

The odds of betting on value: finding true value events in football

Exploring statistics to identify the right events
to wager a true value bet on.

Because value lies not in generous odds.

Marc D. Richter

Data Analyst

Bettors know that the process of value betting means placing a high volume of bets since value betting is a numbers game. To do this value bettors will try to deduct what are true or *fair* odds and whether they can find an edge over the bookmakers.

This is difficult. It requires diligence and knowledge to find the odds to bet on. And: only over time and many bets gains would reliably overtake losses.

I do not even try to assess *fair* odds in this paper. Instead I look at where I can find value to begin with. Because, as we will see, generous odds are more difficult to beat while value is found elsewhere instead.

Basic Concepts

Fair odds (or true probability)

Fair odds imply the actual probability of an outcome. Over time all bets on *fair* odds would return (near) zero profit or loss. All combined probabilities for an event have to add up to 100%. For a coin toss this would mean two equally likely outcomes: Head or Tail, written 50/50 or 1/1. Or in decimal odds: 2.00 for Heads and 2.00 for Tails for every single unit bet. (1/1 meaning just two outcomes, each probability being equally great, 2.00 is the decimal way of writing this same concept. In both cases we win half the time and get paid 2 units for each win so over time we lose as much as we win.)

Overrounding

Bookmakers offering odds that do *not* in fact represent 100% of probability is the most usual practice for bookmakers to make money. This is called Overrounding (or bookie's edge or vig). So the odds bookmakers will offer in reality will likely be 1.95 and 1.94 or similar, lower than *fair* odds (2.00 and 2.00) so over time *they* win and we lose.

Value bets, sharpness and softness

In the case above all odds higher than 2.00 for either event would represent a *value* bet. A *sharp* bookmaker will be as close to the *fair* odds of 2.00, a *soft* bookmaker will move away from that for numerous reasons. Either to attract more bets, avoid bets or - actually make money, even when offering higher odds.

Higher odds

Bookmakers may offer higher odds to offset a mismatch when bets on their book are leaning too far to one side potentially losing them money. They may also try to tickle bettors' biases or typical gamblers' fallacies towards overvaluing current trends or a difficulty to estimate the correct spread for short favorites against long underdogs. And they may do so because the uncertainty inherent in some events allows for it.

Bet types

This paper tests **1X2** or **Three-Way**- bets wagering on the full time result of league games that can end in either a win for the home team (1) a draw with no winner (X) or a win for the away team (2) - with focus only on the Home team wins, so one outcome for each game.

Markets

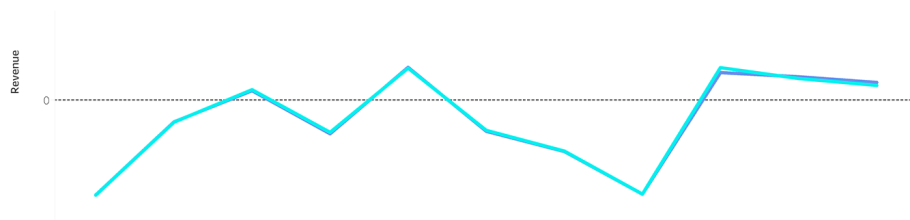
Markets examined here are **European top flight men's football**. This segment consists of five leagues: England's EPL or Premier League, the Spanish La Liga, Italian Serie A, Ligue 1 from France and Bundesliga from Germany). The main focus will be Bundesliga.

Central European bettors who do not use location spoofing via VPN services do not have access to all odds from all countries. Not all odds potentially in the market will be actually available to them.

With Central European odds and only Home Wins logged in, what are the results playing the Median of available odds on all Bundesliga games with single unit bets?

Betting on Median Odds

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Weighted Odds	-52,28	-11,93	5,34	-18,43	18,17	-17,08	-28,22	-51,79	15,32	13,00	9,86
Median Odds	-52,33	-12,06	5,85	-17,68	17,78	-16,60	-28,03	-51,76	18,03	12,05	8,25



The overrounding will creep in and chafe off 3.51% from 3-way bets on Home Wins.

Over time bettors would finish in the negative using real world bookmakers' odds.

Median, Average and "Weighted"

Median odds (so the value that represents the middle of all other values) are different from average odds (the number where all odds' values are added up and divided by the number of odds) - when Median and Average are wider apart it is because there is one or more very high or very low value(s) at the outer end skewing the average away from the middle.

The Average or a value very close to it **might indeed not exist** as it's just a quotient. One or few very high Maxima can distort the Average drastically. So it is important to understand that Averages might be entirely elusive and not have any real world equivalent.

The Median however is the middle value or $\frac{1}{2}$ quantile. It is **very robust** where outliers threaten to skew the data.

"Weighted Odds" is a measure introduced here. It is the arithmetic mean of Median and Average cutting the Maximum from the Median-defining set. If "Weighted" odds' revenue is compared to the revenue with Median odds, "Weighted" hardly differs from Median.

(The closest available real world odds correspond to Weighted odds 99.96% on average - Within single seasons spreads will oscillate between 4.2 and 6.7% total with roughly half that distance from 100%. Outliers are mostly rare, by far the most common spread being +/-0.01.)

Availability of Weighted odds:

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
closest real odds (% of Weighted)	99,93	99,87	99,96	99,97	99,76	100,09	100,02	100,03	99,98	100,06	99,89

It's important to emphasize the point concerning **the purpose of Weighted odds**: actual Maxima are often unavailable due to local blocking, and Averages might have no real-world equivalents. So settling for **attainable odds that reliably actually exist** and are higher than the middle value deliver good odds we can dependably play.

Statistical Claims

“99% of (bettors) lose. And they lose more over time the longer they keep playing.” *

Value betting companies and other services like drop-odds notification services offer pointers as to what *fair* odds might look like and/or where clients might find a higher offer implying value. They **claim annual double digit percentage returns** - the real returns their clients actualize are of course a **black box**. Which odds can be accessed from each bettor's country and a personalized betting strategy may be part of the service or not. All figures on this are either released by the value-bet- services or anecdotal coming from individual bettors. These claims will not be included here.

* (According to: Kenny Alexander, then CEO of Entain (world's 2nd largest gambling conglomerate) in a committee hearing at House of Lords, Feb. 04 2020)

Finding the right metrics

Home wins are the outcome that has the lowest odds overall. Most other bet types can be considered more inherently promising but sticking to lowest-value principle ensures to **identify key indicators for whether a simple metric can identify value bets as well as falsely promising odds**.

Only the top flight leagues will be examined for the following reasons:

- Overrounding is less pronounced in such competitions (see [Patrice Marek's paper](#)).
- Media coverage and comprehensive data tracking lead to high transparency. Truly sharp odds do exist. So baselines for value will be easier to find than for more obscure events.
- Odds for top-flight football are widely available from numerous bookmakers which increases the chance to find value.

Historical threshold

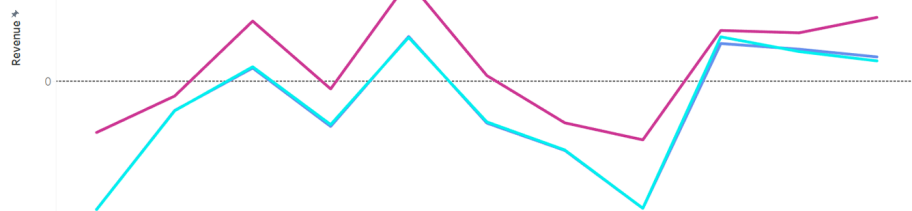
Historic baselines for the probabilities of wins-, draws- or losses are introduced to infer a most-likely outcome from larger trends. This happens either as a score modeling approach or probing for likelihood- for- a- specific- outcome.

Research suggests that **betting by historical metrics** tends to **actually beat the odds** by consistent albeit small margins. (For a hierarchical Bayesian/Poisson model see the work of Leonardo Egidi, Francesco Pauli and Nicola Torelli from their [2018 paper](#))

With a historic score created and applied for each team that has played in the Bundesliga for the past 20 years. (Find the outline of the score below [2]) the results compared to what the previous ones:

Betting with a historic threshold:

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	
Weighted Odds	-52,28	-11,93	5,34	-18,43	18,17	-17,08	-28,22	-51,79	15,32	13,00	9,86	-118,04
Median Odds	-52,33	-12,06	5,85	-17,68	17,78	-16,60	-28,03	-51,76	18,03	12,05	8,25	-116,50
Weighted / Historic Score	-20,93	-6,06	24,44	-3,17	40,61	2,27	-16,99	-23,90	20,63	19,68	25,98	62,56



Using historic scores offsets losses, slightly improves gains and over time produces a slim edge over the odds. This is a small but consistent improvement.

“Great odds”

The threshold of a historic score combined the previously introduced *Weighted* odds will be called “Great Odds”.

Examining a classic value approach: Spreads between Sharp and Soft odds

Value bettors often consult reliably sharp bookmakers (most often *Pinnacle*) for odds. Then look for higher odds available and play them as long as the final payout (often around 95% of the won bet’s value[1]) exceeds *Pinnacle*’s odds multiplied by a single unit bet.

(If Pinnacle set the odds at 2.00, to make upwards of 2,01 in order to beat the tendency to zero out and considering an expected payout of likely 95% one would have to find odds of at least 2.16 .)

The approach faces a few practical issues. It is a **numbers game**. Averaging out becomes a prerequisite and **bettors need to place many bets**. Finding enough qualifying bets is a

problem. Bettors have to either select bets less diligently or may not find enough bets to average out their numbers. In the example above, note the gap between *fair* odds and odds required to reach positive results. Soft bookmakers will mitigate risk by offering **higher odds** because there is **more uncertainty in an event's outcome**.

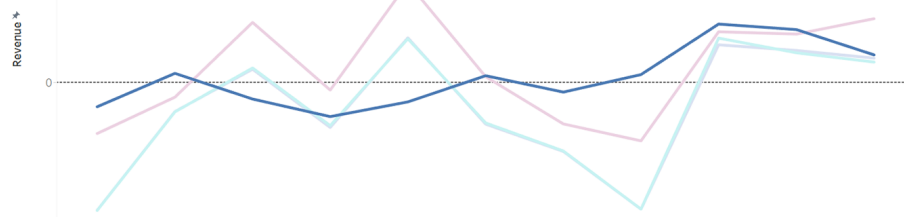
All bets therefore require plausibility checking, slowing down the process. Notification apps or subscription groups for such approaches exist and help identify such odds where soft bookmakers allow enough spread over sharp bookmakers and their odds.

Bets on Big Spreads using Great odds:

To automate the process and only look for attainable odds the prevalent spread between Median and Average over time needs to be taken into account. It moves between 0.21 and 0.29, mostly closer to 0.21. Reflecting the need for higher odds in the marketplace that would skew the average upwards, only odds from such events where Averages were at least 0.3 greater than Median were examined.

Betting on higher odds with Big spreads above Median:

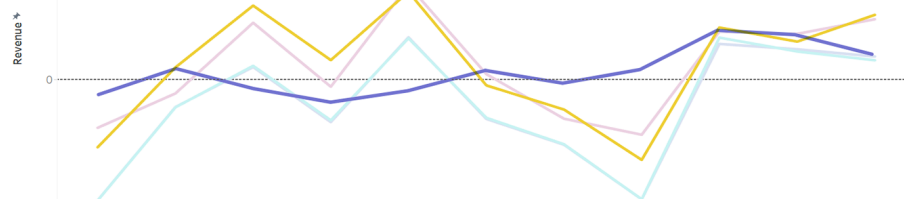
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	
Weighted Odds	-52,28	-11,93	5,34	-18,43	18,17	-17,08	-28,22	-51,79	15,32	13,00	9,86	-118,04
Median Odds	-52,33	-12,06	5,85	-17,68	17,78	-16,60	-28,03	-51,76	18,03	12,05	8,25	-116,50
Weighted / Historic Score	-20,93	-6,06	24,44	-3,17	40,61	2,27	-16,99	-23,90	20,63	19,68	25,98	62,56
Big Spread / Weighted Odds	-10,00	3,65	-6,83	-14,00	-8,00	2,69	-4,00	3,15	23,77	21,52	11,19	23,14



Flattened out major peaks and valleys in the trend leaven only an insubstantial edge. Results from the opposite, a Low-Spread approach with spreads between Median and Average of 0.2 or smaller, and only *Median* odds bet will be less ineffective.

Betting on odds with only small spreads between Average and Median:

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	
Weighted Odds	-52,28	-11,93	5,34	-18,43	18,17	-17,08	-28,22	-51,79	15,32	13,00	9,86	-118,04
Median Odds	-52,33	-12,06	5,85	-17,68	17,78	-16,60	-28,03	-51,76	18,03	12,05	8,25	-116,50
Weighted / Historic Score	-20,93	-6,06	24,44	-3,17	40,61	2,27	-16,99	-23,90	20,63	19,68	25,98	62,56
Big Spread / Weighted Odds	-10,00	3,65	-6,83	-14,00	-8,00	2,69	-4,00	3,15	23,77	21,52	11,19	23,14
Small Spread / Median Odds	-29,33	5,20	31,85	8,32	37,78	-2,60	-13,03	-34,76	22,41	16,30	27,87	70,01

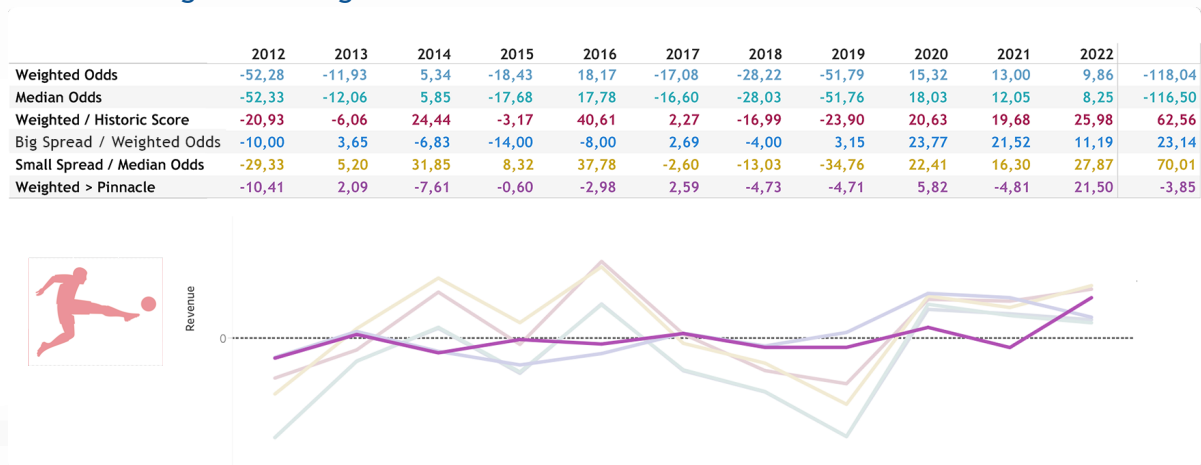


Weighted > Pinnacle's odds

If the **Big-Spread** metric alone seems unproductive one can solve for *Pinnacle's* odds: if Weighted- exceeds *Pinnacle's* odds a bet is placed.

In the Bundesliga there is a significant shortage of sufficient high odds. (This same situation is mirrored when 3 of the other 4 top leagues allow for only very few Big-Spread- but many more Weighted>*Pinnacle* bets.) Most *soft* bookmakers operating in Germany avoid topping the *sharp* ones.

Betting on odds higher than *Pinnacle*:



Higher odds are not inherently more productive. Gains and losses regress to the mean over time, flattening towards zero.

Interim Conclusion: higher odds underperform

Higher odds do not in fact mean value. The higher odds for events with relatively high spreads between Median and Average reflect overall more uncertainty and thus more risky bets. Bookmakers show overall sensible assessment of uncertainty when offering higher odds for higher risk.

Also another phenomenon has to be mentioned here: the **favorite-longshot bias** and its effect: bookmakers' margins or overrounding often gets disproportionately attached to the longshot side while the odds on favorites are much more accurately in line with outcome prediction.

This, along with the higher odds' values to begin with, allows for greater swings in odds over time and keeps the true implied probabilities difficult to deduct.

Betting favorites might therefore lose you less and generous odds may be harder to beat actually.

Over time the stats-lines for higher-odds keep strongly to the mean (zero gains/losses). This indicates that higher odds on riskier bets to look for value only works to level the

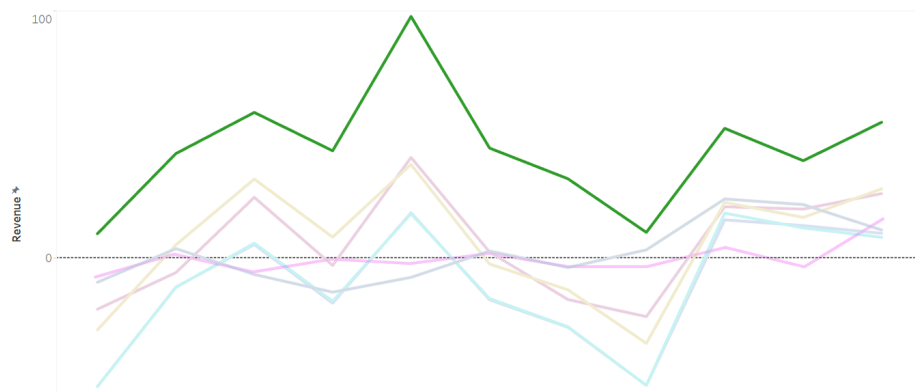
previously inherent favorite-longshot bias thereby **underperforming** simple and prudent metrics that can be easily calculated long before the event.

Examining Small Spreads using Great odds

Using a combination of previously split subsets creates an intersection of two different groups. This can undo the effectiveness of previous sorting or in fact heighten the information gain. If the two simple metrics with gains of note we bet previously - historically validated (“Great Odds”) and 0.20 or smaller spread between odds’ Median and Average (“Small Spread”) are combined - results look like below

Betting on weighted odds with small spreads validated by historic scores:

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	
Weighted Odds	-52,28	-11,93	5,34	-18,43	18,17	-17,08	-28,22	-51,79	15,32	13,00	9,86	-118,04
Median Odds	-52,33	-12,06	5,85	-17,68	17,78	-16,60	-28,03	-51,76	18,03	12,05	8,25	-116,50
Weighted / Historic Score	-20,93	-6,06	24,44	-3,17	40,61	2,27	-16,99	-23,90	20,63	19,68	25,98	62,56
Big Spread / Weighted Odds	-10,00	3,65	-6,83	-14,00	-8,00	2,69	-4,00	3,15	23,77	21,52	11,19	23,14
Small Spread / Median Odds	-29,33	5,20	31,85	8,32	37,78	-2,60	-13,03	-34,76	22,41	16,30	27,87	70,01
Weighted > Pinnacle	-10,41	2,09	-7,61	-0,60	-2,98	2,59	-4,73	-4,71	5,82	-4,81	21,50	-3,85
Small Spread / Great Odds	9,67	42,20	58,85	43,32	97,78	44,40	31,97	10,24	52,41	39,30	54,87	485,01



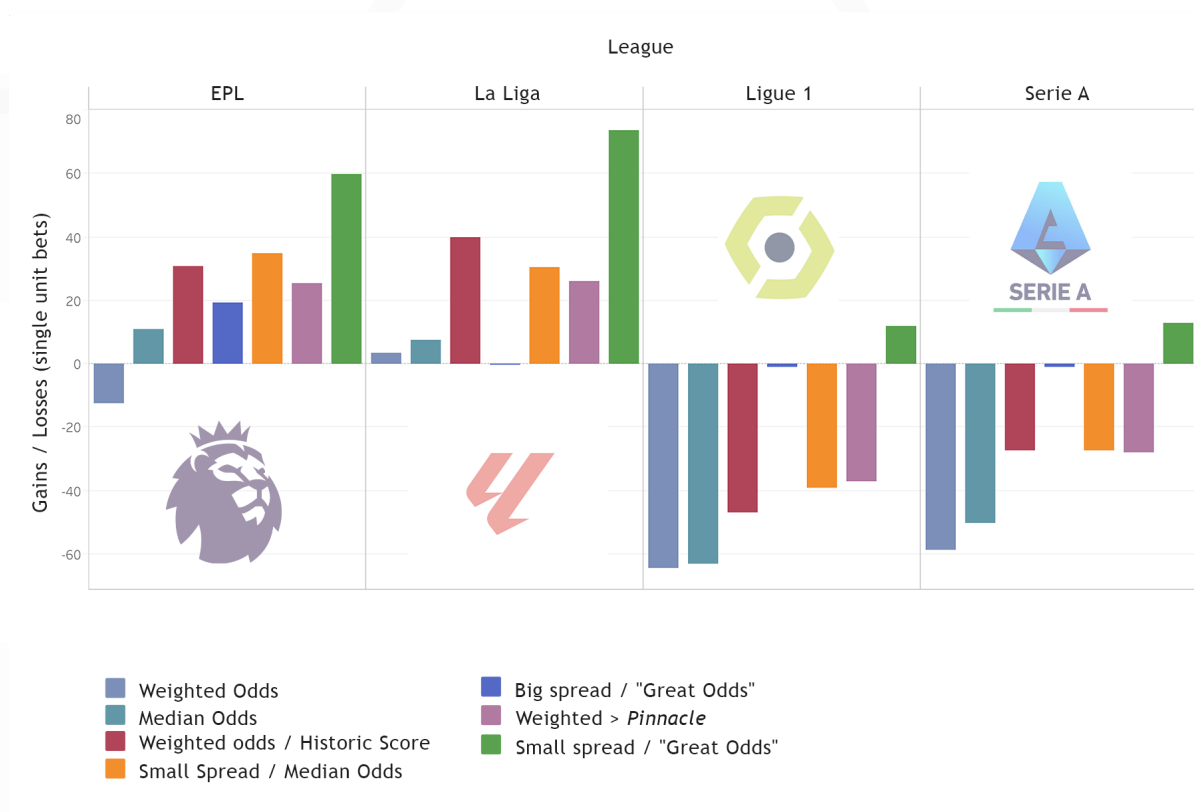
With more bets and a combination of Small Spread and Great Odds we **outperform** other metrics both over time and each season and remain profitable without a single negative season. Profits total at about 17.8% at a rough average of 220 bets per season.

Checking trends in other leagues

Does the overall pattern repeat in the other relevant markets?

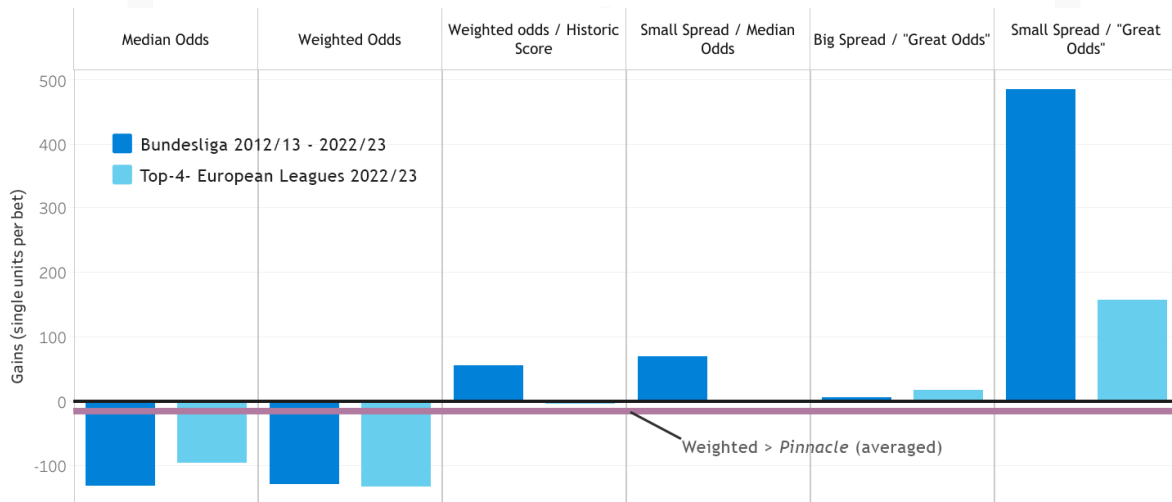
If small spread - great odds are tested in other european top leagues:

	EPL	La_Liga	Ligue_1	Serie_A
Weighted Odds	-12,31	3,27	-64,31	-58,65
Median Odds	10,69	7,51	-63,19	-50,27
Weighted / Historic Score	30,61	39,72	-46,88	-27,24
Big spread / "Great Odds"	19,45	0,00	-1,00	-1,00
Small Spread / Median Odds	34,69	30,51	-39,19	-27,27
Weighted > Pinnacle	25,30	26,07	-37,23	-27,85
SMALL spread / "Great Odds"	59,69	73,51	11,82	12,74



Small Spread / Great Odds outperforms all other metrics and remains consistently profitable over the four probed leagues (full season 2022/23).

Bundesliga profits/losses over time with the probed top leagues by metric:



Conclusion

Odds that are merely higher than a given reference or highest from a line up of odds do not in fact yield higher profits. Bookmakers increase odds only if uncertainty has increased or as bait with previously high longshot-margins. So bettors' gains from longshot bets with increased odds would regress to the mean over time with near-zero profits.

With sharp bookmakers the adherence to outcome probability for odds will make sure they keep a *fair* book. Soft bookmakers will only move upwards of those odds if uncertainty is high enough to safely assume enough bettors will lose out on the more risky bets and there seems enough headroom built in in the overrounding.

Value is instead found where a solid unified theory of an event's outcome with little uncertainty is represented by narrow margins between odds and where large skewing Maxima are absent. Here an upward digression is meaningful and the higher odds therefore valuable. Finding such events is key to uncover true value bets.

A metric that combines solid if broad inference for outcomes with the analysis of odds and their spreads provides a useful baseline for profitable odds.

References

odds and results stats: football-data.co.uk [latest access Oct. 2nd 2023]

additional odds' stats: oddsportal.com [latest access Sep. 30th 2023]

2012-2018 Maximum and Average odds data by betbrain.com [latest access May 31st 2019]

team rankings (full season / in-season): kicker.de [latest access Oct. 6th 2023]

[All LINKS removed]

All sources combined, cleaned, formatted by Marc D Richter. All source- and edited files used in this paper can be downloaded under this LINK (Please acknowledge football-data.co.uk for any odds' or match results-based info).

Repository: https://github.com/richteffektGit/DA_FootballData.git

On bookmakers' efficiency: Marek, Patrice "Bookmakers' Efficiency in English Football Leagues" (2018) 36th International Conference Mathematical Methods in Economics at Jindřichův Hradec; Pilsen

Historic model: Leonardo Egidi, Francesco Pauli and Nicola Torelli "Combining historical data and bookmakers' odds in modelling football scores" (2018) Padova/Trieste

Favorite-Longshot Bias: Joseph Buchdahl - The Favourite-Longshot Bias Revisited (2015)
Football-Data Blog (www.football-data.co.uk/blog/favourite_longshot_bias_revisited.php)

[1] On betting taxes: in most given countries or states betting/gambling legislation is likely to subject sports bets to taxation at a given rate, often a percentage of the bet wagered. Bets lost will usually be covered by the bookmaker keeping the money; won bets' tax burdens will either be shared by bookmakers and bettors or paid in full by the bettor by reducing payout from gains.

recommended for further viewing:

Documentary on bookmakers by director Linda Bendali, written by Laëtitia Giroux
"Paris sportifs, les bookmakers raflent la mise" (or "Sportwetten - Das Milliardenbusiness"), Magneto Presse, ARTE France, CNC (2023) - Available until Dec 22nd 2023 on streaming service Arte.tv

Methodology

[2] This paper's Historic Score:

- the base number of points per active season in the Bundesliga is 33 points, with 1 single point added for each rank that is higher, topped at a maximum 50 points for a season's champion team
- a second tier (2. Bundesliga) tenure followed by promotion to top flight (Bundesliga) will be worth 9 points
- all other non-Bundesliga performances will be worth just 1 point per season
- all teams start out no lower than 9 points to begin with
- ongoing and examined seasons' results will be taken into account, historic rankings will be weighted against current table positions with two incremental steps happening after 10 games played and after half a season (17 games played), upgrading current season's rankings weights by 10% increment each time
- the scores get feature-scaled to represent a multiplier between 1 and 2
- the multiplier is then applied to determine the threshold of probability for the outcome (Home team to win) by calculating the implied probability of a home win based on the underlying historic score of the home team and away team against one another
- First match days of the seasons do not get considered

Bookmakers and odds aggregated by football-data.co.uk

Odds were taken synchronously the afternoon before the day of the event from the bookmakers' website.

Odds Averages and Maxima collected by betbrains.com (2012-2018/19) and football.data.co.uk (2019/20-present)

Bookmakers available historical stats may vary season to season:

Number of seasons per Bookmaker:

11 seasons: Bet365, BWin, Interwetten, Pinnacle, William Hill, VC Bet;

6 seasons: Ladbrokes; 3 seasons Stan James; 1 season; Gamebookers, Blue Square

Marc D. Richter has bet on sports events for roughly 25 years and learned to navigate his cognitive dissonance using data and diligent number crunch. He explores statistical concepts and lateral thinking to refine betting strategies. He could not bet on horses if his life depended on it though. Marc is an outspoken critic of commercial bookmakers and services in the sports betting industry.