Lastly, $u_i$ represents the error of the forecast error of a given bookmaker for a home win, an away win or a draw.

- $\hat{e}_i = \alpha + \beta\tilde{y}_i + \eta_i$

Where $\hat{e}_i$ represents the forecast error of one bookmaker for a home win, an away win or a draw, $\tilde{y}_i$ represents the implied probability forecast of another bookmaker for a home win, an away win or a draw. Lastly, $\eta_i$ represents the error of the regression.

Next, I tabulated the most significant results from the weak forecast efficiency and the forecast encompassing regression tests I carried out and used them to answer my research questions.

# Results

## Weak Forecasting Efficiency

The results of the weak forecasting efficiency tests that I carried out revealed that only 2/13 bookmakers had β coefficients that were statistically significant at the 5% level of significance. This was Interwetten's β coefficient from the regression of their away win forecast errors on their away win implied probability forecasts and Blue Square's β coefficient from the regression of their draw outcome forecast errors on their draw outcome implied probability forecasts. This implies that out of the 39 β coefficients that the weak forecasting efficiency tests I ran produced, Interwetten's away win odds and Blue Square's draw odds are the only odds that on average have been found to have some information that explains their corresponding forecast errors between the 2000/01 to the 2021/22 EPL seasons so far.

My results suggest that overall, the forecast errors of bookmaker odds for match result outcomes can be explained by their implied probability forecasts, but the extent to which they can be explained varies by match result outcome. Most notably the draw outcome forecast errors of bookmakers can be explained by their implied probability forecasts to the greatest extent as 5/13 bookmakers from the data set have coefficients that do not meet the following criterion $H_0$: $\alpha = \beta = 0$ at the 5% level of significance. On the other hand, there are only 1/13 and 3/13 bookmakers that have coefficients that do not meet the following criterion $H_0$: $\alpha = \beta = 0$ at the 5% level of significance for the home win and away win outcomes respectively.

## Forecast Encompassing

The forecast encompassing regression tests that I carried out revealed that between the 2000/01 to the 2021/22 EPL seasons so far, there has been notable instances of forecast encompassing. Blue Square and Pinnacle's draw odds and Pinnacle and Gamebooker's draw odds encompass each other, and the β coefficients from these forecast encompassing regressions are statistically significant at the 1% level of significance as shown in Tables 3, 4 and 5. This implies that Blue Square's draw odds on average between the 2000/01 to the 2021/22 EPL seasons so far, have not considered some information that Pinnacle's draw odds have and vice versa. Additionally, this implies that Pinnacle's draw odds on average have not considered some information that Gamebooker's draw odds have and vice versa. Furthermore, Pinnacle's draw odds have on average been the most inefficient draw odds provided for the EPL between the 2000/01 to the 2021/22 seasons so far with the draw odds of two other bookmaker's able to explain their draw forecast errors.

### The significance of the results

The weak forecasting efficiency tests I have carried out have revealed that there is evidence that the implied probability forecasts of bookmakers for away wins and draws have statistically significant effects on their forecast errors for away wins and draws at the 5% and 10% levels of significance.

Additionally, the forecast encompassing tests I have carried out have revealed that between the 2000/01 to the 2021/22 EPL seasons so far:

- There is evidence that the implied probability forecasts of one bookmaker for home wins has statistically significant effects on the forecast's errors of other bookmakers for home wins at the 5% and 10% levels of significance
- There is evidence that implied probability forecasts of one bookmaker for away wins has statistically significant effects on the forecast's errors of other bookmakers for away wins at the 5% and 10% levels of significance
- There is evidence that the implied probability forecasts of one bookmaker for draws has statistically significant effects on the forecast's errors of other bookmakers for draws at the 1%, 5% and 10% levels of significance

Therefore, the forecasts errors of bookmaker odds for match result outcomes can indeed be explained by their implied probability forecasts and the implied probability forecasts of other bookmakers.

# Conclusion

The main findings of my research are that there has been a degree of inefficiency in the odds set by the bookmakers analysed for home wins, away wins and draws for the EPL between the 2000/01 to the 2021/22 seasons so far according to forecast encompassing tests I have carried out. However, only the draw and away win odds have been found to have displayed some inefficiency by the weak forecasting efficiency tests I have carried out.

My analysis focused on evaluating the efficiency of bookmakers' odds for the EPL. A broader evaluation of the efficiency of betting odds across the top four football divisions in England would enable me to see whether there are notable differences in the efficiency of bookmaker odds across football leagues in England.

In future research I would like to further my investigation of the efficiency of bookmaker odds by evaluating the odds of a larger number of bookmakers to see whether there are notable differences in the efficiency of odds across the broader market for EPL bookmaker odds.

# References


1. Mincer, J, A. and Zarnowitz, V. (1969). 'The evaluation of economic forecasts.' In Economic Forecasts and Expectations*: Analysis of Forecasting Behavior and Performance.* NBER, pp. 1–46.
2. Elaad, G, Reade, J. J and Singleton, C (2020) *Information, prices, and efficiency in an online betting market.* Online at: https://www.sciencedirect.com/science/article/pii/S1544612319306440?via%3Dihub (accessed 30/03/2022).
3. Pope, P. F and Peel, D. A (1989) *Information, Prices, and Efficiency in a Fixed-Odds Betting Market.* Online at*:* https://www.jstor.org/stable/2554281?origin=crossref&seq=1 (accessed 31/03/2022)

**TABLE 1:** Weak forecast efficiency for the away win outcome (first seven regressions only)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | e_away_b365 | e_away_bs | e_away_bw | e_away_gb | e_away_iw | e_away_lb | e_away_ps |
| paway_b365 | 0.006 | | | | | | |
| | (0.027) | | | | | | |
| paway_bs | | -0.02 | | | | | |
| | | (0.05) | | | | | |
| paway_bw | | | 0.015 | | | | |
| | | | (0.028) | | | | |
| paway_gb | | | | 0.006 | | | |
| | | | | (0.037) | | | |
| paway_iw | | | | | 0.059** | | |
| | | | | | (0.027) | | |
| paway_lb | | | | | | 0.007 | |
| | | | | | | (0.03) | |
| paway_ps | | | | | | | 0.017 |
| | | | | | | | (0.037) |
| _cons | -0.025*** | -0.037** | -0.034*** | -0.045*** | -0.057*** | -0.044*** | -0.005 |
| | (0.008) | (0.015) | (0.009) | (0.011) | (0.009) | (0.009) | (0.012) |
| Observations | 7508 | 2280 | 6747 | 4856 | 8250 | 6778 | 3708 |
| R-squared | 0 | 0 | 0 | 0 | .001 | 0 | 0 |
| *Robust standard errors are in parentheses* | | | | | | | |
| **** p<.01, ** p<.05, * p<.1* | | | | | | | |

**TABLE 2:** Weak forecast efficiency for the draw outcome (first seven regressions only)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | e_draw_b365 | e_draw_bs | e_draw_bw | e_draw_gb | e_draw_iw | e_draw_lb | e_draw_ps |
| pdraw_b365 | 0.07 | | | | | | |
| | (0.088) | | | | | | |
| pdraw_bs | | 0.368** | | | | | |
| | | (0.177) | | | | | |
| pdraw_bw | | | 0.069 | | | | |
| | | | (0.092) | | | | |
| pdraw_gb | | | | 0.135 | | | |
| | | | | (0.134) | | | |
| pdraw_iw | | | | | 0.088 | | |
| | | | | | (0.088) | | |
| pdraw_lb | | | | | | 0.165 | |
| | | | | | | (0.107) | |
| pdraw_ps | | | | | | | 0.215* |
| | | | | | | | (0.118) |
| _cons | -0.038 | -0.113** | -0.043* | -0.062* | -0.059** | -0.071** | -0.067** |
| | (0.023) | (0.048) | (0.024) | (0.038) | (0.025) | (0.03) | (0.029) |
| Observations | 7508 | 2280 | 6747 | 4856 | 8250 | 6778 | 3708 |
| R-squared | 0 | 0.001 | 0 | 0 | 0 | 0 | 0.001 |
| *Robust standard errors are in parentheses* | | | | | | | |
| **** p<.01, ** p<.05, * p<.1* | | | | | | | |

**TABLE 3**: Blue Square's draw forecasts errors on probability forecasts of 8 other bookmakers

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| VARIABLES | e_draw_bs | e_draw_bs | e_draw_bs | e_draw_bs | e_draw_bs | e_draw_bs | e_draw_bs | e_draw_bs |
| | | | | | | | | |
| pdraw_b365 | 0.356** | | | | | | | |
| | (0.168) | | | | | | | |
| pdraw_bw | | 0.306* | | | | | | |
| | | (0.171) | | | | | | |
| pdraw_gb | | | 0.349** | | | | | |
| | | | (0.174) | | | | | |
| pdraw_iw | | | | 0.383** | | | | |
| | | | | (0.184) | | | | |
| pdraw_lb | | | | | 0.315* | | | |
| | | | | | (0.173) | | | |
| pdraw_ps | | | | | | 1.147*** | | |
| | | | | | | (0.398) | | |
| pdraw_sb | | | | | | | 0.274 | |
| | | | | | | | (0.196) | |
| pdraw_sj | | | | | | | | 0.292* |
| | | | | | | | | (0.172) |
| Constant | -0.107** | -0.096** | -0.108** | -0.119** | -0.098** | -0.276*** | -0.093* | -0.091** |
| | (0.044) | (0.046) | (0.047) | (0.051) | (0.047) | (0.095) | (0.054) | (0.046) |
| | | | | | | | | |
| Observations | 2,280 | 2,280 | 2,279 | 2,279 | 2,279 | 380 | 1,900 | 2,280 |
| R-squared | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.013 | 0.001 | 0.001 |
| Robust standard errors in parentheses | | | | | | | | |
| *** p<0.01, ** p<0.05, * p<0.10 | | | | | | | | |

**TABLE 4:** Pinnacle's draw forecasts errors on probability forecasts of 6 other bookmakers

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| VARIABLES | e_draw_ps | e_draw_ps | e_draw_ps | e_draw_ps | e_draw_ps | e_draw_ps |
| | | | | | | |
| pdraw_b365 | 0.212* | | | | | |
| | (0.118) | | | | | |
| pdraw_bs | | 1.261*** | | | | |
| | | (0.401) | | | | |
| pdraw_bw | | | 0.216* | | | |
| | | | (0.118) | | | |
| pdraw_gb | | | | 1.099*** | | |
| | | | | (0.406) | | |
| pdraw_iw | | | | | 0.237* | |
| | | | | | (0.123) | |
| pdraw_lb | | | | | | 0.349** |
| | | | | | | (0.168) |
| Constant | -0.067** | -0.310*** | -0.070** | -0.265** | -0.076** | -0.097** |
| | (0.029) | (0.102) | (0.030) | (0.104) | (0.032) | (0.044) |
| | | | | | | |
| Observations | 3,708 | 380 | 3,707 | 379 | 3,707 | 2,279 |
| R-squared | 0.001 | 0.014 | 0.001 | 0.011 | 0.001 | 0.002 |
| Robust standard errors in parentheses | | | | | | |
| *** p<0.01, ** p<0.05, * p<0.10 | | | | | | |

**TABLE 5:** Gamebooker's draw forecasts errors on probability forecasts of 6 other bookmakers

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| VARIABLES | e_draw_gb | e_draw_gb | e_draw_gb | e_draw_gb | e_draw_gb | e_draw_gb |
| | | | | | | |
| pdraw_b365 | 0.094 | | | | | |
| | (0.135) | | | | | |
| pdraw_bs | | 0.378** | | | | |
| | | (0.178) | | | | |
| pdraw_bw | | | 0.032 | | | |
| | | | (0.145) | | | |
| pdraw_iw | | | | 0.127 | | |
| | | | | (0.137) | | |
| pdraw_lb | | | | | 0.165 | |
| | | | | | (0.139) | |
| pdraw_ps | | | | | | 1.110*** |
| | | | | | | (0.404) |
| Constant | -0.048 | -0.116** | -0.029 | -0.062 | -0.070* | -0.266*** |
| | (0.037) | (0.048) | (0.041) | (0.040) | (0.039) | (0.097) |
| | | | | | | |
| Observations | 4,179 | 2,279 | 3,419 | 4,839 | 4,795 | 379 |
| R-squared | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.012 |
| Robust standard errors in parentheses | | | | | | |
| *** p<0.01, ** p<0.05, * p<0.10 | | | | | | |

## Appendix A. Additional tables

TABLE A1: Abbreviations of bookmakers explained

| **Abbreviation** | **Bookmaker** |
|---|---|
| b365 | Bet365 |
| bs | BlueSquare |
| bw | Bet&Win |
| gb | Gamebookers |
| iw | Interwetten |
| lb | Ladbrokes |
| ps | Pinnacle |
| so | Sporting |
| sb | Sportingbet |
| sj | StanJames |
| sy | Stanleybet |
| vc | VC Bet |
| wh | William Hill |