Estimating Expected Loss Rates in Betting Markets: Theory and Evidence

Tadgh Hegarty[*] Karl Whelan[†]

April 2024

Abstract

If betting markets are efficient, in the sense of each bet on a contest having the same expected return, then there is a simple way to use the odds to calculate this common expected return. Guides to sports betting tell bettors to do this calculation to figure out the bookmaker's margin. We show that if bookmakers set higher profit margins for bets with lower probabilities of winning (as implied by the evidence on favorite-longshot bias) then average loss rates across all available bets will be higher than predicted by this widely-recommended calculation. We provide evidence from betting on soccer and tennis to illustrate that average loss rates across all available bets are consistently higher than predicted by the conventional calculation and that the magnitude of this difference is large.

Keywords: Sports Betting, Normalized Probabilities, Favorite-Longshot Bias

[*]tadgh.hegarty@ucdconnect.ie.
[†]karl.whelan@ucd.ie.

1. Introduction

Online sports betting with bookmakers has grown rapidly across the world in recent years, facilitated by mobile internet technologies that allow people to place bets with the touch of a button. Accompanying this has been an explosion in books and websites providing advice on betting. Probably the most common advice from these sources is that bettors should use the odds to estimate the bookmaker's expected gross profit margin on a contest, i.e. the bookmaker's profit rate before accounting for non-payout costs such as salaries or taxes. The higher this margin is, the less likely it is that bettors can earn a profit.

The bookmaker's margin goes by various names—in the US, it is often called the vigorish or "vig", the hold or the juice—and conventions on how to quote odds also vary across countries. So the descriptions can differ in style but the substance of the advice is the same. If bookmakers do not earn any gross profits then decimal odds (the payouts on successful $1 bets) will equal the inverse of the underlying probabilities and the sum of the inverses of the decimal odds will equal one. Given this, bettors are advised to calculate the sum of the inverses of these decimal odds, which is known in the bookmaking business as the "overround". The extent to which the overround is greater than one is determined by the bookmakers margin. More specifically, the inverse of the overround provides an estimate of the expected payout on a $1 bet. This expected payout can then also be used to figure out the so-called "normalized probabilities" of each bet's success implied by the odds.

In this paper, we show that the recommended overround-based formula for the expected payout on bets is correct if a betting market is efficient in the sense that the bookmaker's expected profit margins are equal across bets on each outcome of a game, a condition that Thaler and Ziemba (1988) termed "strong market efficiency." However, there is a large literature, dating back to Griffith (1949), demonstrating that sports betting markets tend to exhibit favorite-longshot bias: Losses from betting on longshots are larger than from betting on favorites. This bias is well known and it has been documented several times that the normalized probabilities implied by the assumption of market efficiency are too low for favorite bets and too high for longshots.[1]

In contrast, we believe the implications of favorite-longshot bias for expected payout calculations are not known. Normalized probabilities are, by construction, correct on average because they sum to one. However, we show that if bookmakers have higher profit margins for bets that are less likely to win, then the average loss rate across all available bets will be higher than implied by the overround formula. We illustrate this result using large datasets on odds and outcomes from betting on soccer and tennis. We show that the magnitude of this bias is large. Across all bets in our soccer data set, average loss rates for betting on soccer are one-fifth higher than implied by the overround formula and while the corresponding figure for the tennis data set is forty percent.

[1] See, for example, Berkowitz, Depken and Gandar (2018) and Hegarty and Whelan (2023a).

2. The Overround Formula with an Efficient Betting Market

Unlike pari-mutuel betting, which pools all bets and pays the funds out (minus a fraction to cover costs and profits) to those who picked the winner in proportion to the size of their bet, the modern online betting industry offers fixed-odds bets offered by bookmakers. Bookmakers make offers such as "You get back \$3 if your bet wins and lose your \$1 bet otherwise" and this offer is not affected by the actions of subsequent bettors. In this example, 3 is the decimal odds.

There are many online resources aimed at informing people about how fixed-odds betting markets work, most of them containing advertising for betting websites. These online resources place a key emphasis on the need to calculate the bookmaker's margin or "vig" when evaluating a bet. Discussions of this issue vary in their sophistication. To illustrate, consider a sporting event with N possible outcomes in which bookmakers offer decimal odds O_i on outcome i occurring. The less sophisticated resources tell bettors to calculate the margin by subtracting one from the overround (the sum of the inverses of the decimal odds)[2]

$$m = \sum_{i=1}^{N} \frac{1}{O_i} - 1 \tag{1}$$

So, for example, if the overround is 1.045, bettors can infer that the bookmaker's margin is 4.5%. This suggests that, for every dollar placed, the bookmaker stands to earn an average of 4.5 cents. The more sophisticated resources instead tell bettors to calculate the bookmaker's margin as[3]

$$m = 1 - \frac{1}{\sum_{i=1}^{N} \frac{1}{O_i}} \tag{2}$$

In this case, if the overround is $v = 1.045$, the bookmaker's margin is $1 - \frac{1}{1.045} = 0.043$. As we show below, under specific conditions, this second formula correctly predicts that the expected return for a bookmaker on each dollar staked by bettors. Under these conditions, the expected payout to a bettor on a one dollar bet is

$$\pi = 1 - m = \frac{1}{\sum_{i=1}^{N} \frac{1}{O_i}} \tag{3}$$

We will term this "the overround formula" for the expected payout. For relatively small margins the calculations from equations 1 and 2 will be very similar because for low values of x, the approximation $x \approx 1 - \frac{1}{1+x}$ will work well.

To derive the conditions under which the calculated margin in equation 2 is correct, we will assume that bookmakers know the true probabilities P_i that outcome i will occur and that the bookmaking market corresponds to Thaler and Ziemba's (1988) definition of strong-form efficiency which

[2] Here is an example https://www.legalsportsreport.com/sports-betting/vigorish/

[3] Here, for example, https://bookies.com/guides/what-is-the-vigorish

implies that all bets on the same event should have the same expected rate of return.[4] This means that bookmakers set decimal odds so that the expected payout on each bet is given by

$$P_i O_i = \mu \quad i = 1, ..., N \tag{4}$$

where μ is the common expected payout across all bets on the event. The requirement that the probabilities sum to one gives us the following

$$\sum_{i=1}^{K} P_i = \sum_{i=1}^{K} \frac{\mu}{O_i} = 1 \tag{5}$$

which can be re-expressed as

$$\mu = \frac{1}{\sum_{i=1}^{K} \frac{1}{O_i}} = \pi \tag{6}$$

In other words, the actual expected payout on all bets (μ) equals the overround-based calculation of equation 3. The underlying probabilities can also then be estimated correctly as the "normalized" probabilities defined as

$$P_k = \frac{\pi}{O_k} \tag{7}$$

3. Implications of Favorite-Longshot Bias

Here, we describe how favorite-longshot bias impacts the accuracy of both normalized probabilities (the inaccuracy of which has previously been documented) and the overround-based estimate of the expected payout (which has not been previously shown).

3.1. Normalized Probabilities

The accuracy of the overround formula for the expected payout relies on the assumption that betting markets feature strong-form efficiency. However, there is a large literature documenting that bookmakers tend to make bigger profits from bets on longshots than bets on favorites. Many different explanations have been offered but, from our perspective, the key point is just that such a pattern exists.[5] We provide our own examples of this pattern from data on soccer and tennis betting below.

To illustrate the implications of this pattern, assume now that odds are determined by the book-

[4]Technically, Thaler and Ziemba (1988) defined a strong form of efficiency for a betting market as being the property that "*All bets should have expected values equal to (1–t) times the amount bet*" where t was the track tack from pari-mutuel betting, which was the focus of their research. However, the generalization to betting markets with odds set by bookmakers is clear.

[5]Snowberg and Wolfers (2008) and Ottaviani and Sørensen (2008) are excellent surveys of the theoretical and empirical literature on the favorite-longshot bias.

maker according to

$$O_i = \frac{\mu_i}{P_i} \quad \text{where} \quad \frac{d\mu_i}{dP_i} > 0 \quad i = 1, ..., N \tag{8}$$

so there are separate expected payout rates for each bet and the payout rates μ_i depend positively on the P_i. In this case, bookmakers explicitly set odds to make higher profit margins on bets with lower probabilities of success.

Now consider first the estimated probabilities based on the assumption of market efficiency. These are calculated by using the overround to estimate the expected payout rate, which we will now denote as $\hat{\mu}$. With favorite-longshot bias, the overround-based estimate of the expected payoff can be expressed in terms of the probabilities and expected payouts (which are unobserved to bettors) as

$$\pi = \frac{1}{\sum_{i=1}^{N} \frac{P_i}{\mu_i}} \tag{9}$$

The normalized probabilities can be re-expressed as follows:

$$\hat{P}_i = \frac{\pi}{O_i} = \frac{\pi}{\frac{\mu_i}{P_i}} = \frac{\pi}{\mu_i} P_i = \left(\frac{1}{\mu_i \sum_{j=1}^{N} \frac{P_j}{\mu_j}} \right) P_i \tag{10}$$

The term in the denominator of the fraction multiplying P_i can be written as

$$\mu_i \sum_{j=1}^{N} \frac{P_j}{\mu_j} = P_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} P_j \frac{\mu_i}{\mu_j} \tag{11}$$

Now consider the implications of favorite-longshot bias for this calculation. It calculates a probability weighted average of 1 and a set of terms of the form $\frac{\mu_i}{\mu_j}$. Suppose outcome i has the lowest probability and thus the lowest value of μ_i. Then the terms in the $\frac{\mu_i}{\mu_j}$ will all be less than one and the overall sum in equation 11 will be less than one. This will imply $\hat{P}_i > P_i$. The same logic says that $\hat{P}_i < P_i$ for the outcome with the highest probability and that the size and sign of the bias in probability estimates will depend monotonically on the size of the underlying probability. While the normalized probabilities are biased in one direction for low true probabilities and in the other direction for high true probabilities, the fact that they sum to one means that they will at least on average be correct.

We will note that while Strumbelj (2014), Berkowitz, Depken and Gandar (2018) and Hegarty and Whelan (2023a) have all presented empirical evidence of biases in normalized probabilities in various fixed-odds betting markets, this formal proof that these biases are caused by there being higher margins for lower probability bets, is to our knowledge new.

3.2. Expected Payouts

What are the properties now of the standard overround-based calculation of the expected payout? Normalized probabilities, while biased, will on average be correct but this is not the case for expected payouts. Equation 9 is a complex function of the N separate payout rates, μ_i set by the bookmaker. It is the inverse of a weighted sum of the inverses of the payout rates, where the probabilities of the outcomes are the weights (technically it is a probability-weighted harmonic mean of the payout rates).

We want to compare π with the average payout rate across all bets, which can be calculated as a simple average of the separate payout rates, μ_i. We might hope in the complexity of equation 9—in which the inverse of the expected payouts are weighted by probabilities and then the inverse is taken—that the two inverse operations essentially cancel, so that π can be well approximated as a simple linear function of the expected payouts. We show in an appendix that this is indeed the case. When there is favorite-longshot bias, the overround formula for the expected payout can be approximated by the probability weighted mean of the payout rates, which we will denote $\bar{\mu}^p$

$$\pi \approx \sum_{i=1}^{N} P_i \mu_i = \bar{\mu}^p \tag{12}$$

The appendix shows that this approximation works well as long as

$$\epsilon = \sum_{i=1}^{N} P_i \left(\frac{\mu_i - \bar{\mu}^p}{\bar{\mu}^p} \right)^2 \tag{13}$$

is small. We show below, using the variations in observed payout rates ranging from favorites to longshots in our data as proxies for the μ_i values and the resulting normalized probabilities as proxies for the P_i values, that the average value for ϵ appears to be very small, so the approximation works well in practice.

The implication of these calculations is that when bettors calculate expected returns using the overround formula, they are unknowingly placing more weight on the expected return from betting on favourites than on the expected return on longshots. We can compare the expected payout implied by the overround formula with the average expected payout rate across all available bets. Favorite-longshot bias means P_i and μ_i are positively correlated, so we can conclude that

$$\frac{1}{N} \sum_{i=1}^{N} \mu_i < \sum_{i=1}^{N} P_i \mu_i = \bar{\mu}^p \approx \pi \tag{14}$$

because $\bar{\mu}^p$ places more weight on the higher values of μ_i than the simple average. This means the average payout across all available bets is less than suggested by the overround formula.

An alternative possibility could be that the overround formula's expected payout rate represents the average payout across all bets that have actually been placed, rather than the simple average across all available bets. This average payout across bets placed is not generally observable because bookmakers do not publish data on betting volumes. However, it is unlikely that betting volumes are strictly proportional to the underlying probabilities. While people often think that the highest volume of bets must be placed on favorites, this doesn't necessarily confirm with either theory or practice. In the baseline case where markets are efficient, the odds for each bet should be equally attractive, suggesting an equal split among bets as a reasonable baseline outcome. And the few studies that we do have on volumes in fixed-odds betting markets, such as Strumpf (2003), Levitt (2004) and Flepp, Nüesch and Franck (2016) have shown that odds are set in a more complex way than just to align the fraction of bets placed with the underlying probabilities. It seems likely, then, that the overround formula's expected payout rate also overstates the average payout on bets placed.

4. Evidence From Betting on Soccer and Tennis

Here we describe two datasets with odds and outcomes from sporting events. We first show both datasets exhibit a favorite-longshot bias and then confirm the predictions just derived for estimated normalized probabilities and the estimated payout rates.

4.1. Data and Favorite-Longshot Bias

We use two datasets made available by gambling expert Joseph Buchdahl. From www.football-data.co.uk, we obtain outcomes and average closing odds on home wins, away wins and draws across a wide range of bookmakers for 84,230 European professional soccer matches, spanning the 2011/12 to 2021/22 seasons for 22 European soccer leagues (listed in the appendix). From www.tennis-data.co.uk, we have outcomes and average closing odds for 55,988 professional men's and women's tennis matches on the ATP and WTA tours between 2011 and 2022.

We divided all possible bets (252,690 on soccer and 111,976 on tennis) into deciles according to their decimal odds. The charts in Figure 1 show average payouts for a unit bet for each decile of odds. The data clearly suggest that these betting markets do not satisfy strong-form efficiency. Average payouts decline as the odds rise, dropping off particularly for the upper deciles of odds.

This pattern of low average payouts for longshot bets is somewhat larger for tennis than for soccer. For soccer, bets in the highest odds decile have an average payout on a \$1 bet of only \$0.83 (meaning an average loss of 17%) while bets in the lowest odds decile have only a 3% average loss rate. For tennis, the pattern is even more extreme, with bets in the bottom decile losing 26% on average while bets in the top decile lose only 3%. Standard t tests for differences of means across these deciles strongly reject the hypotheses of the mean payouts for the lower deciles of odds being the same as for the higher deciles. This evidence confirms the existing findings using smaller datasets of Angelini and de Angelis (2019) for soccer and Forrest and McHale (2007) for tennis.

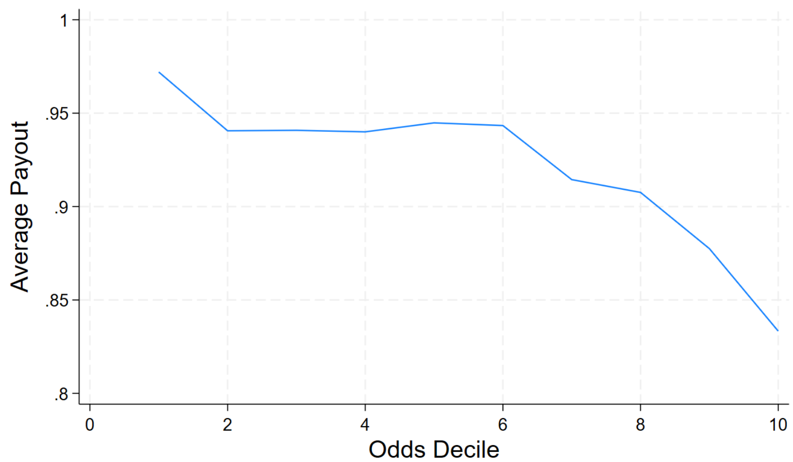
4.2. Normalized Probabilities

Figure 2 confirms the result derived above that the presence of favorite-longshot bias means normalized probabilities are too high when the true probability is low and too low when the true probability is high. The charts divide all bets in the datasets into 20 quantiles organized by normalized probabilities and calculates the actual fraction of winning bets for each quantile. If the probabilities were accurate then the outcomes would follow a 45 degree line. However, we can see from both the upper panel (for the soccer data) and the lower panel (the tennis data) that actual win rates lie below the 45 degree line for low probabilities and above it for high probabilities.

The deviations of these probability estimates from the 45 degree line may seem small but they are statistically significant for both high and low estimated probabilities. Also, for low values, the deviations are a big percentage of the estimated probabilities, consistent with the large average loss estimates above. Ultimately, the favorite-longshot bias in payouts occurs because longshot bets don't win as often as the odds suggest they should.

Figure 1: Average payout rates for bets by deciles of the decimal odds (1 = lowest odds, 10 = highest odds)

(a) Soccer



(b) Tennis

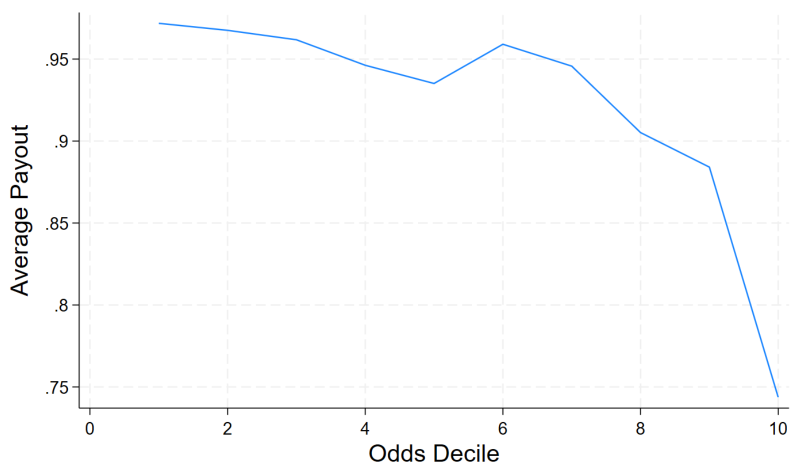
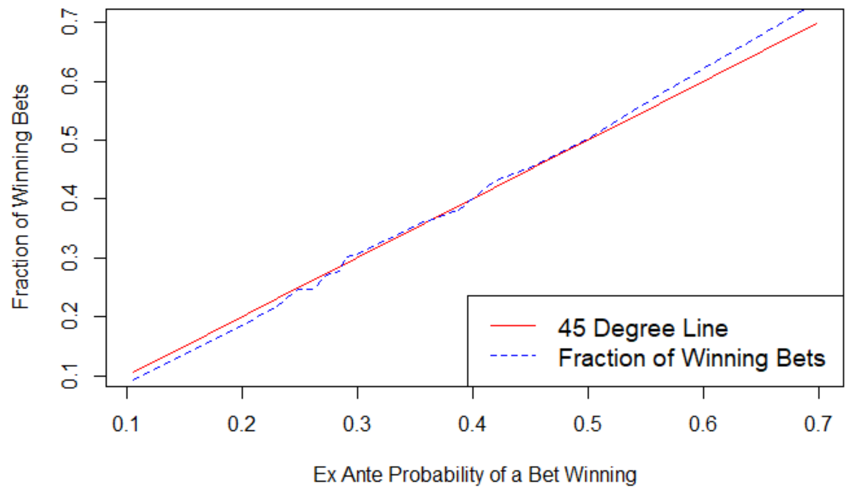
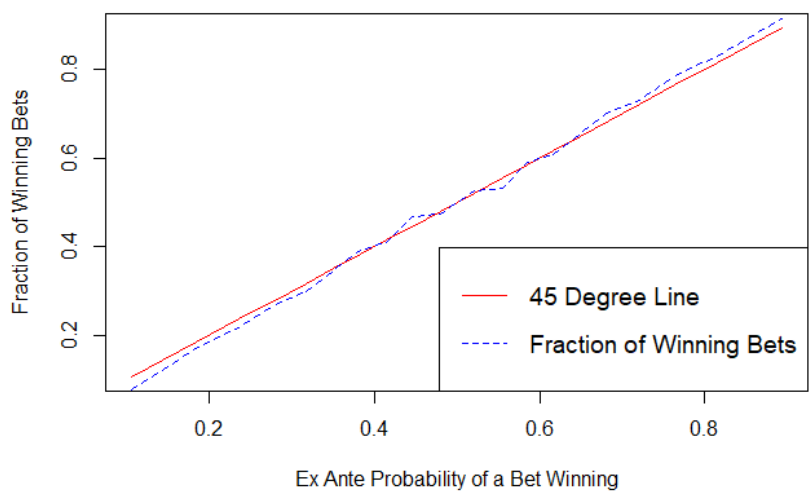


Figure 2: Actual Fraction of Wins Sorted by Normalized Probabilities

(a) Soccer



(b) Tennis



4.3. Expected Payouts

We start with a method for checking whether the approximation we derived for the expected payout in equation 12 is accurate. In the appendix, we show the accuracy of this approximation depends on ϵ as defined in equation 13. This approximation error, which will vary for each match, depends on the probabilities P_i of each outcome in a match as well as the expected payout μ_i for each bet. We do not observe either of these but we do observe a consistent pattern of average payouts declining as the odds increase, as documented in Figure 1. Given this, we can estimate the size of the approximation error by assigning an expected payout to each bet based on which decile its odds are in. With an assumed expected payout for the bets, we can directly derive the probability that the bet will win from the observable odds because by assumption $P_i = \frac{\mu_i}{O_i}$. Based on this, we can construct $\bar{\mu}^p$ as defined in equation 12 and then use this to construct an ϵ as defined in equation 13 for each match.

For the soccer data, the average value of ϵ is 0.0012. From equation B.5 in the appendix, this means the expected payout rate implied by the overround formula will be about 0.1% below the probability weighted sum of the probability-specific payout rates. The average value of ϵ for the tennis data was 0.0016, again implying the approximation error is very small.

These small approximation errors mean average payout rates estimated by the overround formula will be very well approximated by a weighted average of expected payout rates, where the weights are the probabilities P_i of the bets being successful. This means average payout rates across all bets will tend to be lower than predicted by the overround formula. Table 1 (for soccer) and Table 2 (for tennis) confirm this prediction. For soccer, the average loss rate predicted by the overround formula is 6.5% while the actual average loss rate across all bets is 7.8%, so losses are twenty percent higher than predicted. For tennis, the average loss rate predicted by the overround formula is 5.4% while the actual average loss rate across all bets is 7.4%, so losses are almost forty percent higher than predicted. In both cases, t-tests strongly reject the hypotheses that the means of the actual loss distributions are equal to the means obtained from the overround equation.

The tables also show this pattern has been relatively stable over time. Both average realized loss rates and the loss rates predicted by the overround formula have fallen over the past decade, perhaps reflecting greater competition in the sports betting market. However, for each year, realized average loss rates across all bets have been larger than predicted by the overround formula.

Figure 3 further illustrates this finding by sorting all matches in the two samples into 20 quantiles according to their predicted average loss rate from the overround formula and displaying their actual average loss rates across all bets. Across the full range of quantiles (apart from the bottom soccer quantile) the actual average loss rates are larger than the expected loss rates implied by the overround formula. The larger deviations of outcomes from those predicted by the overround formula for tennis in the bottom deciles are consistent with its pattern of favorite-longshot bias being stronger.

Table 1: Average loss rates across all available soccer bets compared with loss rates implied by overround formula
(N = number of matches)

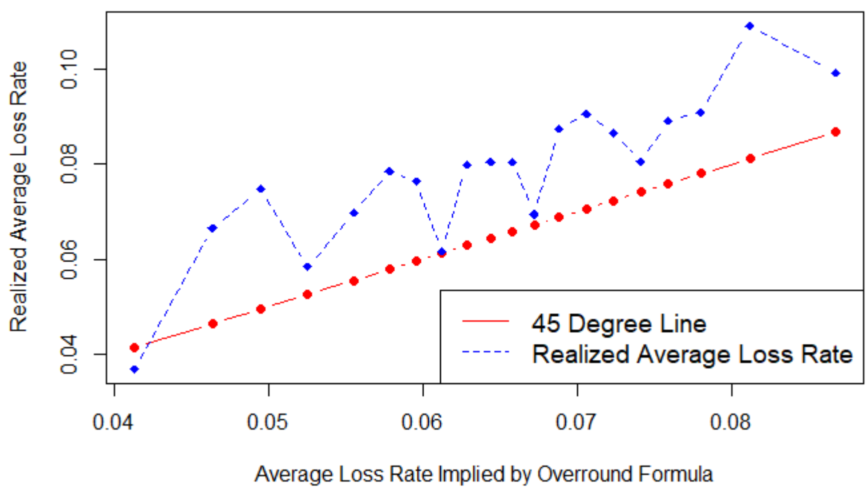
Season	Loss Rates Implied by Overround Formula	Realized Average Loss Rates	N
All Seasons	6.5%	7.8%	84,230
2011 / 2012	7.5%	9.2%	7,694
2012 / 2013	7.0%	7.7%	7,705
2013 / 2014	6.9%	8.6%	7,616
2014 / 2015	6.6%	8.1%	7,841
2015 / 2016	6.6%	7.7%	7,801
2016 / 2017	6.6%	8.1%	7,841
2017 / 2018	6.4%	8.5%	7,794
2018 / 2019	6.0%	7.4%	7,661
2019 / 2020	5.9%	6.1%	6,893
2020 / 2021	5.8%	7.0%	7,644
2021 / 2022	5.6%	7.5%	7,740

Table 2: Average loss rates across all available tennis bets compared with loss rates implied by overround formula
(N = number of matches)

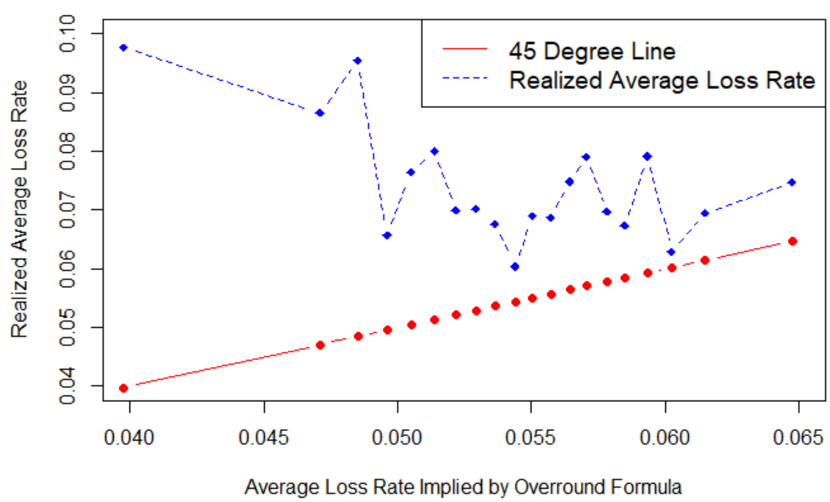
Year	Loss Rates Implied by Overround Formula	Realized Average Loss Rates	N
All Years	5.4%	7.4%	58,112
2011	6.0%	9.5%	5,124
2012	5.8%	8.4%	5,011
2013	5.7%	8.5%	5,066
2014	5.6%	7.5%	5,071
2015	5.7%	8.3%	5,145
2016	5.5%	6.8%	5,141
2017	5.3%	5.8%	5,127
2018	4.9%	6.8%	5,104
2019	5.0%	6.9%	5,080
2020	5.0%	7.1%	2,321
2021	5.1%	7.7%	4,929
2022	5.2%	5.5%	4,993

Figure 3: Average loss rates across all available bets compared with loss rates implied by overround formula: Sorted by overround formula loss rate into 20 quantiles

(a) Soccer



(b) Tennis



4.4. On Shin Probabilities

Strumbelj (2014) and others have noted that normalized probabilities appear to be biased for many fixed-odds betting markets. Strumbelj documented that an alternative set of estimated probabilities, based on Shin's (1993) model, do better in predicting actual win rates. The reason for this is that Shin's model incorporates a fraction of bettors who are insiders who know what the outcome of the event is going to be and this feature imparts a favorite-longshot bias on the odds. Cain, Law and Peel (2001) showed how to use Shin's model to take odds for an individual game and simultaneously estimate the fraction of insiders and the win probabilities of the contestants. These probability estimates "unwind" the favorite-longshot pattern in the odds when mapping them into probabilities.

Cain, Law and Peel's method shows one way to recover probabilities that have a better predictive performance than normalized probabilities. However, Shin's model is not useful in recovering better estimates of the average expected payouts on bets because it assumes that competition among bookmakers results in gross profit margins of zero, which is clearly counter-factual. It may be possible, however, to apply a theoretical model that predicts a pattern of favorite-longshot bias in the odds and predicts non-zero gross profit margins and use such a model to improve estimated average payout rates. The monopoly bookmaker model presented by Hegarty and Whelan (2023b), where disagreement among bettors results in a favorite-longshot bias and average margins depend on the elasticity of demand, may be a promising approach.

5. Conclusions

Betting on sports is growing rapidly around world. Many guides exist to help those new to sports betting to understand how it works. A key element of their guidance is that bettors should use the overround formula to calculate the bookmaker's profit margin and thus the amount that bettors should expect to lose.

We have shown that when bookmakers set higher profit margins for bets with a lower likelihood of winning—as is the case in many betting markets such as the ones for soccer and tennis reported here—the overround formula understates the average loss rates across all available bets. In our examples, actual average loss rates across all available bets are one-fifth higher than predicted for betting on soccer and forty percent higher for betting on tennis. We recommend that advice for those interested in gambling on sports should be updated to inform people that they will likely lose more on average on the bets offered by bookmakers than is indicated by the calculation that is currently widely recommended.

References

[1] Angelini, Giovanni and Luca De Angelis (2019). "Efficiency of online football betting markets," *International Journal of Forecasting*, Volume 35, pages 712–721.

[2] Berkowitz, Jason, Craig A. Depken and John M. Gandar (2018). "The conversion of money lines into win probabilities: Reconciliations and simplifications," *Journal of Sports Economics*, Volume 19, pages 990-1015.

[3] Cain, Michael, David Law and David Peel (2001). "The Incidence of Insider Trading in Betting Markets and the Gabriel and Marsden Anomaly," *Manchester School*, Volume 69, pages 197-207.

[4] Flepp, Raphael, Nüesch, Stephan and Egon Franck (2016). "Does Bettor Sentiment Affect Bookmaker Pricing?" *Journal of Sports Economics*, Volume 17, pages 3-11.

[5] Forrest, David and Ian Mchale (2007) "Anyone for tennis (betting)?" *The European Journal of Finance*, Volume 13, pages 751-768.

[6] Griffith, Richard (1949). "Odds adjustment by American horse race bettors," *American Journal of Psychology*, Volume 62, pages 290–294.

[7] Hegarty, Tadgh and Karl Whelan (2023a). Forecasting soccer matches with betting odds: A tale of two markets. CEPR discussion paper No. 17949.

[8] Hegarty, Tadgh and Karl Whelan (2023b). Disagreement and market structure in betting markets: Theory and evidence from European Soccer. CEPR Discussion Paper No. 18144.

[9] Levitt, Steven (2004). "Why Are Gambling Markets Organised So Differently from Financial Markets?" *The Economic Journal*, Volume 114, pages 223-246.

[10] Ottaviani, Marco and Peter Norman Sørensen (2008). "The favorite-longshot bias: An overview of the main explanations," in *Handbook of Sports and Lottery Markets*, Elsevier.

[11] Shin, Hyun Song (1993). "Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims," *Economic Journal*, Volume 103, pages 1141-1153.

[12] Strumpf, Koleman (2003). Illegal sports bookmakers. Unpublished manuscript. Available at https://users.wfu.edu/strumpks/papers/Bookie4b.pdf

[13] Snowberg, Erik and Justin Wolfers (2008). "Examining explanations of a market anomaly: Preferences or perceptions?" in *Handbook of Sports and Lottery Markets*, Elsevier.

[14] Strumbelj, Erik (2014). "On Determining Probability Forecasts from Betting Odds, " *International Journal of Forecasting*, Volume 30, pages 934-943.

[15] Thaler, Richard and William Ziemba (1988). "Anomalies: Parimutuel betting markets: racetracks and lotteries," *Journal of Economic Perspectives*, 2, 161-174.

A Soccer Leagues in the Dataset

Table 3: The 22 soccer leagues in the dataset

Nation	Number of Divisions	Division(s)
England	5	Premier League, Championship, League 1 & 2, Conference
Scotland	4	Premier League, Championship, League 1 & 2
Germany	2	Bundesliga 1 & 2
Spain	2	La Liga 1 & 2
Italy	2	Serie A & B
France	2	Ligue 1 & 2
Belgium	1	First Division A
Greece	1	Super League Greece 1
Netherlands	1	Eredivisie
Portugal	1	Primeira Liga
Turkey	1	Super Lig

B Approximation Result

We obtain the approximation described in equation 12 with an approach used to derive Jensen's inequality. Using Taylor series, we can write provide a second-order approximation of any function of the individual payouts, μ_i as

$$F(\mu_i) \approx F(\bar{\mu}^p) + F'(\bar{\mu}^p)(\mu_i - \bar{\mu}^p) + \frac{F''(\bar{\mu}^p)(\mu_i - \bar{\mu}^p)^2}{2} \tag{B.1}$$

The inverse of the overround-based estimated of the expected payout π is given by

$$\frac{1}{\pi} = \sum_{i=1}^{N} \frac{P_i}{\mu_i} \tag{B.2}$$

Applying the Taylor series approximation in equation B.1 to $F(x) = \frac{1}{x}$ around the point $\bar{\mu}^p$, we get

$$\frac{1}{\pi} \approx \frac{1}{\bar{\mu}^p} - \frac{\mu_i - \bar{\mu}^p}{(\bar{\mu}^p)^2} + \frac{(\mu_i - \bar{\mu}^p)^2}{(\bar{\mu}^p)^3} \tag{B.3}$$

Taking expectations using the P_i terms as probabilities and using the definition of $\bar{\mu}^p$ in equation 12 , the middle term on the right equals zero and we get

$$\frac{1}{\pi} \approx \frac{1}{\bar{\mu}^p} + \sum_{i=1}^{N} \frac{P_i (\mu_i - \bar{\mu}^p)^2}{(\bar{\mu}^p)^3} \tag{B.4}$$

The inequality $\frac{1}{\pi} < \frac{1}{\bar{\mu}^p}$ that this implies is an application of Jensen's inequality for convex functions because $F(x) = \frac{1}{x}$ is convex for positive x. This can be re-written as

$$\pi \approx \frac{\bar{\mu}^p}{1 + \epsilon} \tag{B.5}$$

where

$$\epsilon = \sum_{i=1}^{N} P_i \left(\frac{\mu_i - \bar{\mu}^p}{\bar{\mu}^p} \right)^2 \tag{B.6}$$

If ϵ is sufficiently small, then we can write this approximation as

$$\pi \approx \bar{\mu}^p \tag{B.7}$$