DISCUSSION PAPER SERIES

DP17948

CALCULATING THE BOOKMAKER'S MARGIN: WHY BETS LOSE MORE ON AVERAGE THAN YOU ARE WARNED

Tadgh Hegarty and Karl Whelan

ASSET PRICING

ISSN 0265-8003

CALCULATING THE BOOKMAKER'S MARGIN: WHY BETS LOSE MORE ON AVERAGE THAN YOU ARE WARNED

Tadgh Hegarty and Karl Whelan

Discussion Paper DP17948
Published 02 March 2023
Submitted 26 February 2023

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

CALCULATING THE BOOKMAKER'S MARGIN: WHY BETS LOSE MORE ON AVERAGE THAN YOU ARE WARNED

Abstract

If betting markets are efficient, then the expected loss rate on all bets on a game can be calculated from the quoted odds. Guides to sports betting tell bettors how to do this calculation of the predicted average loss rate. 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 bets will be higher than predicted by this widely-recommended calculation. We provide evidence from betting on soccer and tennis to illustrate that average realized loss rates on bets are consistently higher than predicted by the conventional calculation.

JEL Classification: G14, L83, Z20

Keywords: N/A

Tadgh Hegarty - tadgh.hegarty@ucdconnect.ie
University College Dublin

Karl Whelan - karl.whelan@ucd.ie
University College Dublin and CEPR

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

Tadgh Hegarty* Karl Whelan†

University College Dublin

February 2023

Abstract

If betting markets are efficient, then the expected loss rate on all bets on a game can be calculated from the quoted odds. Guides to sports betting tell bettors how to do this calculation of the predicted average loss rate. 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 bets will be higher than predicted by this widely-recommended calculation. We provide evidence from betting on soccer and tennis to illustrate that average realized loss rates on bets are consistently higher than predicted by the conventional calculation.

Keywords: Market Efficiency, Sports Betting, 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 2022, thirty states had legalized sports betting in various forms. Betting in these markets is already large, with $186 billion placed in legal sports betting markets in the time between the Supreme Court ruling and the end of 2022.[1] The huge amount of money being spent on advertising on sports betting in the US suggests that the newly-legalized bookmaking firms believe this market is going to grow substantially over the next few years.[2]

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 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 game, 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 juice or the hold—and conventions on how to quote odds 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.[3] This calculation tells bettors how much they will lose on average unless they have a better judgement than the bookmakers on the probabilities of potential outcomes.

The overround formula is based on the assumption that the 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. In this paper, we show that if bookmakers have higher profit margins for bets that are less likely to win, then the average loss rate across all bets will be higher than implied by the overround formula. We illustrate this result with examples from betting on soccer and tennis.

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

[2] Forbes have reported that total advertising spending by US sportsbooks was likely around $1.8 billion in 2022 and is projected to grow to $2.9 billion by 2024. https://www.forbes.com/sites/bradadgate/2022/09/15/sports-betting-is-revving-up-ad-spending-for-fourth-quarter/

[3] Here is a good example www.thesportsgeek.com/blog/calculating-margins-in-sports-betting/

2. How Bettors Calculate The Bookmaker's Margin

Consider a sporting event with N possible outcomes, each with probability P_i. An efficient betting market will have the property that the expected return to betting on each outcome will be the same. Market efficiency would imply that discrepancies in expected returns would be eroded by bettors switching away from bets with bad expected returns to those with better returns and by bookmakers under-cutting those who offered poor odds on certain bets.

Bookmakers earn profits and have to cover costs, so the common expected payout on a $1 bet, which we denote as μ, must be less than one. The decimal odds O_i, the total payout from betting $1 on outcome i when this outcome occurs, must satisfy

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

Combined with the condition that the probabilities sum to one, this provides $N + 1$ linear equations which solve to give a unique set of $N + 1$ unknown values, namely the N probabilities and an expected return μ. Specifically, μ is given by

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

and the so-called normalized probabilities can then be calculated directly from equation 1. The expected payout is determined by the sum of the inverses of the odds, i.e. the overround. As a shorthand, we will call equation 2 the overround formula.

Almost every resource aimed at informing people about sports betting places a key emphasis on using the overround to estimate the bookmaker's margin.[4] Bettors are told to use this calculation to screen out bets that are bad value and to see which bookmakers are offering the most attractive odds across the range of options. While the calculation is relatively easy to do, bettors can also use one of the internet's many betting margin calculators, which compute the margin for you once you have typed in the odds. Most of these sites provide links to online bookmakers.

Importantly, while bettors hope they are better at assessing the likely outcome of games than bookmakers, they can use the overround formula to calculate how much they can expect to lose on average on bets if bookmakers are correct in their evaluation of the probabilities. Since it is pretty unlikely that the average bettor is more qualified than bookmakers to assess the likelihood of outcomes, the overround formula provides a baseline for what people can expect the average loss rate to be across all available bets.

[4] In many cases, these guides tell people to calculate the bookmaker's profit margin as $\sum_{i=1}^{N} \frac{1}{O_i} - 1$ rather than use equation 2. This is not technically correct but, in practice, for the range of margins seen in online betting, these two calculations are generally very close. For example, for a bet in which the odds for both teams in a game pay a profit of $100 on bets of $110, the bookmaker's gross profit margin is 4.54% while this alternative calculation gives 4.76%.

3. Favorite-Longshot Bias

The overround formula relies on the assumption that betting markets are perfectly efficient. 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.

We will assume now that odds are determined by the bookmaker according to

$$O_i = \frac{\mu_i}{P_i} \qquad i = 1, ..., N \tag{3}$$

where the payout rates μ_i depend positively on the P_i. Applying the overround formula to estimate the expected payout rate, which we will now denote as $\hat{\mu}$, we get

$$\hat{\mu} = \frac{1}{\sum_{i=1}^{N} \frac{P_i}{\mu_i}} \tag{4}$$

The overround formula estimates the expected payout rate as a probability-weighted harmonic mean of the individual payout rates.

We want to compare $\hat{\mu}$ with the average payout rate across all bets, which can be calculated as a simple average of the separate payout rates. However, the calculation of $\hat{\mu}$ is quite complex. To derive a result that allows for a direct comparison of these two calculations, we will use an approximation for the overround formula. Specifically, we show that for sports betting markets, the probability weighted harmonic mean $\hat{\mu}$ is well approximated by the probability weighted arithmetic mean of the payout rates, which we will denote $\bar{\mu}^p$

$$\hat{\mu} \approx \sum_{i=1}^{N} P_i \mu_i = \bar{\mu}^p \tag{5}$$

In general, $\bar{\mu}^p$ will not equal the average payout rate for bettors across all the bets they have placed. This is because betting volumes in fixed-odds markets do not have to be 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 fixed odds “money line” bets on US sports from Moscowitz and Vasudevan (2022) shows the number of bets placed as being relatively equal across

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

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.

We obtain the approximation described in equation 5 with an approach used to derive Jensen's inequality. Using Taylor series, we can write any function of the individual payouts, μ_i as

$$F(\mu_i) = F(\bar{\mu}^p) + F'(\bar{\mu}^p)(\mu_i - \bar{\mu}^p) + \frac{F''(\mu_i^*)(\mu_i - \bar{\mu}^p)^2}{2} \tag{6}$$

where μ_i^* is a value between μ_i and $\bar{\mu}^p$. The inverse of $\hat{\mu}$ is given by

$$\frac{1}{\hat{\mu}} = \sum_{i=1}^{N} \frac{P_i}{\mu_i} \tag{7}$$

Applying the Taylor series expansion in equation 6 to $F(x) = \frac{1}{x}$ around the point $\bar{\mu}^p$, we get

$$\frac{1}{\mu_i} = \frac{1}{\bar{\mu}^p} - \frac{\mu_i - \bar{\mu}^p}{(\bar{\mu}^p)^2} + \frac{(\mu_i - \bar{\mu}^p)^2}{(\mu_i^*)^3} \tag{8}$$

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

$$\frac{1}{\hat{\mu}} = \frac{1}{\bar{\mu}^p} + \sum_{i=1}^{N} \frac{P_i(\mu_i - \bar{\mu}^p)^2}{(\mu_i^*)^3} \tag{9}$$

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. Approximating the μ_i^* terms with $\bar{\mu}^p$, this can be re-written as

$$\frac{1}{\hat{\mu}} \approx \frac{1}{\bar{\mu}^p} + \frac{\text{Var}(\mu_i)}{(\bar{\mu}^p)^3} \tag{10}$$

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

$$\hat{\mu} \approx \frac{\bar{\mu}^p}{1 + \frac{\text{Var}(\mu_i)}{(\bar{\mu}^p)^2}} \tag{11}$$

we can see that the overround-based estimated payout rate $\hat{\mu}$ will be smaller than $\bar{\mu}^p$. But if the variance in the bookmaker's profit margins across bets is sufficiently small, then we can write this approximation as

$$\hat{\mu} \approx \bar{\mu}^p \tag{12}$$

As we discuss below, the variation of payout rates in actual sports betting markets is small enough for this approximation to work well in practice.

When this approximation holds, we can compare the expected payout implied by the overround formula and the actual average payout rate across bets. Favorite-longshot bias means P_i and μ_i are positively correlated, so we can conclude that

$$\frac{1}{N}\sum_{i=1}^{N}\mu_i < \sum_{i=1}^{N}P_i\mu_i = \bar{\mu}^p \approx \hat{\mu} \tag{13}$$

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.

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 and outcomes on 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 the appendix. From www.tennis-date.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 a contest.

Figure 1 displays two bar charts that divide all bets in our samples into deciles by their predicted probability of success according to the standard approach described above.[6] 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 much smaller datasets of Angelini and de Angelis (2019) for soccer and Forrest and McHale (2007) for tennis.

Second, we can use the distributions in Figure 1 to estimate the size of the variance-related term in equation 11 in each case. For the soccer data, the average expected payout from the overround for-

[6]The presence of favorite-longshot bias means the probability estimates are biased but 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.

mula is 0.935—we can use this as an approximation for $\bar{\mu}^p$—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

$$\frac{\text{Var}\,(\mu_i)}{(\bar{\mu}^p)^2} = \left(\frac{0.039}{0.935}\right)^2 = 0.0017 \tag{14}$$

This means the loss percentage implied by the overround formula will be less than 0.2% below the probability weighted sum of the various markups.[7] This calculation may actually over-estimate the average gap between $\hat{\mu}$ and $\bar{\mu}^p$. Bootstrapped simulations that replicate both the sample of estimated probabilities and the pattern of favorite-longshot bias in the soccer betting data indicate the average gap between $\hat{\mu}$ and $\bar{\mu}^p$ as being 0.0009. The variance in the loss percentages 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 suggest that favorite-longshot bias should imply that the average loss rates across all available bets are higher than predicted by the overround formula. Table 1 (for soccer) and Table 2 (for tennis) confirm this prediction. For soccer, the average expected loss rate predicted by the overround formula is 6.5% while the actual average loss rate for across all bets is 7.8%, so losses are twenty percent higher than predicted. For tennis, the average expected loss rate predicted by the overround formula is 5.4% while the actual average loss rate for 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 expected and actual loss distributions are equal.

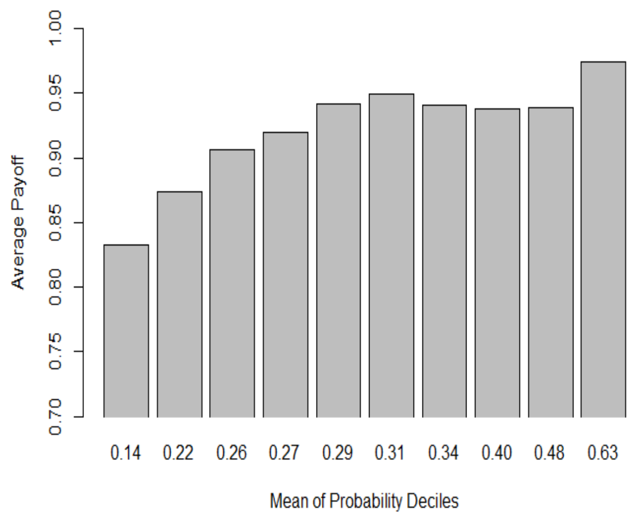
The tables also show this pattern has been relatively stable over time. Both average realized loss rates and the loss rates predicted by the overround formula have fallen over the past decade, perhaps reflecting greater competition in the sports betting market. However, for each year, realized average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

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

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

(a) Soccer



(b) Tennis

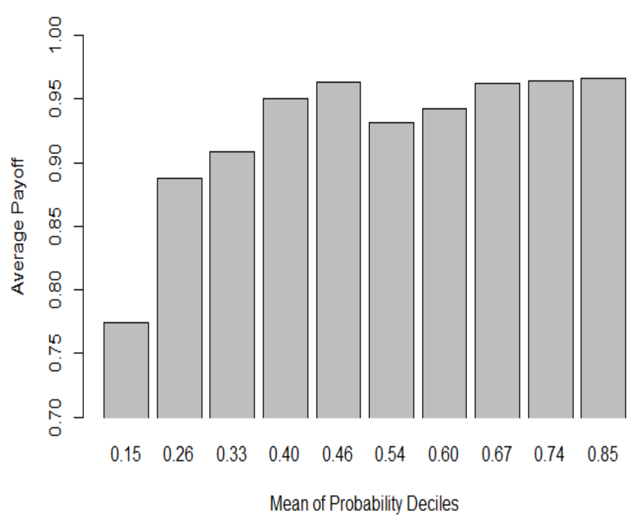


Table 1: Average loss rates on all soccer bets compared with loss rates implied by overround formula (N = number of matches)

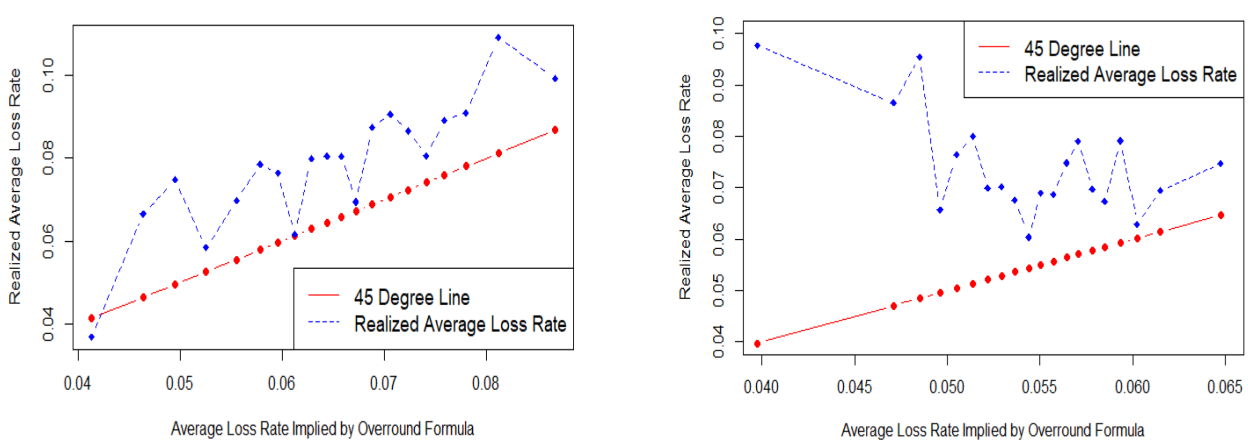
Season	Mean Loss Rates Implied by Overround Formula	Mean Realized Average Loss Rates	N
All Seasons	6.5%	7.8%	84,230
2011 / 2012	7.5%	9.2%	7,694
2012 / 2013	7.0%	7.7%	7,705
2013 / 2014	6.9%	8.6%	7,616
2014 / 2015	6.6%	8.1%	7,841
2015 / 2016	6.6%	7.7%	7,801
2016 / 2017	6.6%	8.1%	7,841
2017 / 2018	6.4%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

Year	Mean Loss Rates Implied by Overround Formula	Mean 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 on all 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—then the overround formula understates the average loss rates incurred across all bets. In our example, actual average loss rates 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.

A Soccer Leagues in the Dataset

The soccer dataset contains the following leagues.

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