Information, prices and efficiency in an online betting market

Guy Elaad*
Ariel University

J. James Reade
University of Reading

Carl Singleton
University of Reading

April 2019
Link to latest version

Abstract

We study the odds (or prices) set by fifty-one online bookmakers for the result outcomes in over 16,000 association football matches in England since 2010. Adapting a methodology typically used to evaluate forecast efficiency, we test the Efficient Market Hypothesis in this context. We find odds are generally not biased when compared against actual match outcomes, both in terms of favourite-longshot or outcome types. But individual bookmakers are not efficient. Their own odds do not appear to use fully the information contained in their competitors' odds.

Keywords: prediction markets, Efficient Market Hypothesis, favourite-longshot bias
JEL codes: C53, G14, Z29

*guyel@ariel.ac.il; Department of Economics and Business Management, Ariel University, P.O. Box 3, 40700, Ariel, Israel

1 Introduction

Betting markets have long been viewed as a testing ground for Fama's 1970 Efficient Market Hypothesis (henceforth EMH; e.g. Thaler and Ziemba, 1988). Participants in these markets are knowledgeable and motivated, and information is timely, reliable, widely available at next to no cost, and straightforward to interpret. In theory, this informational efficiency should imply betting odds, or prices, supply accurate forecasts of future outcomes. Testing the EMH should be more straightforward in betting markets than in typical financial markets because there is a fixed time when a bet's value is revealed, i.e. when an event ends. There are other particular features of betting markets, most relevant being the nature of bookmakers to exploit systematically biases among their customers (Kuypers, 2000; Levitt, 2004). If this is the case, then odds could in fact imply inefficient forecasts.

We study the online betting market for 16,000 English professional football match results between 2010 and 2018. We chose this market since there are many bookmakers providing odds over several years, and this data is readily available, meaning any findings should not be driven by some narrow period or specific bookmaker. It is also a high volume betting market, with £millions traded on individual match results at the most prominent exchanges.

Despite a literature looking at the topic of efficiency in these betting markets (e.g. Goddard and Asimakopoulos, 2004; Vlastakis et al., 2009), there are few studies formally testing whether prices for home-draw-away result outcomes are efficient. Pope and Peel (1989) studied the odds for English matches in 1981/82 from four high-street bookmakers, finding some evidence of inefficiency, in particular for the draw outcome. Kuypers (2000) similarly found evidence of inefficiency by studying the odds from one bookmaker on English matches in 1993/95, and Reade (2014) did for the top five divisions in English football between 2011 and 2014. Angelini and De Angelis (2019) studied a larger sample of matches and sources of online odds for results in the top European professional leagues. They tested market efficiency by adapting the Mincer and Zarnowitz (1969) forecast evaluation framework, asking whether average market prices for home and away wins deviate from efficiency due to a favourite-longshot bias. They found different degrees of efficiency across markets. We extend this approach by modelling the heterogeneity between bookmakers in their profit margins and considering whether odds also imply too few drawn (tied) matches. We test for a version of semi-strong efficiency, i.e. whether there is any significant opportunity for bettors to do better on average than just losing the bookmaker's profit margin because of some predictable bias in the odds. We also check for differences in efficiency between the tiers of English football, since in the lower leagues competition in the online market is less, the volume of money staked is lower and the bettors are likely to be more informed.

2 Method

Let y_i equal one if some combination of match and outcome result (home win, draw or away win) happened and zero otherwise, where $i = 1, \ldots, I$. Let p_{ij} be the unobserved beliefs of bookmaker $j = 1, \ldots, J$ about the probability of y_i happening beforehand. The bookmaker gives decimal odds o_{ij} on the outcome, meaning that on taking a £1 bet they return o_{ij} to the bettor if the outcome happens and gain £1 if it doesn't. Let $z_{ij} = 1/o_{ij}$ be the inverse odds or implied odds-based probability forecast of the bookmaker. For any match, summing z_{ij} over the three possible outcomes will give a value greater than one, which reflects the bookmaker's expected rate of commission or profit margin κ_j, also know as the 'overround' or 'vig'. This implies $z_{ij} = p_{ij} + \kappa_j$. If we denote $\varepsilon_{ij} = y_{ij} - z_{ij}$, then an efficient bookmaker-specific market requires $E_i\left[\varepsilon_{ij}\right] = -\kappa_j$. In other words, the market is efficient if the bookmaker makes some average level of commission across matches and outcomes, and no other information can predict ε_{ij}, since it will already be priced into the odds.

We consider two general departures from the EMH in a betting market. First, we ask whether the odds systematically over- or under-predict any particular result. There is evidence that individuals tend to under-predict significantly the draw outcome of football matches, which is consistent with the psychological concept of splitting, or 'black and white' thinking (e.g. Na et al., 2018). Second, there is an empirical irregularity in some prediction markets known as the favourite-longshot bias, whereby odds appear to underestimate the chances of the most (least) expected outcomes over the least (most), making bets on favourites more profitable than on longshots or vice versa (see summary by Ottaviani and Sørensen, 2008).

We test whether betting markets are efficient by adapting the standard Mincer and Zarnowitz (1969) forecast efficiency evaluation framework further than Angelini and De Angelis (2019), and estimating the following model using least squares:

$$\varepsilon_{ij} = \beta_h h_{ij} + \beta_a a_{ij} + \beta_z z_{ij} + \phi_{t(ij)} + \alpha_j + \upsilon_{ij}, \quad E\left[\upsilon_{ij} \mid h_{ij}, a_{ij}, z_{ij}, \phi_t, \alpha_j\right] = 0\,, \tag{1}$$

where h_{ij} and a_{ij} are dummy variables indicating whether the odds are for a home or away win. To address heterogeneity in the bookmaker margins and how this might be correlated with their tendency to reflect a favourite-longshot bias in their odds, α_j give bookmaker fixed effects. We address whether these margins have in general changed by estimating season fixed effects in $\phi_{t(ij)}$, where $t(ij)$ indicates the season when a match took place. The remaining heterogeneity is left in the residual term υ_{ij}.

The betting market is efficient according to the following sufficient condition, or null hypothesis, $H_0 : \beta_h = \beta_a = \beta_z = 0$. If we find that estimates of β_h, β_a or $-(\beta_h + \beta_a)$ are significantly positive or negative, then this would imply that home, away or draw results are under- or over-predicted by bookmaker odds, respectively. Similarly, a positive estimate of β_z would be consistent with a favourite-longshot bias being prevalent in market prices. Crucially, we estimate standard errors which are robust to clusters at the match level, as ε_{ij} will be highly correlated over bookmakers and outcomes for any given match. Not addressing this, or clustering in some other inappropriate dimension, such as at the bookmaker level, will lead to spuriously precise estimates and possible false rejection of the null hypothesis.[1]

We also test whether the information contained in linear combinations of odds from other bookmakers $k = 1,...,J-1 \neq j$ can significantly explain ε_i for any given bookmaker j. To do so, we apply the forecast encompassing method of Chong and Hendry (1986), estimating the following using least squares for each bookmaker j:

$$\varepsilon_i = \alpha + \beta_h h_i + \beta_a a_i + \beta_z z_{ij} + \sum_{\substack{k=1 \\ k\neq j}}^{J-1} \beta_k z_{ik} + \phi_{t(i)} + \eta_i, \quad E\left[\eta_i \mid \alpha, h_i, a_i, z_{ij}, z_{ik}, \phi_t,\right] = 0\,. \tag{2}$$

Under the null hypothesis that an individual bookmaker operates an efficient market, $H_0 : \beta_1 = ... = \beta_k = 0$. If we find that the linear combination of other bookmaker odds significantly explains another bookmaker's values of ε_i, then the odds of the latter are not taking into account relevant information, and this bookmaker cannot be operating an efficient market.

3 Data and Results

Results of the four leading English professional leagues, Premier League (PL), Championship (CH), League One (L1) and League Two (L2), were collected for the 2010/11 to 2017/18 seasons from Soccerbase.com. There were 16,407 matches: 3,039 in the PL; and 4,456 in each of the other leagues. Online betting odds from immediately before each match began and for each result outcome were extracted from Oddsportal.com for fifty-one bookmakers.[2] This matched panel of odds and matches is unbalanced, since not all bookmakers set odds for each match. The coverage of these bookmakers over matches was increasingly complete over the seasons, most likely reflecting the increasing competition and globalisation of online betting markets. In what follows, we focus on a reduced sample of thirteen bookmakers. These were selected using an arbitrary rule, whereby they set odds for at least 90% of the matches taking place in every season and each league. We check whether our findings generalise to the larger sample of fifty-one bookmakers.

To begin, we focus on the PL market, estimating variants of Equation (1). The results are summarised in Table 1. Column (I) shows results from estimating the regression model excluding the favourite-longshot bias term. The positive signs of $\widehat{\beta}_h$ and $\widehat{\beta}_a$ suggest home and away wins are over-predicted, relative to the draw, but there is no significant evidence that bookmaker odds excessively price in any particular result outcomes. Column (II) considers the favourite-longshot

[1] In practice, almost all the variance in the estimated residuals is between match-outcomes rather than within, i.e. $var_{ij}(\widehat{\overline{\upsilon}}_i)/var_{ij}(\widehat{\upsilon}_{ij})$>0.999, where $\widehat{\overline{\upsilon}}_i$ is the mean estimated residual of match-outcome i.

[2] Oddsportal.com also collect odds, and associated betting volumes for a small number of betting exchanges. These enable us to check the relative importance in terms of betting volumes of the different divisions in English football. On average volumes are much larger in the PL (mean £23,262, median £10,111) than in the CH (mean £3,044, median £1,540), L1 (mean £1,030, median £616) or L2 (mean £878.20, median £559).

bias only. The estimate is negative, consistent with odds favouring the longshot, but this is not significant, as found by Angelini and De Angelis (2019). Columns (III) and (IV) estimate Equation (1) looking at both potential sources of inefficiency, for the selected thirteen bookmakers and the larger set of fifty-one, respectively. There is no evidence in any of these estimates to reject the null hypothesis that the PL results betting market has been efficient over the past eight seasons.

Table 2 presents equivalent estimates of Equation (1) for each league as per column (III) of Table 1, with the PL results included again for comparison. In each of the leagues over this period, there is no significant evidence to reject the EMH in this context. In the language of the Mincer and Zarnowitz (1969) forecast efficiency framework, the odds give unbiased forecasts of football match outcomes and are generally efficient, subject to the bookmaker's rate of commission.[3]

Competition in betting markets has been increasing as transaction costs have fallen and the industry has increasingly moved online (Forrest, 2008). In this case, we would expect the average commission rate to be decreasing. Figure 1 plots the values of $\widehat{\phi}_j$ for each division using the model estimates from Table 2, thus describing the percentage point change in average commission rates since the 2010/11 season. The level of commission ranged between 5% in the Premier League and 7% in League 2 in 2010/11. In all four divisions and for the fixed sample for 13 bookmakers, commission rates significantly decreased by 0.3-0.5 points by the end of the 2017/18 season.

Finally, we estimate Equation (2) for each of the thirteen bookmakers in the reduced sample over all divisions, sample years and match outcomes, covering the 12,855 matches where all set odds, and testing whether a linear combination of the odds from the other 12 bookmakers can significantly help to predict ε_i. Table 3 summarises these results by displaying the p-values for the relevant Wald tests for each bookmaker.[4] We find that for most bookmakers a linear combination of the odds from competitors in the market would significantly predict their own odds-implied forecast errors. In this way, we can reject the EMH for seven of the thirteen bookmakers at the 0.1% level, 9 at the 1% level, and 12 at the 5% level of significance. Betting market prices for English football match results do not incorporate readily available and timely information which can be used to forecast outcomes.

4 Conclusion

At the overall market level, we found no statistically significant evidence which could reject an efficient market hypothesis of the online betting market for English football match results. The odds offered by online bookmakers were generally not biased towards any particular result outcome, nor did they feature the favourite-longshot bias, which has been documented in other betting markets. But individual bookmaker-specific markets are not efficient, since they fail to use

[3] Strong-form efficiency would dictate that there is no information not currently in the bookmaker's information set which could predict the odds-based forecast errors (Nordhaus, 1987).

[4] We do not display the other estimated model coefficient estimates as these are by construction identical across individual bookmaker regressions and uninteresting or insignificant, as per the general results.

the information contained in their competitors' odds. There is also suggestive evidence that the increased competition facing online bookmakers has reduced commission rates and profit margins.

Acknowledgements

We would like to thank Luca De Angelis for helpful comments.

Funding: Singleton is supported by the Economic and Social Research Council (UK) under Grant No. ES/J500136/1.

References

Angelini, G., and L. De Angelis. 2019. "Efficiency of online football betting markets." *International Journal of Forecasting*, 35(2): 712 – 721.

Chong, Y. Y., and D. Hendry. 1986. "Econometric evaluation of linear macro-economic models." *Review of Economic Studies*, 53(4): 671–690.

Fama, E. F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance*, 25(2): 383–417.

Forrest, D. 2008. "Soccer Betting in Britain." In *Handbook of Sports and Lottery Markets*. Eds. by D. B. Hausch, and W. T. Ziemba, San Diego Elsevier, 421 – 446.

Goddard, J., and I. Asimakopoulos. 2004. "Forecasting football results and the efficiency of fixed-odds betting." *Journal of Forecasting*, 23(1): 51–66.

Kuypers, T. 2000. "Information and efficiency: an empirical study of a fixed odds betting market." *Applied Economics*, 32(11): 1353–1363.

Levitt, S. D. 2004. "Why are gambling markets organised so differently from financial markets?" *Economic Journal*, 114(495): 223–246.

Mincer, J., and V. Zarnowitz. 1969. "The evaluation of economic forecasts." In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. NBER, 1–46.

Na, S., Y. Su, and T. Kunkel. 2018. "Do not bet on your favourite football team: the influence of fan identity-based biases and sport context knowledge on game prediction accuracy." *European Sport Management Quarterly*, 0(0): 1–23.

Nordhaus, W. D. 1987. "Forecasting Efficiency: Concepts and Applications." *The Review of Economics and Statistics*, 69(4): 667–674.

Ottaviani, M., and P. N. Sørensen. 2008. "The Favorite-Longshot Bias: An Overview of the Main Explanations." In *Handbook of Sports and Lottery Markets*. Eds. by D. B. Hausch, and W. T. Ziemba, San Diego Elsevier, 83 – 101.

Pope, P. F., and D. A. Peel. 1989. "Information, Prices and Efficiency in a Fixed-Odds Betting Market." *Economica*, 56(223): 323–341.

Reade, J. J. 2014. "Information and predictability: Bookmakers, prediction markets and tipsters as forecasters." *The Journal of Prediction Markets*, 8(1): 43–76.

Thaler, R., and W. T. Ziemba. 1988. "Parimutuel Betting Markets: Racetracks and Lotteries." *Journal of Economic Perspectives*, 2(2): 161–74.

Vlastakis, N., G. Dotsis, and R. N. Markellos. 2009. "How efficient is the European football betting market? Evidence from arbitrage and trading strategies." *Journal of Forecasting*, 28(5): 426–444.

TABLE 1: Model estimates and tests of betting market efficiency for English Premier League match results, 2010/11-2017/18

	(I)	(II)	(III)	(IV)
Home win: $\widehat{\beta}_h$	-0.005		-0.001	0.002
	(0.014)		(0.015)	(0.016)
Away win: $\widehat{\beta}_a$	-0.012		-0.011	-0.010
	(0.013)		(0.013)	(0.013)
Inverse odds: $\widehat{\beta}_z$		-0.016	-0.018	-0.013
		(0.032)	(0.035)	(0.036)
F-test: $\widehat{\beta}_h, \widehat{\beta}_a, \widehat{\beta}_z$			0.817	0.851
F-test: $-(\widehat{\beta}_h + \widehat{\beta}_a)$	0.484		0.619	0.733
F-test: $\widehat{\phi}_t$	0.000	0.000	0.000	0.000
Bookmaker FEs: $\widehat{\alpha}_j$	No	Yes	Yes	Yes
Bookmakers: J	13	13	13	51
Odds: N	117,285	117,285	117,285	325,644

Notes: First F-test shows p-values of the Wald test for restriction $H_0 : \beta_h = \beta_a = \beta_z = 0$. Second F-test shows p-values of the Wald test for restriction $H_0 : \beta_d = -(\beta_h + \beta_a) = 0$. Second F-test shows p-value of the Wald test for restriction $H_0 : \phi_{2010} = \cdots = \phi_{2017} = 0$.
*,** indicate significance at the 5%, and 1% levels, two-sided tests. Estimated standard errors in parentheses are robust to clusters at the match level.

TABLE 2: Estimates and tests of betting market efficiency in English professional football: comparison of preferred model over divisions, 2010/11-2017/18

	PL	CH	L1	L2
Home win: $\widehat{\beta}_h$	-0.001	-0.014	0.000	-0.013
	(0.015)	(0.015)	(0.014)	(0.014)
Away win: $\widehat{\beta}_a$	-0.011	0.001	0.008	0.011
	(0.013)	(0.011)	(0.011)	(0.011)
Inverse odds: $\widehat{\beta}_z$	-0.018	0.048	0.000	0.031
	(0.035)	(0.054)	(0.053)	(0.060)
F-test: $\widehat{\beta}_h, \widehat{\beta}_a, \widehat{\beta}_z$	0.817	0.748	0.901	0.416
F-test: $-(\widehat{\beta}_h + \widehat{\beta}_a)$	0.619	0.563	0.721	0.928
F-test: $\widehat{\phi}_t$	0.000	0.000	0.000	0.000
Bookmaker FEs: $\widehat{\alpha}_j$	Yes	Yes	Yes	Yes
Bookmakers: J	13	13	13	13
Odds: N	117,285	170,760	168,615	169,677

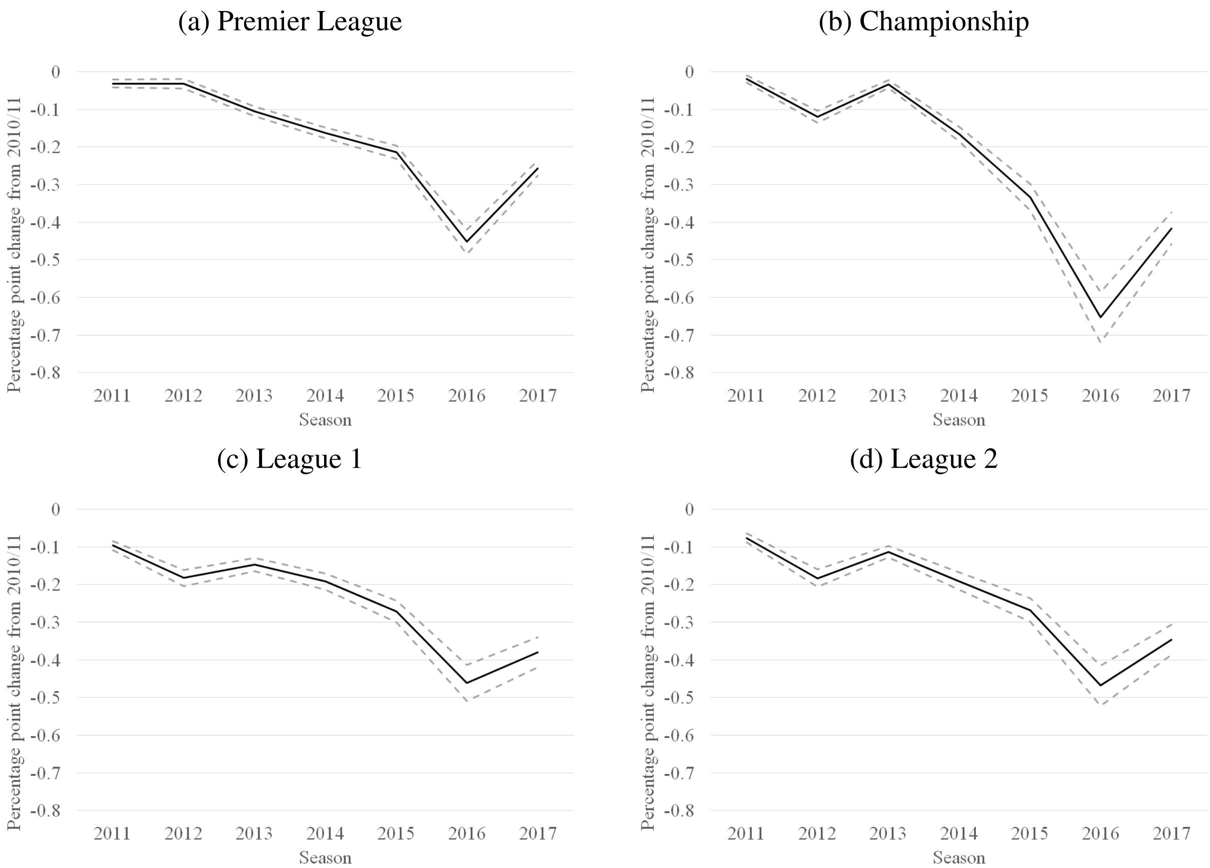
Notes: see Table 1
*,** indicate significance at the 5%, and 1% levels, two-sided tests. Estimated standard errors in parentheses are robust to clusters at the match level.

TABLE 3: Tests of individual bookmaker market efficiency, all divisions 2010/11-2017/18: do other bookmakers' odds add information?

Bookmaker	F-test: $\widehat{\beta_k}$
10Bet	0.0015
12Bet	0.0021
188BET	0.0000
5Dimes	0.0266
BetVictor	0.0000
Jetbull	0.0000
Leonbets	0.0930
Pinnacle	0.0137
SBOBET	0.0000
Titanbet	0.0001
William Hill	0.0000
bet365	0.0405
youwin	0.0000
Odds: N	38,655

Notes: F-test shows p-value of the Wald test for restriction $H_0 : \beta_1 = ... = \beta_k = 0$.

FIGURE 1: Estimated change in bookmaker commission rate on markets since the 2010/11 season



Notes.- dashed lines indicate 95% confidence intervals, estimated with standard errors robust to clusters at the match level.