Calculating The Bookmaker's Margin: Why Bets Lose More On Average Than You Are Warned

Tadgh Hegarty* Karl Whelan†

University College Dublin

December 2023

Abstract

Guides to sports betting tell bettors how to use quoted odds to calculate the expected loss rate on bets due to the bookmaker's margin. We show that if betting markets are efficient, in the sense of each bet on a contest having the same expected return, then the recommended calculation is correct. However, we also 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 on available bets are consistently higher than predicted by the conventional calculation.

Keywords: Sports Betting, Gambling Losses, Favorite-Longshot Bias

JEL Classification: G14, L83, Z20

*tadgh.hegarty@ucdconnect.ie.

†karl.whelan@ucd.ie.

1. Introduction

Online sports betting has grown rapidly across the world in recent years. This trend is particularly evident in the US after a 2018 Supreme Court decision declared the federal prohibition on betting on sports to be unconstitutional. By 2023, 38 states had legalized sports betting in various forms.[1] Betting in these markets is already large, with $277 billion placed in legal sports betting markets in the time between the Supreme Court ruling and October 2023.[2] The huge amount of money being spent on advertising on sports betting in the US suggests the newly-legalized bookmaking firms believe this market is going to grow substantially over the next few years.[3]

Unlike pari-mutuel racetrack 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. In other words, they 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 known as the "decimal odds" for this bet.

The rise in online sports betting has been accompanied by an explosion in books and websites providing advice on fixed-odds betting. One of the key pieces of advice from these sources is that bettors should use the odds to calculate the bookmaker's expected gross profit margin on a contest, i.e. the bookmaker's profit earned on bets before accounting for costs such as salaries or taxes. This 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: Bettors should calculate the sum of the inverses of the decimal odds, known as the "overround". The inverse of the overround will then tell them the expected payout on a $1 bet, which will be a number less than one.

In this paper, we show that the recommended overround-based formula for the expected payout on bets is correct if the betting market is efficient in the sense that the bookmaker's expected profit margins are equal across bets on each outcome of a game. 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. While this bias is well known, we believe its implications for expected payout calculations are not. 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 the odds and outcomes from betting on soccer and tennis.

[1] See https://www.americangaming.org/research/state-gaming-map/

[2] Data from https://www.legalsportsreport.com/sports-betting/revenue/

[3] Forbes have reported that total advertising spending by US sportsbooks is projected to be about $2 billion in 2023. https://www.forbes.com/sites/bradadgate/2023/09/01/more-sportsbook-ads-are-running-in-non-sports-programming/

2. The Overround Formula with an Efficient Betting Market

Consider a sporting event with N possible outcomes. Bookmakers offer decimal odds O_i on outcome i occurring, meaning O_i is the total payout (inclusive of the original stake) from betting $1 on outcome i when this outcome occurs. For "spread bets" on two-outcome sports, which have traditionally been popular in the US, the gambler bets on a scoreline adjusted to add points to the underdog's score and equal odds are set for both bets. However, internationally, most sports betting involves betting on the potential outcomes of the actual contest, with higher odds offered on the less likely outcomes. This type of "moneyline" betting is now a major feature of the newly-legal US sportsbooks.

There is a wealth of online resources aimed at informing people about how fixed-odds betting markets work, most of it containing advertising for betting websites. These resources place a key emphasis on the need to calculate the bookmaker's margin or "vig" when evaluating a bet. Discussions of these issues vary in their sophistication. The less sophisticated resouces tell bettors to calculate the margin by subtracting one from the overround (the sum of the inverses of the decimal odds)[4]

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

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%. The more sophisticated resources instead tell bettors to calculate the bookmaker's margin as[5]

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

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 will be 4.3% with the same figure being the expected loss rate for the bettor. Under these conditions, the expected payout 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 book-

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

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

making market corresponds to Thaler and Ziemba's (1988) definition of strong-form efficiency which implies that all bets on the same event should have the same expected rate of return."[6] 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

$$P_k = \frac{1}{\sum_{i=1}^{K} \frac{1}{O_i}} \frac{1}{O_k} \tag{7}$$

3. Favorite-Longshot Bias

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.[7] We provide our own examples of this pattern from data on soccer and tennis betting below.

We will assume now that odds are determined by the bookmaker 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

[6] 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.

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

probabilities of success.

What are the properties now of the standard overround-based calculation of the expected payout? In this case, the calculation can be expressed in terms of the probabilities and expected payouts (which are unobserved to bettors) as

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

This 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 complex calculation 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 $\bar{\mu}^p$

$$\pi \approx \sum_{i=1}^{N} P_i \mu_i = \bar{\mu}^p \tag{10}$$

and this approximation works well as long as $\frac{\text{Var}(\mu_i)}{(\bar{\mu}^p)^2}$ is small. Using the variations in observed payout rates ranging from favorites to longshots in our datasets below as proxies for the μ_i values, we have found that this calculation produces a small number so the approximation works well in practice.

This approximation allows us to compare the expected payout implied by the overround formula and 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

$$\frac{1}{N} \sum_{i=1}^{N} \mu_i < \sum_{i=1}^{N} P_i \mu_i = \bar{\mu}^p \approx \pi \tag{11}$$

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 hope may 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. 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. While there do appear to be inefficiencies in fixed-odds betting markets, the odds for each bet still have to be attractive enough to get people to bet on them and if they are not, bookmakers will adjust them upwards. Indeed, recent evidence on moneyline bets on US sports from Moscowitz and Vasudevan (2022) shows the number of bets placed as being relatively equal across the different deciles by estimated win probability. This suggests the average payout rate for bettors across bets placed will be closer to an equally weighted payout rate across all bets. It seems likely, then, that the overround formula's expected payout rate also overstates the average payout obtained by bettors.

4. Evidence From Betting on Soccer and Tennis

To provide empirical examples of the discrepancy between expected loss rates implied by the overround formula and realized average loss rates across all available bets, we use two datasets both made publicly available by gambling expert and author, Joseph Buchdahl. From www.football-data.co.uk, we obtain betting odds for each possible outcome (home win, away win and draw) and outcomes for 84,230 European professional soccer matches, spanning the 2011/12 to 2021/22 seasons for 22 European soccer leagues across 11 different nations as described in Table 1. From www.tennis-data.co.uk, we have odds and outcomes for 58,112 professional men's and women's matches played across the world on the ATP and WTA tours. Our measure of betting odds is the average closing odds across a wide range of online bookmakers surveyed by Buchdahl. While his sites also report the maximum odds available from bookmakers on each match, these odds tend to only be available as promotional bets with limited stakes allowed and they do not represent the typical market odds available for most bets on an event.

Figure 1 displays two bar charts that divide all bets in our samples into deciles by their predicted probability of success according to the normalized probability method described in equation 7.[8] For each decile, it displays the average payout on these bets per dollar staked. This figure is useful for two purposes. First, despite these betting markets having high volumes and many different competing providers, the odds clearly display an important inefficiency with a clear pattern of favorite-longshot bias evident. For soccer, bets in the lowest decile for estimated probability of success have an average payout on a $1 bet of only $0.83 (meaning an average loss of 17%) while bets in highest decile have only a 2% average loss rate. For tennis, the pattern is even more extreme, with bets in the bottom decile losing 23% on average while bets in the top decile lose only 3%. This evidence confirms the existing findings using smaller datasets of Angelini and de Angelis (2019) for soccer and Forrest and

[8] In Appendix B, we show that the presence of favorite-longshot bias means normalized probability estimates are biased: when the true probabilities are low, normalized probabilities over-estimate them with the opposite applying when the true probabilities are high. However, there is still a monotonic relationship between the estimated probabilities and the true probabilities, so the pattern reported in the bar chart would not be affected by this bias.

McHale (2007) for tennis.

Second, we can use the data to estimate the size of the approximation error for equation 10. In the appendix, we show that this depends on the size of $\frac{\text{Var}(\mu_i)}{(\bar{\mu}^p)^2}$. We can use the variance in payout rates across the deciles illustrated in Figure 1 to estimate a typical value for Var (μ_i) and use the expected payout from the overround formula to estimate a typical value for $\bar{\mu}^p$. For the soccer data, the average expected payout from the overround formula is 0.935 and the standard deviation in estimated payout rates across bets illustrated in Figure 1 is 0.039. So we can estimate the variance-related term as

$$\frac{\text{Var}\,(\mu_i)}{(\bar{\mu}^p)^2} = \left(\frac{0.039}{0.935}\right)^2 = 0.0017 \tag{12}$$

From equation A.18 in the appendix, this means the expected payout rate implied by the overround formula will be less than 0.2% below the probability weighted sum of the probability-specific payout rates.[9] The variance in payout rates implied by the tennis data is larger but still implies the overround formula will be less than 0.35% below the probability weighted sum of the various markups.

These small approximation errors mean average payout rates estimated by the overround formula will be well approximated by a weighted average of expected payout rates, where the weights are the probabilities P_i of the bets being successful. As explained above, this means average payout rates across all bets will tend to be lower than predicted by the overround formula. Table 2 (for soccer) and Table 3 (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 7.5%, 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 2 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 one for the soccer data) 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.

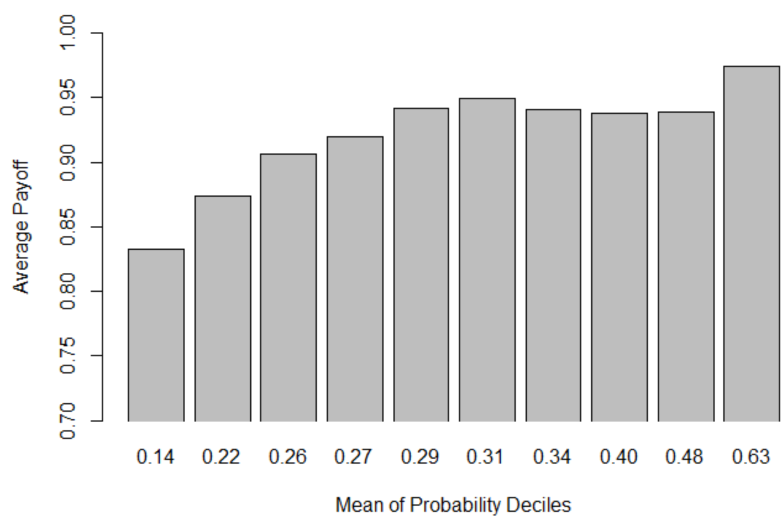
[9] This calculation is 0.0018 if we used the actual average ex post payout rate of 0.922.

Table 1: 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

Figure 1: Average payout rates for bets by deciles of estimated values of the probability the bet will win

(a) Soccer



(b) Tennis

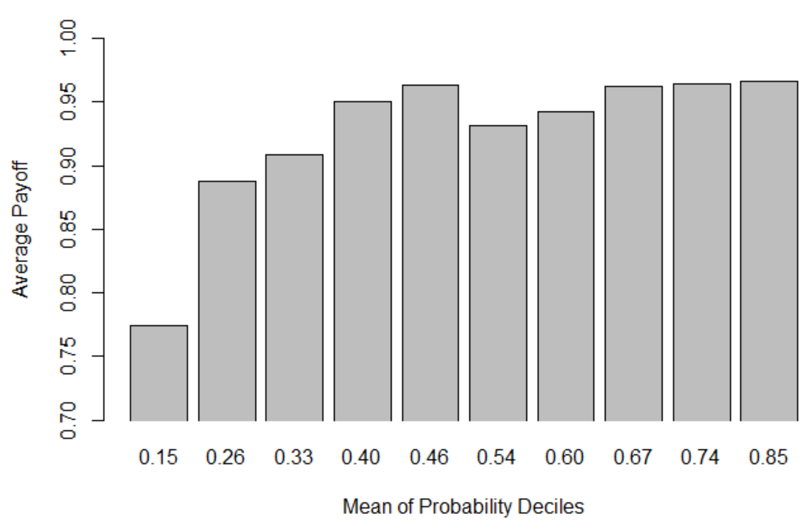


Table 2: Average loss rates across all available soccer bets compared with loss rates implied by overround formula
(N = number of matches)

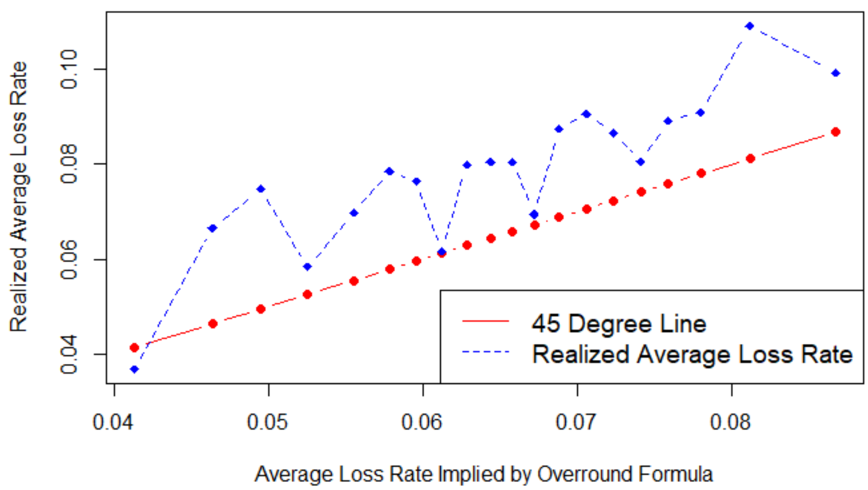
Season	Loss Rates Implied by Overround Formula	Realized Average Loss Rates	N
All Seasons	6.5%	7.8%	84,230
2011 / 2012	7.5%	9.2%	7,694
2012 / 2013	7.0%	7.7%	7,705
2013 / 2014	6.9%	8.6%	7,616
2014 / 2015	6.6%	8.1%	7,841
2015 / 2016	6.6%	7.7%	7,801
2016 / 2017	6.6%	8.1%	7,841
2017 / 2018	6.4%	8.5%	7,794
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 3: Average loss rates across all available tennis bets compared with loss rates implied by overround formula
(N = number of matches)

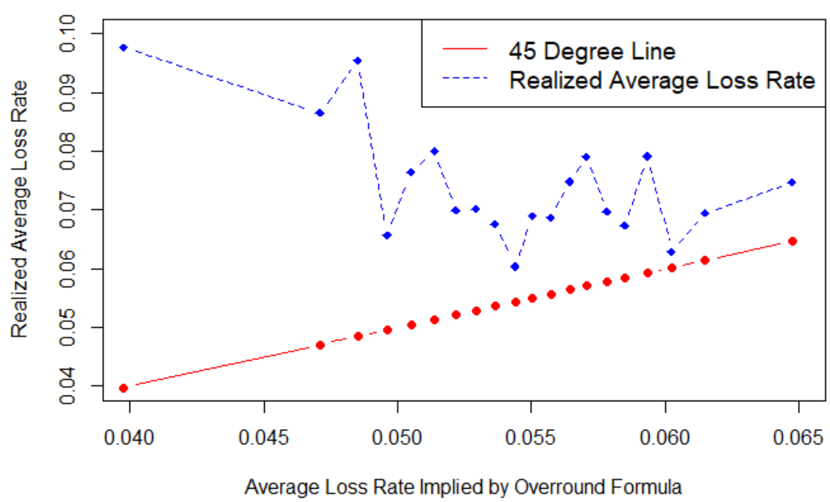
Year	Loss Rates Implied by Overround Formula	Realized Average Loss Rates	N
All Years	5.4%	7.4%	58,112
2011	6.0%	9.5%	5,124
2012	5.8%	8.4%	5,011
2013	5.7%	8.5%	5,066
2014	5.6%	7.5%	5,071
2015	5.7%	8.3%	5,145
2016	5.5%	6.8%	5,141
2017	5.3%	5.8%	5,127
2018	4.9%	6.8%	5,104
2019	5.0%	6.9%	5,080
2020	5.0%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

Figure 2: 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



5. Conclusions

Betting on sports is growing rapidly around the world, most notably in the United States. 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] Forrest, David and Ian Mchale (2007) "Anyone for tennis (betting)?" *The European Journal of Finance*, Volume 13, pages 751-768.

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

[4] Moscowitz, Tobias and Kaushik Vasudevan. (2022). Betting without beta. Yale University Working paper.

[5] 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.

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

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

A Approximation Result

We obtain the approximation described in equation 10 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

$$F(\mu_i) \approx F(\bar{\mu}^p) + F'(\bar{\mu}^p)(\mu_i - \bar{\mu}^p) + \frac{F''(\bar{\mu}^p)(\mu_i - \bar{\mu}^p)^2}{2} \tag{A.13}$$

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

$$\frac{1}{\pi} = \sum_{i=1}^{N} \frac{P_i}{\mu_i} \tag{A.14}$$

Applying the Taylor series approximation in equation A.13 to $F(x) = \frac{1}{x}$ around the point $\bar{\mu}^p$, we get

$$\frac{1}{\pi} \approx \frac{1}{\bar{\mu}^p} - \frac{\mu_i - \bar{\mu}^p}{(\bar{\mu}^p)^2} + \frac{(\mu_i - \bar{\mu}^p)^2}{(\bar{\mu}^p)^3} \tag{A.15}$$

Taking expectations using the P_i terms as probabilities, the middle term on the right equals zero and we get

$$\frac{1}{\pi} \approx \frac{1}{\bar{\mu}^p} + \sum_{i=1}^{N} \frac{P_i (\mu_i - \bar{\mu}^p)^2}{(\bar{\mu}^p)^3} \tag{A.16}$$

The inequality $\frac{1}{\hat{\mu}} < \frac{1}{\bar{\mu}^p}$ 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

$$\frac{1}{\pi} \approx \frac{1}{\bar{\mu}^p} + \frac{\text{Var}(\mu_i)}{(\bar{\mu}^p)^3} \tag{A.17}$$

where $\text{Var}(\mu_i)$ is the variance of payout rates. Re-writing this as

$$\pi \approx \frac{\bar{\mu}^p}{1 + \frac{\text{Var}(\mu_i)}{(\bar{\mu}^p)^2}} \tag{A.18}$$

we can see that the overround-based estimated payout rate $\hat{\mu}$ will be smaller than $\bar{\mu}^p$. But if term $\frac{\text{Var}(\mu_i)}{(\bar{\mu}^p)^2}$—the variance in the bookmaker's profit margins across bets divided by the square of the probability weighted average of the payouts—is sufficiently small, then we can write this approximation as

$$\pi \approx \bar{\mu}^p \tag{A.19}$$

B Bias in Normalized Probabilities

Here we show that when the bookmaker sets odds to have a favorite-longshot bias, the traditional normalized probabilities are biased estimates of the true probabilities. When there is a favorite-longshot bias, normalized probabilities can be expressed as

$$\hat{P}_i = \frac{\pi}{O_i} = \frac{\pi}{\frac{\mu_i}{P_i}} = \frac{\pi}{\mu_i} P_i = \left(\frac{1}{\mu_i \sum_{j=1}^{N} \frac{P_j}{\mu_j}} \right) P_i \tag{B.20}$$

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

$$\mu_i \sum_{j=1}^{N} \frac{P_j}{\mu_j} = P_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} P_j \frac{\mu_i}{\mu_j} \tag{B.21}$$

We can now examine the implications of favorite-longshot bias for this calculation. It calculates a 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 B.21 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.