Wharton SPORTS ANALYTICS Student Research Journal

2023, Spring Edition

wsb.wharton.upenn.edu/student-research-journal

An Investigation of Sports Betting Selection and Sizing

John Beggy*, W'23*
The Wharton School of the University of Pennsylvania, Philadelphia PA, USA

Daniel Kim*, C'23*
University of Pennsylvania, Philadelphia PA, USA

Kejdi Mucaj*, W'23*
The Wharton School of the University of Pennsylvania, Philadelphia PA, USA

James Nordell*, W'23*
The Wharton School of the University of Pennsylvania, Philadelphia PA, USA

Advisor: Abraham Wyner, PhD
The Wharton School of the University of Pennsylvania, Philadelphia PA, USA

Abstract

Across the many sports betting book, there is variability in the betting lines. Using the average implied odds as a proxy for actual probability, strategic betters are able to identify mispriced lines. These mispriced lines can yield positive expected value opportunities. This work explores bet sizing strategies. The commonly used full Kelly and constant bet size strategies are not effective in practice. Full Kelly scenarios was observed to not work in a realistic betting environment; it led to bankruptcy in 100% of the scenarios. The study recommends partial Kelly with coefficient 0.50 and a conservative 10% threshold as the most profitable strategy, limiting variance and avoiding bankrupting bets while still taking advantage of positive ROI opportunities. The results showed that this strategy could generate an annual rate of return of about 80% over 11 years of betting on the exact matches in the dataset.

key words: betting, Kelly criteria

Introduction

For our previous project, we investigated the availability of positive ROI opportunities in the sports betting industry. We looked through the paper “Beating the bookies with their own numbers - and how the online sports betting market is rigged” and were able to come to a similar conclusion that the authors did. There are dozens of sportsbooks available that offer differing odds, and as a result some are more favorable than others. By using the average implied odds as a proxy for actual probability, strategic betters are able to identify mispriced lines that, in many cases, yield positive expected value opportunities.

With these opportunities available, there are certainly ways to generate positive returns over a period of time. Given this information and conclusion, we wanted to further investigate how various betting strategies would fare in the long run in order to determine how best to select and size bets. The researchers whose work we attempted to recreate used a rather primitive betting strategy of placing a $50 wager on each bet that they identified as profitable. However, there are other methods that we believe would be much stronger at generating returns in the long run while still accounting for risk of ruin and protecting from crippling downswings.

Data Cleaning and Bet Selection

Data Cleaning

We used the same data set as used in our previous reproduction. Although the data itself contains everything needed, further inspection shows that the dataset is perhaps too broad, resulting in a fairly large number of abnormalities that complicate the results of our betting strategies. The dataset included over 100 international leagues, ranging from the English Premier League to the Yemeni League. Some of the most abnormal observations were data from extremely obscure leagues. These match-ups often featured tremendous skill gaps between teams, and odds were typically offered by only one or two sportsbooks. These obscure leagues have exceedingly low betting volume and a dearth of advanced data for modeling odds, which undermines our assumption that average odds are a good proxy for the true probabilities of match outcomes. This introduces an unmanageable amount of uncertainty to our betting strategies, and coupled with the logistical difficulty of placing bets on these markets in the real world, presents a strong point of concern in the dataset. To mitigate this issue, we decided to limit the dataset to the top five European soccer leagues: England, France, Germany, Italy, and Spain. We are confident that matches in these leagues are sufficiently fairly officiated, highly scrutinized, and well approximated by average odds that are offered from a plethora of Europe's most reputable sportsbooks and oddsmakers.

We were also compelled to consider what to do with the abundant arbitrage opportunities present in the dataset. Arbitrage opportunities are indicated by maximum odds that differ significantly from the average odds for that event. This is typically indicative of sportsbooks that have encountered an error in posting or forming odds, or are gravely late to update odds following critical news, such as injury updates. We identified these occurrences by summing the implied probabilities of the maximum odds for each match; any matches whose cumulative implied probabilities summed to less than 1.0 must have mispriced odds. We decided to filter out matches whose probabilities were under 0.80, under the assumption that some of these matches may have been improperly scraped or have mispriced odds that are nearly impossible to capitalize on, as they are corrected almost immediately. We chose to retain the other arbitrage opportunities in the interest of practicality and realism. The average bettor is far more likely to capture moderately mispriced odds before they are corrected. Additionally, sports betting arbitrage is indeed a viable betting strategy that has paid dividends for the tech- and data-savvy bettor. To remove these would be ignorant of the reality of smart betting.

Finally, we also removed matches that featured decimal odds above 4.0 (25% implied probability of any particular event). This was less theory-driven but moreso a decision informed by empirical results. Of course, it is still possible for large underdogs to offer positive expected value so long as the odds are favorable. Still, we found that inclusion of such large underdogs corresponded with markedly increased bankroll variance, which typically resulted in ruin far more quickly. As such, we limited maximum odds to 4.0 in the interest of lowering variance without excessively trimming down good-value bets.

Bet Selection

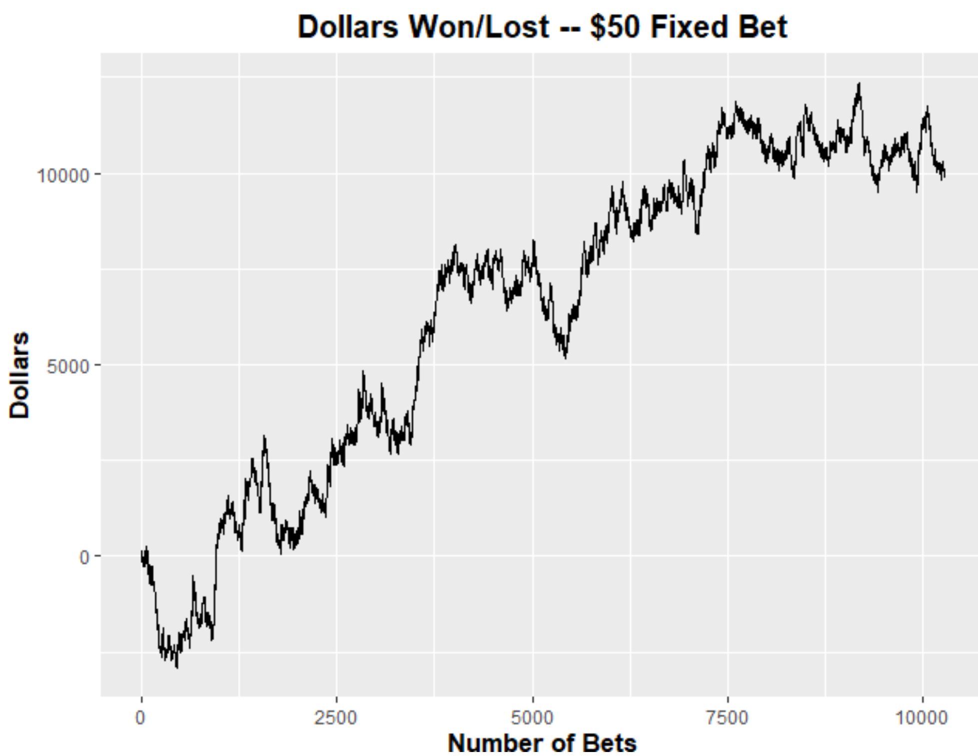
Before sizing could be taken into account, the positive expected value bets had to be identified. Our criteria for bet selection was based on a standard vig of 10% juice total, or the implied probabilities of the adding up to 110% rather than a fair 100%, as is generally industry standard. Therefore, each betting line should contain about 3.3% juice (given that soccer betting lines are three-way). Utilizing a minimum threshold of a 0.05 expected return threshold (which we refer to as bet “margin”) yielded profitable results which is what we opted to use for our selected betting methods that contained a fixed threshold. As previously mentioned, we would take an average of all of the lines across the various sportsbooks to use as a proxy for the true fair odds (with vig still in them) and would compare individual lines to this average in order to determine whether or not there was the minimum required value. Filtering based on these criteria resulted in 121,507 selected lines, of which 10,275 met the 0.05 threshold.

Bet Sizing

With this framework in place to choose our bets, we attempted to model the performance of different sizing strategies: constant sizing, Full Kelly, Partial Kelly, and varying EV margin thresholds.

Constant Bet Size

We used a constant bet sizing strategy as a benchmark, as completed in the prior project. We utilized the 0.05 threshold in order to identify the bets we wanted to place and then wagered $50 on each of them. The strategy was tremendously successful, as it yielded $10,000 profit when applied to the 10,275 placed bets out of the 121,507 total matches in our pool.

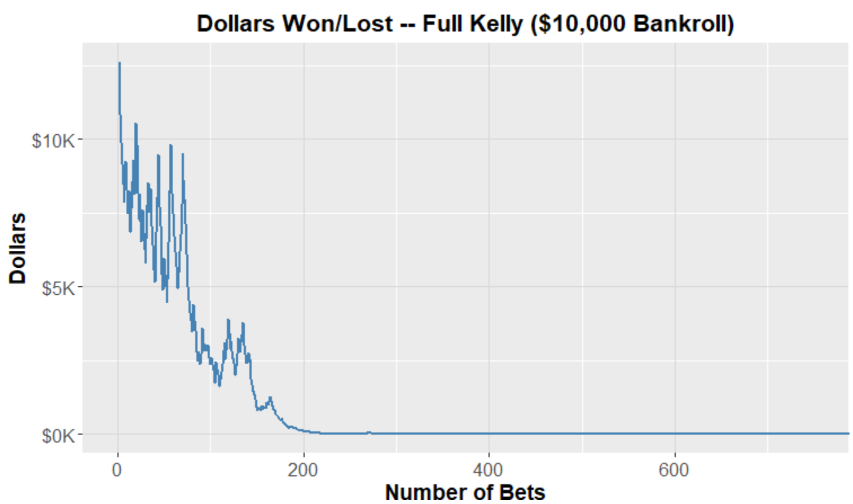


Full Kelly (Margin = 0.05)

The first attempt at a strategic bet sizing strategy was the use of Full Kelly, which maximizes the expected logarithm of the returns. It is a very strong system but there can be significant downside risk as it is not built to consider variance thresholds and does not account for the uncertainty surrounding the implied probability of the event, which can never be fully known.

$$f^* = \frac{bp - q}{b} = \frac{p(b+1) - 1}{b}$$

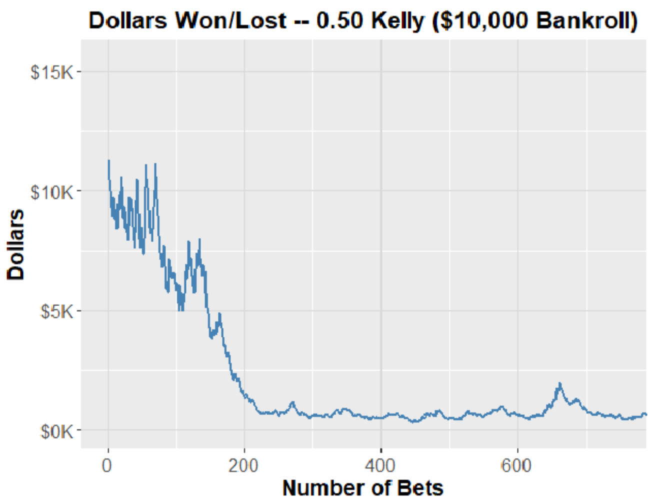
Starting with a bankroll of $10,000, we employed this strategy that calculates bet sizing based on the formula above where f* is the fraction of the current bankroll to wager, b is the net odds received on the wager (profit if win), p is the probability of winning, and q is the probability of losing (1 − p). Generally it can generate impressive results, but when applied to our data set its fatal flaws were revealed. One bet was deemed tremendously advantageous and Kelly determined a size of 90% of the bankroll, and bad luck and a highly improbable loss completely ruined it. Even when we placed a cap on maximum Kelly size at 20% of bankroll, the bankroll went to zero as Full Kelly is simply too aggressive. Multiple of these instances led to the bankroll asymptotically approaching zero, proving that it is not a viable strategy in practice for bettors with a finite bankroll that cannot be easily replenished (as is the case in all real-world scenarios).

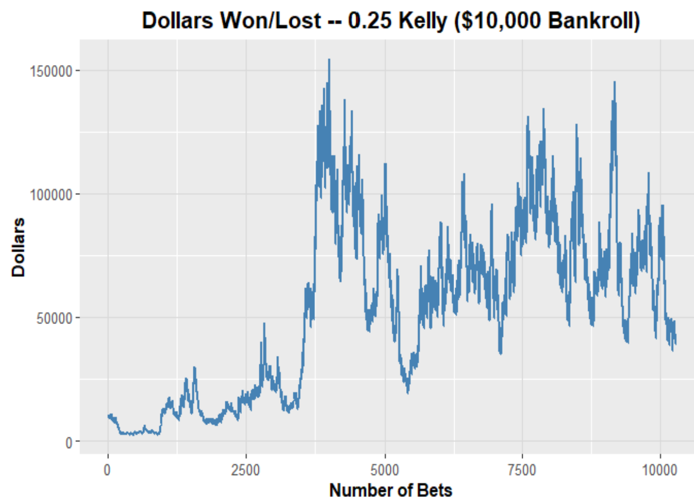


Partial Kelly

A common adjustment made to Kelly is to employ “Partial Kelly”, which aims to reduce variance by scaling down the bet sizes. It is commonly used in practice by bettors and investors because Kelly is far too aggressive to be sustainable for people managing a bankroll in almost all situations. The partial Kelly strategy adjusts for the uncertainty of the underlying probability, which of course can never be fully known in sports betting. Unsurprisingly, this resulted in superior outcomes for our algorithm, although some coefficients fared better than others. Half Kelly still resulted in ruin in every simulation, but the ruin did occur later than in full Kelly.

Quarter Kelly showed far better results, yielding a profit of $40,374 over 10,275 bets. As we continuously reduced the coefficient, the profits grew faster and faster. Partial Kelly works because it accounts for the possibility of oddsmaker error; After all, just because oddsmakers seem to have reached some consensus on an event does not mean the underlying probability has been identified with full confidence. The Full Kelly bettor who believes this without hesitation will surely end in ruin as the inevitable bad run wipes out their bankroll.





Marginal Threshold

In addition to employing different betting strategies, we also experimented with varying the expected value threshold criterion for our bet selection. In the previous section, we operated under a 0.05 margin, meaning the best odds must have been five percentage points better than the average odds for us to bet the market. In exploring the impact of margin size, we simulated each strategy under a more conservative 10% margin and a more aggressive 2.5% margin. The impact was profound; 79,860 bets were placed with a 2.5% margin, falling to 10,275 bets for a 5% margin, and ultimately dropping to only 369 bets when using a 10% margin.

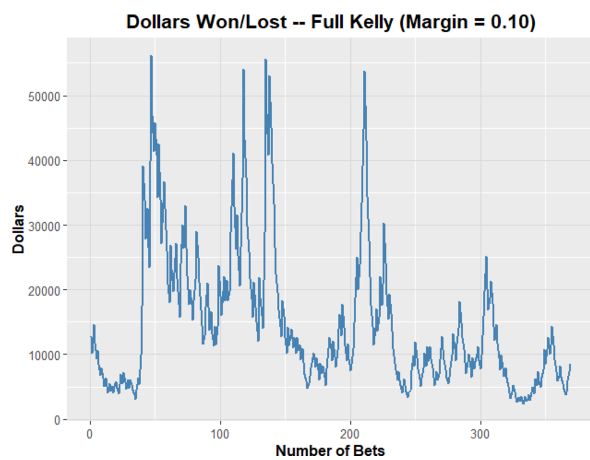
The conservative strategy had the greatest returns. The $50 fixed bets returned less total profit, an unsurprising result due to the low number of bets. The results for Kelly betting were far better. Full Kelly avoided ruin and only resulted in a $1,549 loss, whereas Half Kelly was astoundingly profitable with a $115,097 profit. Quarter Kelly remained profitable with a $62,425 return. The aggressive approach resulted in the most extreme results, where $50 fixed bets resulted in ruin, and all Kelly strategies similarly left us bankrupt. A full summary of our results can be found in the tables below.

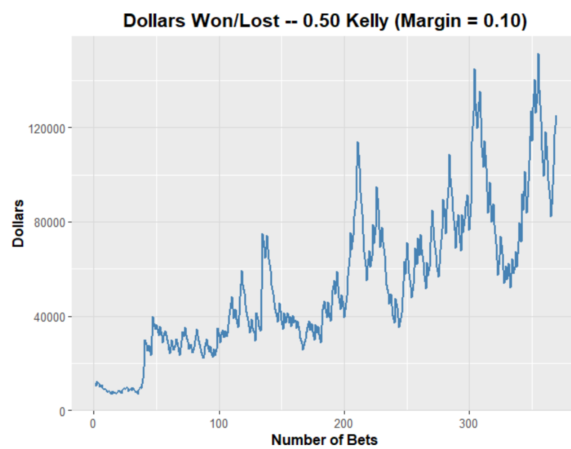
Margin = 0.05 (Baseline Strategy)			
$50 Fixed Bets	**Full Kelly**	**0.50 Kelly**	**0.25 Kelly**
$9,956.50	Bankrupt	Bankrupt	$40,374.00
Number of Bets: 10,275			

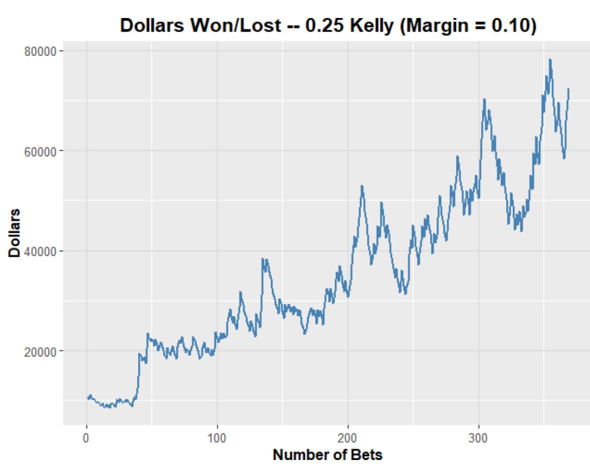
Margin = 0.10 (Conservative Strategy)			
$50 Fixed Bets	**Full Kelly**	**0.50 Kelly**	**0.25 Kelly**
$2,934.50	-$1,549.00	$115,097.00	$62,425.00
Number of Bets: 369			

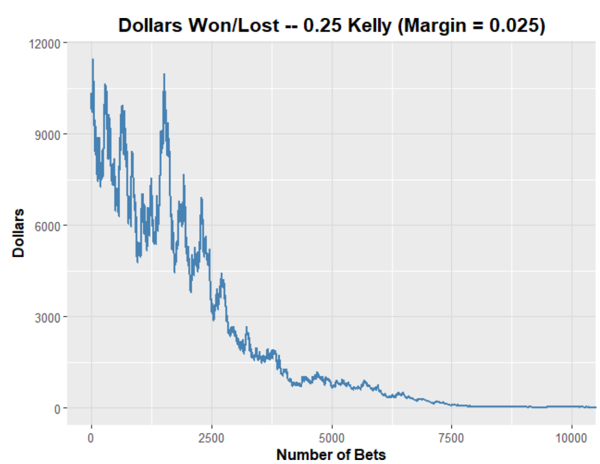
Margin = 0.025 (Aggressive Strategy)			
$50 Fixed Bets	**Full Kelly**	**0.50 Kelly**	**0.25 Kelly**
Bankrupt	Bankrupt	Bankrupt	Bankrupt
Number of Bets: 79,860			

Unsurprisingly, using a conservative bet strategy brings the highest returns but is complemented by a notably low betting pool, whereas using higher margins in 5% and 2.5% provides a much larger pool but comes with an increased risk of diminished returns, as indicated by our poor results in those scenarios. We noted a key relationship throughout these simulations: as the margin of safety increases, we are more confident in our bets and thus a more aggressive Kelly strategy can be employed to maximize returns. This is why Half Kelly is most profitable under a 10% margin, but Quarter Kelly is superior with a 5% margin. This analysis shows again the dangers of using Full Kelly regardless of the threshold as we lost money in all 3 scenarios and points to using a stricter margin, prioritizing quality bets over higher quantity through a larger betting volume.









Conclusions

Full Kelly simply does not work in a realistic betting environment such as the one we set up. It led to bankruptcy in 100% of the scenarios to which it was applied. This is common knowledge in the world of sports betting so we were satisfied to come to the finding using data.

With an aggressive (low) margin, a constant bet size strategy will never be profitable. We found convincing evidence that “vig” not only makes the offered lines mathematically unfair to the bettor but also makes it impossible to profit in practice. With an aggressive bet selection criteria such as the 2.5% threshold, placing $50 bets over and over again will leave you deeply in the red. Conversely, a conservative margin can be very profitable in theory. Our 10% threshold resulted in a profit of $2,934 after placing a total of $18,450 in bets; that’s about a 16% return on investment!

Partial Kelly with coefficient 0.50 and a conservative 10% threshold is the most profitable strategy we found. This strategy limits variance, avoids bet sizes that could bankrupt you, and stays away from games where the book has enough juice to eat your entire margin. Betting on the exact matches in our data set would take you 11 years, with an annual rate of return of about 80%.