Spring 5-1-2024

A Comparative Risk Analysis of Sportsbooks and Options Sellers

Daniel Aybar
daybar@skidmore.edu

<u>A Comparative Risk Analysis of Sportsbooks and Options Sellers</u>

This thesis is submitted in partial fulfillment of the requirements for the course Senior Seminar (EC 375), during the Spring Semester of 2024.

While writing this thesis, I have not witnessed any wrongdoing, nor have I personally violated any conditions of the Skidmore College Honor Code.

Name: Daniel Aybar

Signature: Daniel Aybar

Abstract

Both options sellers and sportsbooks are price-setters in growing speculative markets. With the increased availability of sports betting and options trading, potential inefficiencies in risk-taking behavior could be exposed in these markets. This paper serves to compare the risk taking behavior of sportsbooks and options sellers. Two proposed statistics were used to measure risk in these speculative industries. Risk premium, based on concepts from Arscott (2022) and Moskowitz (2021), measured risk from a per-dollar returns framework. Tail Risk Ratio is a novel addition to previous literature which compares the tail risk of sellers and buyers in speculative markets. The relationship between these two statistics was also explored. The findings of this paper show that options sellers take on significantly more risk than sportsbooks in both an average and volatility-based framework. In addition, it was found that there was no significant relationship between the two proposed risk statistics.

Introduction

Up until the late 20th century, stock options trading was only accessible to a select group mainly consisting of stock brokers, investment bankers, and other finance connoisseurs (Optionstrading, 2022). Modern software platforms have reduced the barriers to entry in this market significantly, giving almost anybody the ability to buy and sell options. The most popular of these platforms is RobinHood, which gained a lot of attention following the GameStop stock frenzy (Thorbeck, 2021). In truth, options trading had been on the rise before the incident, but the increased publicity of trading accessibility caused by the saga truly transformed the options betting market. From 2013 to 2020, the number of options contracts traded increased from 9.4 billion to 21.3 billion; a strong 236% increase in 7 years (FIA, 2023). In 2023, however, the number of options contracts traded ballooned to 54.5 billion - a 256% increase from 2020 and a 579% increase from ten years before in 2013 (FIA, 2023).

The sports betting industry is another market that has grown rapidly over the past few years. The main catalyst for this was the landmark legal case *Murphy vs National Collegiate Athletic Association* which saw the federal legalization of sports betting in 2018 (US Supreme Court, 2018). Since then, 35 states (and Washington DC) have struck down laws pertaining to sports betting, with more soon to follow (Bengel & McCarriston, 2023). With these changes in legislation, it is now estimated that roughly half of the legal population in the US can sports bet (Dixon, 2023). Additionally, roughly $220 billion worth of sports bets were placed in the US from 2018 to 2023, amounting to $132.57 bet per year per capita (Parry, 2023). The market size of sports betting is continually growing, not just through legislation, but also through advertising. In 2021-2023, the sports betting industry spent a combined $4.7 billion on advertisements, $1.9 billion of which were spent during 2023 alone (Garcia, 2024). This has been both a product and a driver of increased industry revenues. In 2023, the gross revenue reported by sportsbooks in the US amounted to $8.76 billion (McClymont, 2024). This is an almost 2,000% increase from the $441 million in gross revenue reported in 2018. This growth is predicted to be sustained well into the future, with the sports betting market having a projected value of $40 billion by 2030 (Dixon, 2023).

Growth in these markets has presented options traders and sportsbooks with the new challenge of maintaining profitability at a large scale. This is especially pertinent for options sellers, who have uncapped potential payouts. While options selling has proven to be profitable, it could be that sellers are taking on excess risk due to the high volatility of these potential payouts. Similarly, sportsbooks that provide high payout options for heavy underdogs could be exposing themselves to unforeseen risks.

There is an extensive amount of literature evaluating risk in the options market. For example, a study by Kappaguntula & Ruddy (2007) estimated a parameter that shows the risk of an option given the premium received by the seller. A comparable study by Fedania & Grammaticos (1992) found that the risk of an option was directly related to the bid-ask spread, which is inherently related to the premium. Another study by Galai & Masulis (1975) looked at the options risk problem through the lens of the CAPM model as well as the Black-Scholes option pricing model to find the risk of equity. More similarly to the tasks of this paper, Pan (2002) identified that higher premiums are charged for options in markets with higher expected volatility. Because of this, the expected returns from high volatility jumps risks were lower than the expected returns due to diffusive risks, implying that the options selling market charges a significant risk premium for contracts in more volatile settings. A more recent study by Bollerslev & Todorov (2011) following the 2008 financial crisis similarly found that during times of higher stock market volatility and investor uncertainty, the risk-premia for options increased significantly.

While there are many studies on sports betting risk from the perspective of bettors, there is significantly less literature on risk from the perspective of sportsbooks. A study conducted by Arscott (2022) found that legal sportsbooks operated in a more risk-seeking way than illegal sportsbooks. The concept of commission, which assesses the average gain from a betting transaction for sportsbooks, used by Arscott (2022) will be used and further developed in this paper to understand the risk-taking behaviors of these institutions. Berkowitz et al. (2017) conducted related research. In their 2017 paper, they assessed money line data from multiple professional sports leagues to calculate win percentages at each money line level using several different methods. This paper will also identify the win percentage at each money line level but will use sample data instead of the calculation methods provided by Berkowitz et al. (2017).

There is one study that serves to compare investing and betting markets. Moskowitz (2021) compared pricing in the sports betting market to asset pricing in the financial market. The significance of momentum, size, and characteristic of betting contracts were assessed when examining pricing. Moskowitz (2021) looked at pricing inefficiencies from the perspective of buyers of these assets, as opposed to the sellers. Theoretical betting strategies were used to test the validity of investing strategies in a separate speculative market, where stronger assumptions could be made about the asset. While Moskowitz (2021) found that there was evidence of momentum-based premia in sports betting, one major flaw is that they were unable to draw any conclusions on risk comparisons between the two industries.

Many previous papers have attempted to calculate risk and evaluate risk mitigation strategies within these two industries individually, but there has yet to be literature that analyzes the risk-taking behavior in the options and sports betting industries side-by-side. Comparing within-industry risk is important when assessing actions at an individual level, but when looking at an aggregate level, comparing price-setting actors in separate speculative industries can reveal market-level inefficiencies. This paper will serve to bridge the gap in this literature by comparing the risk taken on by options sellers and sportsbooks. The main reason sports betting and options trading were chosen as complementary markets is that the structure of buy options and money line sports bets are innately similar, allowing for comparative risk calculations. This will be discussed further in the background section. In addition, both sportsbooks and options sellers are price setters, allowing for the implementation of favorable pricing strategies. Furthermore, the pricing in both markets is dependent on consumer expectations. For example, if everybody expects the price of Apple to increase, the value of call options increases. Similarly, if everybody expects Skidmore soccer to beat Vassar, the odds on the game will shift, changing the potential payout outcomes.

The reason sports betting was chosen over other gambling markets, like casino gambling, is that casino gambling outcomes have definite probabilities. This is to say there is a guarantee of how much money a casino will win on average when gamblers play games such as slots, roulette, and black-jack. Because these probabilities are known, there is no real speculative aspect to these markets in aggregate. In sports betting, however, there is no known true probability that an underdog will beat a favorite. Instead, these probabilities are estimated, creating potential misspecification in price setting. Similarly, the success of options trades are dependent on the probability that a stock price moves a certain amount. While this probability can be estimated, the true probability remains unknown. Another key difference is that the pricing and payout of casino games are not dependent on consumer expectations.

The data used to analyze risk between these two markets includes betting data from 61,424 Division I college basketball games from 2007 to 2019, stock market data from over 200 randomly picked stocks from 1997-2017, and 15,173 single-day options from 2021-2022. College basketball data was used for the sportsbook data due to the variability in money lines, availability of large quantities of data, and the large scale of March Madness-related bets. Money line data was chosen because it entails uneven payout structures related to the probability of an individual event occurring, which is similar to that of options. This will be clarified in the background and conceptual framework sections. Single-day options were chosen to represent the options market for a few reasons. Firstly, part of pricing options includes extrinsic value, which is based on the time remaining on the contract among other things. Having a set of data with the same time remaining thus reduces the variability of the extrinsic value of the options, meaning that pricing differences will mainly be due to their intrinsic value. Secondly, it is much easier to calculate potential payouts over a one-day period than over varying amounts of time. Lastly, single-day options represent day trading, which significantly changes the risk preference of the investor, aligning it more closely to that of a sports gambler. The idea that the decision-making behavior of investors in short-term markets is similar to those in gambling markets is empirically proven by Arthur, et al. (2016).

Two risk statistics, called risk premium and tail risk ratio, are calculated for each industry. Risk premium is based on average returns from a trade while tail risk ratio is dependent on the volatility of potential returns. It is hypothesized that these two variables will have a positive correlation with one another because people will be willing to take on more tail risk at higher expected return levels. These statistics are compared in each market using a difference in means and difference in medians test. It was found that for both statistics and using both methods of comparison, options traders took on significantly more risk than sportsbooks. In addition, there was no significant relationship between the two risk statistics.

This paper will be organized into several sections. The upcoming Background section will give more detailed information on how options trading and sports betting function. It will also include an explanation of how the two fields are compatible at all. The Conceptual Framework section will follow with the derivation of the risk calculation methods. A Data section will then provide more information on the data used and how it relates to the risk calculation methods. The Methodology section will then outline how the data will be used to execute the conceptual framework, as well as the method of comparison between the risk statistics in the two industries. This will be followed by a Results section which will show the outcome of these calculations and how they compare to the initial hypotheses. The Discussion section will explain the results and the possible reasonings behind them.

Background

I. Options:

Stock options give people the right, but not the obligation, to buy or sell stocks at a certain pre-agreed-upon price, called the strike price. In exchange for this right, option buyers pay a premium to the sellers, which is called the ask price. The total amount of premiums paid by the buyer is equal to the ask price times the number of total contracts of the option purchased. Many things impact the ask price of the option. This includes the distance between the strike price and the current stock price, the time remaining on the option before expiry, and the volatility of the stock amongst other things. There are also many advanced metrics that determine the "fair" price of an option, commonly referred to as "The Greeks" because they are all denoted as Greek letters. There are plenty of papers that derive different options pricing models using these metrics, such as Black & Scholes (1968), Merton (1973), Johnson & Stutz (1987), Bates (1996, 2003), Linetsky (1997), Gârleanu et al. (2009), and Aguilar et al. (2019) among others. This paper will not attempt to develop an options pricing model but rather assume that previous pricing methods are being applied to the sample data.

There are two types of options: call and put options. A call option gives the buyer of the option the right to buy the stock at the strike price before the expiration of the contract. This would mean that the buyer expects the market price of the stock to be higher than the strike price before the expiration of the option. For example, a call option with a strike price of $102 and a current stock price (referred to as the spot price) of $100 would give the buyer of the option the right to buy the stock at any point during the contract for $102. Assume the ask price for this option is $1 per share. If the market price of the stock increases to $105, the buyer can execute the option to buy at $102 and sell their shares in the open market for $105 with a $2 profit per share ($3 subtracted by the $1 premium per share). If the market price of the stock does not exceed $102, though, the buyer will let the option expire, as it would not make sense to buy the stock for $102 when they could get it on the open market for cheaper.

A put option gives the buyer of the option the right to sell a stock at the strike price before the expiration of the contract. This means the buyer anticipates that the stock price will be lower than the strike price at some point when they hold the contract. Using the same example as above, assume the strike price of the option is $102 and the spot price is $100. This would give the buyer of the contract the right to sell the stock at $102 per share, which is higher than the current market price of $100 per share. Thus, for the contract to be worth anything for the seller, the premium must be higher than the $1 per share set above, otherwise the buyer of the option will simply execute the option immediately and make a $1 profit per share. Assume now that the premium is set at $4 per share by the seller. This would mean the market price of the stock would need to decrease to $98 for the option to break even for the buyer. If the market price increases beyond the strike price, though, the option will not be profitable, and the buyer will be forced to incur losses equal to the $4 premium per share times the number of shares the option contained.

Using the above two examples, a formula can be created to denote the gains and losses of both call and put options for the buyer of the contract. It is important to note that in the below formula, the profit for the buyer is equivalent to the loss of the seller and vice-versa:

1. ***Call Option Returns***

$$\text{Call Options: Profit} = 1_{\text{Strike} \leq \text{Market Price}} * ((\text{Market Price} - (\text{Strike} + \text{Ask})) * \text{Q}) - 1_{\text{Strike} > \text{Market Price}} * (\text{Ask} * \text{Q})$$

2. ***Put Option Returns***

$$\text{Put Options: Profit} = 1_{\text{Strike} \geq \text{Market Price}} * (((\text{Strike} - \text{Ask}) - \text{Market Price}) * \text{Q}) - 1_{\text{Strike} < \text{Market Price}} * (\text{Ask} * \text{Q})$$

It is assumed that the secondary options market does not exist, meaning buyers of options cannot sell their options to other buyers before expiration. One reason for this is that the sample of options data used contains only single-day options, meaning it would be more difficult to manage buying and selling the option for profit over one day. Additionally, this paper focuses on the seller of the option, who must pay the equivalent losses of the option regardless of who holds it. Similarly, the paper will assume that risk mitigation techniques used by options sellers, such as purchasing a stock as it approaches the strike price to reduce losses, are not used in the sample data. One reason for this is that there is no data to ensure that these risk mitigation measures took place. Thus, it can be deduced that the losses and profits of the seller presented in the option are equal to their total losses and profits on the trade.

II. Sports Betting:

There are many different ways to bet on sports, including player prompts, spreads, over/unders, and parlays which combine multiple bets. This paper, however, will solely focus on money line bets. As previously mentioned, the main reason for this is that the payout structure is very similar to that of options. A money line bet is a bet placed on Team A to beat Team B. Not all teams are evenly matched, which means that betting for teams with a worse probability of winning will pay out higher amounts than betting on favorites. These different payout structures are defined in what is called the "money line odds". A positive money line represents a more than 1:1 payout and is typically assigned to the underdogs. For example, if Team A is an underdog and Team B is a favorite, a +200 money line denotes that a $100 bet on Team A returns an additional $200 if Team A beats Team B (2:1 payout). A negative money line is given to favorites and constitutes a less than 1:1 payout. For example, a -200 money line would signify a $200 bet that returns an additional $100 if Team B beats Team A (1:2 payout). If the team that is bet on loses, in either case, the losses are equal to the amount that was bet. It will be assumed that risk mitigation techniques employed by gamblers and sportsbooks, such as cashouts, are not available. Using these two examples, a money line payout formula can be produced. Again, it is important to note that the profits of bettors are equivalent to the losses of sportsbooks and vice-versa:

3. *Positive Money Line Returns*

$$\text{Returns}_{+\text{ML}} = 1_{\text{Team A beats Team B}} * \left(\text{ML}/100 * \text{bet sizeA}\right) - 1_{\text{Team B beats Team A}} * \left(\text{bet sizeA}\right)$$

4. *Negative Money Line Returns*

$$\text{Returns}_{-\text{ML}} = 1_{\text{Team B beats Team A}} * \left(-100/\text{ML} * \text{bet sizeB}\right) - 1_{\text{Team A beats Team B}} * \left(\text{bet sizeB}\right)$$

5. *Total in-Game Returns*

$$\text{Total Returns} = \text{Returns}_{+\text{ML}} + \text{Returns}_{-\text{ML}}$$

It is visible in equations 1 through 4 that there is an innate similarity in the structure of returns for these investing and betting tools. The returns of options are inherently tied to the probability that the market price moves above or below the strike price, just as the returns of money line bets depend on the probability of Team A beating Team B. There is also a similarity in that the potential losses are equal to the probability of the necessary event *not* occurring times the price of the bet or premium.

III. Betting and Options Behavior Similarities:

While these two fields have similar payout structures, there may be differing behaviors among bettors and investors. There is existing evidence that people act differently in investing settings versus gambling settings (Deck et al, 2014). In Deck et al. (2014)'s study, people were put in either a casino setting or an investment setting. In each case, participants were given the choice between two assets. Both assets had the same expected return but the first option was lower risk compared to the second. Deck et al. (2014) found that the people placed in the casino setting were more likely to choose the riskier asset, implying that the risk preference of gamblers is higher than that of investors. Another study by Humphries et al. (2013) revealed that sports gamblers tend to be more interested in consumption benefits than wealth maximization, showing that they tend to make irrational decisions simply for the fun of it.

While these studies imply incompatibility between sports betting and investing behaviors, they lack in certain areas. Deck et al. (2014) found that there were only minor differences in risk-seeking activity based on the domain setting. In addition, most sports betting and day trading of options occur online, meaning the investor or bettor is always in the same domain when making their trade or bet. Humphries et al. (2013) showed that bettors act irrationally when betting due to consumption benefits. However, the researchers neglected to mention how certain short-term investments (namely day-trading activities) are mostly speculative and not rationally motivated by long-term wealth maximization either. Arthur et al. (2016) revealed that, while investing and gambling behaviors tend to be distinct, gambling and speculative behaviors tend to be very similar. This is to say that people investing in long-term investments, such as pension funds, are likely to be very risk averse. Meanwhile, gamblers and speculative investors are more likely to engage in risk-seeking activities, even in circumstances that diminish expected wealth. A similar study by Ryu et al. (2019) found that speculative behavior in the Bitcoin market led investors to lose self-control related to investment decisions. Additionally, speculative behavior led to a significant decrease in returns among these investors (Rye et al., 2019).

The options trading market has been taken over by single-day options in the past few years. It was estimated that in 2023, roughly 43% of all SPX index options were single-day expiry, as opposed to only 17% in 2020 (CBOE, 2023). The chosen options data for this paper consists of only single-day options, which implies that every one of them was day traded. Due to the short-term nature of these contracts, it is much more likely that these were used as vehicles of speculation as opposed to investment. Additionally, the very nature of options are speculative, as they bet on the movement of individual stock prices over time. While there have been countless attempts to predict the movement of individual stocks (as well as the stock market as a whole), there has yet to be anyone successful at it. Even Warren Buffet has gone on to say that people would be better off investing in index funds than trying to beat the market through the purchase of individual stocks (Malkeil, 1973). Because of this, the findings of this paper will be based on random walk theory, supported by Burton Malkeil in his 1973 book *A Random Walk Down Wall Street*, which implies that the movement of individual stock prices are determined by the previous price of the stock and some random error term (Malkeil, 1973). An equation denoting this is shown below. This theory has been supported empirically in the US stock market by Van Horne & Parker (1967), Fama (1995), and Meerschaert & Scalas (2006) among others. With this assumption in hand, it can be stated that all single-day options trades, and options trades of any length, are speculative, as they depend primarily on random error.

6. *Random Walk Theory*

$$\text{Random Walk: } y_t = y_{t-1} + \varepsilon_t$$

IV. Betting and Options Differences:

While it has been determined that there are many similarities between the sports betting and options trading markets, there are several key differences as well. The first and most pressing is that sports bets have capped payout structures while options contracts do not. This means that sportsbooks know exactly how much money they will have to payout to winning bettors, equal to the money line odds. Options traders, on the other hand, have unbounded potential payouts determined by the movement of stock prices. Because the movement of stocks is random, options sellers do not know in advance how much they will lose on a trade that they default on. This will have a significant impact on the volatility of expected returns for options sellers, who could face a wide variety of different payout scenarios.

An important difference between these two fields is that, on average, the quantity of money involved in individual options trades is much higher than in individual sports bets. This potentially creates asymmetry between the expected returns and risk profiles between these two markets. The proposed risk statistics account for this difference, as the quantity of money bet in the trade is canceled out in both calculations. This will be more evident when looking at the derivations of these equations in the Conceptual Framework section.

Another difference that is accounted for in this study is that sportsbooks are always price setters while options sellers can be price setters or takers. This has a significant impact on the premium price paid in an options transaction. In settings where the buyer of the option is the price setter, this price is bound to be lower than in cases where it is determined by the seller. Because of this, the ask price, as opposed to the bid price of the option, is used to determine the premium price paid by the option buyer. This implies that the option seller is the price setter, meaning they are able to charge a price that is advantageous to them, just as sportsbooks do when they set their odds.

Conceptual Framework

A universal equation that needs to be understood about speculative and investment assets pertains to the expected returns of these assets. Equation 7 shows that expected returns are equal to expected gains subtracted from expected losses.

7. *General Expected Returns*

$$\mathrm{E(Returns)} = \mathrm{P(gain)} * \mathrm{E(payoff|gain)} - \mathrm{P(loss)} * \mathrm{E(payout|loss)}$$

This equation can be applied to previously derived equations that showed the payout structure of sports betting and options trading. Equations 1 and 2 showed that both the gains and losses were present in the calculation of total returns for options buyers. Similarly, equations 3 and 4 showed the gains and losses from the perspective of bet-takers. Because the rest of this paper will look at these markets from the perspective of sellers, the following breakdown of gains and losses in these markets will be retooled to look at the returns of sellers. Again, it should be remembered that the gains of sellers are equal to the losses of buyers and vice-versa. Additionally, the market price variable will be replaced with a variable showing a percent change in the spot price. This change was made to better suit the concepts and data used in later sections.

In the derivations, blue text represents gains while red text represents losses.

Expected Returns of Sportsbooks:

8. *Positive Money Lines*

$$Returns\ (+ML) = 1_{TeamB\ wins} * bet_A - 1_{TeamA\ wins} * ML/100 * bet_A$$

$$E(Returns) = E(1_{TeamB\ wins} * bet_A) - E(1_{TeamA\ wins} * ML/100 * bet_A)$$

$$E(Returns) = \mathbb{P}(TeamB\ wins) * bet_A - \mathbb{P}(TeamA\ wins) * ML/100 * bet_A$$

9. *Negative Money Lines*

$$Returns\ (-ML) = 1_{TeamA\ wins} * \mathrm{bet_B} - 1_{TeamB\ wins} * -100/\mathrm{ML} * \mathrm{bet_B}$$

$$E(Returns) = E\left(1_{TeamA\ wins} * bet_B\right) - E\left(1_{TeamB\ wins} * -100/\mathrm{ML} * bet_A\right)$$

$$E(Returns) = \mathbb{P}(TeamA\ wins) * bet_B - \mathbb{P}\left(TeamB\ wins\right) * (-100/\mathrm{ML}) * bet_B$$

10. *Total*

$$Total\ Returns = 1_{TeamB\ wins} * bet_A + 1_{TeamA\ wins} * \mathrm{bet_B} - 1_{TeamA\ wins} * ML/100 * bet_A - 1_{TeamB\ wins} * -100/\mathrm{ML} * \mathrm{bet_B}$$

$$\mathrm{E(Returns)} = \mathbb{P}\left(TeamB\ wins\right) * bet_A + \mathbb{P}\left(TeamA\ wins\right) * bet_B - \mathbb{P}\left(TeamA\ wins\right) * ML/100 * bet_A - \mathbb{P}\left(TeamB\ wins\right) * (-100/\mathrm{ML}) * bet_B$$

Expected Returns of Options Sellers

11. *Call Options*

$$\mathrm{Returns} = \sum_{i=1}^{n} 1_{\mathrm{R_i} < (\mathrm{strike}+\mathrm{ask})/\mathrm{spot}} * \mathrm{Ask} * \mathrm{Q} - \sum_{i=1}^{n} 1_{\mathrm{R_i} \geq (\mathrm{strike}+\mathrm{ask})/\mathrm{spot}} * ((1+R_i) * spot - (strike+ask)) * Q$$

$$E(Returns) = E(\sum_{i=1}^{n} 1_{R_i < (\mathrm{strike}+\mathrm{ask})/\mathrm{spot}} * \mathrm{Ask} * Q) - E(\sum_{i=1}^{n} 1_{R_i \geq (\mathrm{strike}+\mathrm{ask})/\mathrm{spot}} * ((1+R_i) * spot - (strike+ask)) * Q)$$

$$E(Returns) = \sum_{i=1}^{n} P(R_i < (strike+ask)/spot) * Ask * Q - \sum_{i=1}^{n} P(R_i \geq (strike+ask)/spot) * ((1+R_i) * spot - (strike+ask)) * Q$$

$\mathrm{R_i}$ represents all possible daily returns of stocks, following the true distribution of possible returns

12. *Put Options*

$$\text{Returns} = 1_{R > (\text{strike} - \text{ask})/\text{spot}} * \text{ask} * Q - 1_{R \leq (\text{strike} - \text{ask})/\text{spot}} * ((strike - ask) - (1 + R_i) * spot) * Q$$

$$E(Returns) = E(\sum_{i=1}^{n} 1_{R_i > (\text{strike} - \text{ask})/\text{spot}} * ask * Q) - E(\sum_{i=1}^{n} 1_{R_i \leq (\text{strike} - \text{ask})/\text{spot}} * ((strike - ask) - (1 + R_i) * spot) * Q)$$

$$E(Returns) = \sum_{i=1}^{n} P(R_i > (strike - ask)/spot) * ask * Q - \sum_{i=1}^{n} P(R_i \leq (strike - ask)/spot) * ((strike - ask) - (1 + R_i) * spot) * Q$$

I. Risk Premium

The first risk calculation method used is “Risk Premium”. To prevent confusion between the proposed “Risk Premium” statistic and the premium paid for an option, the former will always be referred to as risk premium while the latter will be called the premium price. The point of this value is to show what the option seller or sportsbook charges the buyer per dollar to hold risk. This calculation is highly dependent on the average returns of the seller, and essentially shows the percent profit they make from a dollar bet by the buyer. Higher values represent lower risk for the seller, as it shows that their average return from a bet is high. The possible values range from -∞ to 1. A value of -2, for example, would show that for every $1 bet by the buyer, the seller would lose an additional $2 on average. For sportsbooks, this could happen in the case of a +200 payout that occurs 100% of the time. In addition, a value of 1 represents an average return of $1 per dollar bet by the buyer, meaning that the seller makes 100% profit on whatever is bet by the buyer. This could take place when someone places a bet on something that has a 0% probability of occurring. This statistic is the same as the concept of commissions used by Arscott (2022) and per-dollar expected returns used by Moskowitz (2021). The following equations show the derivations of the formula for both general and domain-specific cases.

13. General Risk Premium

$$E(R) = E(G) - E(L)$$

$$RP = \frac{E(R)}{bet} = \frac{E(G) - E(L)}{bet}$$

14. Risk Premium for Sportsbooks

$$RP = \frac{E(\text{Returns})}{\text{bet}_A + \text{bet}_B} = \frac{\mathbb{P}\left(TeamB\ wins\right) * bet_A + \mathbb{P}\left(TeamA\ wins\right) * bet_B - \mathbb{P}\left(TeamA\ wins\right) * ML/100 * bet_A - \mathbb{P}\left(TeamB\ wins\right) * (-100/\text{ML}) * bet_B}{\text{bet}_A + \text{bet}_B}$$

$$RP = \mathbb{P}\left(TeamB\ wins\right) * \frac{\text{bet}_A}{\text{bet}_A + \text{bet}_B} + \mathbb{P}\left(TeamA\ wins\right) * \frac{bet_B}{bet_A + bet_B} - \mathbb{P}\left(TeamA\ wins\right) * ML/100 * \frac{bet_A}{bet_A + bet_B} - \mathbb{P}\left(TeamB\ wins\right) * (-100/\text{ML}) * \frac{bet_B}{bet_A + bet_B}$$

15. *Risk Premium for Call Options*

$$E(Returns) = \sum_{i=1}^{n} P\left(R_i < (strike + ask)/spot\right) * \text{ask} * Q - \sum_{i=1}^{n} P\left(R_i \geq (strike + ask)/spot\right) * \left(\left(1 + R_i\right) * spot - (strike + ask)\right) * Q$$

$$RP = \frac{E(Returns)}{Q * ask} = \sum_{i=1}^{n} P\left(R_i < (strike + ask)/spot\right) - \frac{\sum_{i=1}^{n} P\left(R_i \geq (strike + ask)/spot\right) * \left(\left(1 + R_i\right) * spot - (strike + ask)\right)}{\text{ask}}$$

16. *Risk Premium for Put Options*

$$E(Returns) = \sum_{i=1}^{n} P\left(R_i > (strike - ask)/spot\right) * ask * Q - \sum_{i=1}^{n} P\left(R_i \leq (strike - ask)/spot\right) * \left((strike - ask) - \left(1 + R_i\right) * spot\right) * Q$$

$$RP = \frac{\text{E(Returns)}}{\text{ask} * \text{Q}} = \sum_{i=1}^{n} P\left(R_i > (strike - ask)/spot\right) - \frac{\sum_{i=1}^{n} P\left(R_i \leq (strike - ask)/spot\right) * \left((strike - ask) - \left(1 + R_i\right) * spot\right)}{ask}$$

To understand what equation 13 means, it is important to think from the perspective of a risk taker in a betting situation. For example, two people who place a bet on the outcome of a coin flip. If the coin comes up heads, person A pays $1 to person B. If it comes up tails, person B pays $1 to person A. This is an example of a fair bet. Both people have a 50% chance of default (i.e. losing the bet) and a payout that doubles their money if they win. Thus, the expected return, as well as risk premium, will be equal to 0 for both people. Assume now that person A and person B engage in a new bet using a random number generator with possible outcomes of 1-1,000. If person A guesses the correct number, person B will pay them $999. If person A guesses the wrong number, they will pay person B $1. This is another example of a fair bet, but person B is now exposed to significantly more payout risk than person A. Because of this, they change the bet. Now person A will only receive $990 if they guess the right number. Using equation 13, it can be found that person B is now charging a risk premium of 0.9% for taking on this excess risk. This means that the expected return for person B in this bet is positive in the long run ($0.009 per dollar bet by person A). Price setters in speculative markets use risk premiums to create a platform of long-run profitability over time. Comparing their risk premiums will give more insight into which industry charges a sufficient premium.

It is hypothesized that there is no significant difference between the risk premium charged by sports books and single-day options sellers in this study. This is because, in the data used, both sports books and options sellers are price setters in their markets, allowing them to determine a fair risk premium based on consumer behavior. It has already been determined that consumer behavior should be relatively similar in these two markets, and because of this, the risk premiums should be relatively similar as well. Arscott (2022) found that, on average, the value of "commissions" charged by sportsbooks was roughly 4.6%. Because of this, it will be assumed that both options sellers and sportsbooks charge a risk premium at around this rate.

II. Tail Risk Ratio

The second risk measurement is called "Tail Risk Ratio". While risk premium was a statistic that depended on the average gains and losses of a given bet or investment, tail risk ratio is a measure that captures risk based on the volatility of expected returns. This is necessary to implement when looking at tail risk events, which may be unlikely on average but pose a problem when they occur. A real-world example of tail risk causing industry problems can be seen when looking at insurance companies. Natural disasters are rare events historically, and because of that, predicting the probability that they occur is very difficult. Small changes in the probability of these events massively change the potential losses of insurance companies that have to cover property damage during extreme weather events. This is mainly because the cost of repairing property after natural disasters is huge, especially when whole cities are damaged.

Thus, insurance companies had to stop insuring certain areas where there was too much risk of natural disaster (Jacobsen, 2024). Essentially, they did not account for tail risk in their models until very recently - when it started hurting them. Using the context of options and sports betting, tail events are still prevalent. For options, massive stock price movements are rare, but when they occur, options sellers could have losses that are orders of magnitude larger than their previous expected gains. Sportsbooks have much more limited tail risk since they have capped payouts (equal to the money line odds). However, tail risk regarding extremely high money line odds persists due to the random nature of sports and the high payout possibilities attributed to these money lines.

When calculating tail risk, more weight should be assigned to higher potential payout values. One way to do this is to square each loss value and calculate the expected-squared payouts. Using Jensen's inequality, it can be seen that this will always result in a higher value than by taking the expected payout and squaring it. Taking the difference between these two values (expected-squared losses and squared expected losses) allows the tail risk of a trade or bet to be assessed. The problem with this method is that it does not account for the potential gains of the trade. To fix this, the tail risk ratio calculation will contain expected squared gains subtracted by squared expected gains in the denominator. The reason a ratio was chosen instead of subtracting gains from losses again is that these values tend to be extremely high, meaning a ratio will be much more interpretable than a difference. When trying to interpret the values of the tail risk ratio, it is easier to think of the expected gains of sellers in terms of expected losses for buyers. Values greater than 1 imply that the tail risk for sellers (numerator) is larger than the tail risk for buyers (denominator), while values between 0 and 1 would indicate the opposite. In all, tail risk ratio compares the tail risk of buyers and sellers to see which party has a volatility-based risk advantage. The general and domain-specific tail risk ratio calculations, influenced by the concept of “Risk Edge” proposed by Taleb (2012) and Jensen’s inequality, are shown below. This calculation differs from Taleb’s risk edge in two main ways. The first is that it is from the perspective of sellers in speculative markets as opposed to buyers. Secondly, it is interpretable as a ratio containing expected gains and losses instead of a difference pertaining to expected returns. This is an important distinction to make when comparing the risk in trades with positive and negative expected returns. Using Taleb’s method, trades with the same absolute value of expected returns will have the same risk edge, regardless of the sign. Tail risk ratio fixes this issue by breaking down gains and losses individually and comparing them using a ratio.

17. *Tail Risk Ratio for Sportsbooks*

$$E(Loss) = \left(P\left(TeamA\ wins\right) * ML/100 * bet_A\right) + \left(P\left(TeamB\ wins\right) * -100/ML * bet_B\right)$$

$$E(Loss^2) = \left(P\left(TeamA\ wins\right) * (ML/100 * bet_A)^2\right) + \left(P\left(TeamB\ wins\right) * (-100/ML * bet_B)^2\right)$$

$$E(Loss)^2 = \left(P\left(TeamA\ wins\right) * ML/100 * bet_A\right)^2 - \left(P\left(TeamB\ wins\right) * -100/ML * bet_B\right)^2$$

$$E(Gains) = \left(P\left(TeamA\ wins\right) * bet_B\right) + \left(P\left(TeamB\ wins\right) * bet_A\right)$$

$$E(Gains^2) = \left(P\left(TeamA\ wins\right) * (bet_B)^2\right) + \left(P\left(TeamB\ wins\right) * (bet_A)^2\right)$$

$$E(Gains)^2 = \left(P\left(TeamA\ wins\right) * bet_B\right)^2 + \left(P\left(TeamB\ wins\right) * bet_A\right)^2$$

$$Tail\ Risk\ Ratio = \frac{E\left(Loss^2\right) - E(Loss)^2}{E\left(Gains^2\right) - E(Gains)^2}$$

$$Tail\ Risk\ Ratio = \frac{\left(P\left(TeamA\ wins\right) * \left(ML/100 * bet_A\right)^2\right) + \left(P\left(TeamB\ wins\right) * \left(-100/ML * bet_B\right)^2\right) - \left(P\left(TeamA\ wins\right) * ML/100 * bet_A\right)^2 - \left(P\left(TeamB\ wins\right) * -100/ML * bet_B\right)^2}{\left(P\left(TeamA\ wins\right) * \left(bet_B\right)^2\right) + \left(P\left(TeamB\ wins\right) * \left(bet_A\right)^2\right) - \left(P\left(TeamA\ wins\right) * bet_B\right)^2 + \left(P\left(TeamB\ wins\right) * bet_A\right)^2}$$

18. Tail Risk Ratio for Call Options

$$\mathrm{E(Loss)} = (\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i \geq (strike + ask)/spot\right) * ((1+\mathrm{R_i}) * \mathrm{spot} - (\mathrm{strike}+\mathrm{ask})) * Q)$$

$$\mathrm{E(Loss^2)} = \sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i \geq (strike + ask)/spot\right) * \left(\left(\left(1+\mathrm{R_i}\right) * \mathrm{spot} - (\mathrm{strike}+\mathrm{ask})\right) * Q\right)^2$$

$$\mathrm{E(Loss)}^2 = \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i \geq (strike + ask)/spot\right) * \left(\left(1+\mathrm{R_i}\right) * \mathrm{spot} - (\mathrm{strike}+\mathrm{ask})\right) * Q\right)^2$$

$$\mathrm{E(Gains)} = (\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i < (strike + ask)/spot\right) * ask * Q)$$

$$\mathrm{E(Gains^2)} = \sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i < (strike + ask)/spot\right) * \left(ask * Q\right)^2$$

$$\mathrm{E(Gains)}^2 = \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i < (strike + ask)/spot\right) * ask * Q\right)^2$$

$$\text{Tail Risk Ratio} = \frac{\mathrm{E(Loss^2)} - \mathrm{E(Loss)}^2}{\mathrm{E(Gains^2)} - \mathrm{E(Gains)}^2}$$

$$\text{Tail Risk Ratio} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i \geq (strike + ask)/spot\right) * \left(\left(\left(1+\mathrm{R_i}\right) * \mathrm{spot} - (\mathrm{strike}+\mathrm{ask})\right) * Q\right)^2 - \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i \geq (strike + ask)/spot\right) * \left(\left(1+\mathrm{R_i}\right) * \mathrm{spot} - (\mathrm{strike}+\mathrm{ask})\right) * Q\right)^2}{\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i < (strike + ask)/spot\right) * \left(ask * Q\right)^2 - \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i < (strike + ask)/spot\right) * ask * Q\right)^2}$$

19. Tail Risk Ratio for Put Options

$$\mathrm{E(Loss)} = (\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i \leq (strike - ask)/spot\right) * ((\mathrm{strike}-\mathrm{ask}) - (1+\mathrm{R_i}) * \mathrm{spot}) * Q)$$

$$\mathrm{E(Loss^2)} = \sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i \leq (strike - ask)/spot\right) * \left(\left((\mathrm{strike}-\mathrm{ask}) - \left(1+\mathrm{R_i}\right) * \mathrm{spot}\right) * Q\right)^2$$

$$\mathrm{E(Loss)}^2 = \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i \leq (strike - ask)/spot\right) * \left((\mathrm{strike}-\mathrm{ask}) - \left(1+\mathrm{R_i}\right) * \mathrm{spot}\right) * Q\right)^2$$

$$\mathrm{E(Gains)} = (\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i > (strike - ask)/spot\right) * ask * Q)$$

$$\mathrm{E(Gains^2)} = \sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i > (strike - ask)/spot\right) * \left(ask * Q\right)^2$$

$$\mathrm{E(Gains)}^2 = \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i > (strike - ask)/spot\right) * ask * Q\right)^2$$

$$\text{Tail Risk Ratio} = \frac{\mathrm{E(Loss^2)} - \mathrm{E(Loss)}^2}{\mathrm{E(Gains^2)} - \mathrm{E(Gains)}^2}$$

$$\text{Tail Risk Ratio} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i \leq (strike - ask)/spot\right) * \left(\left((\mathrm{strike} - \mathrm{ask}) - \left(1 + \mathrm{R_i}\right) * \mathrm{spot}\right) * Q\right)^2 - \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i \leq (strike - ask)/spot\right) * \left((\mathrm{strike} - \mathrm{ask}) - \left(1 + \mathrm{R_i}\right) * \mathrm{spot}\right) * Q\right)^2}{\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i > (strike - ask)/spot\right) * \left(ask * Q\right)^2 - \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} P\left(R_i > (strike - ask)/spot\right) * ask * Q\right)^2}$$

Using the previous examples with person A and person B, these equations can be better understood. It was found that the risk premium was 0 during the coin flip bet and the initial number generator bet. The tail risk ratio, however, is quite different among these two examples. In the coin flip bet, the expected losses of person B are equal to the 50% chance they lose times the $1 they pay given that they lose. This gives an expected squared loss - E(loss^2) - of $0.5, and a squared expected loss - E(loss)^2 - of $0.25. The values for the expected gains are the same, resulting in a tail risk ratio of 1. In the random number generator bet, though, there is a 0.1% chance that person B loses and a $999 payout given that they lose. This would give an expected squared loss of 998.01 and a squared expected loss of 0.998. The gains are dependent on the 99.9% chance of losing the bet and the $1 they pay when they lose. This returns an expected squared gain of 0.999 and a squared expected gain of 0.998. By taking the ratio of the differences, the tail risk ratio is equal to roughly 997,998. It can be seen that, while the risk premium is the same for both of these examples, the tail risk ratio is quite different because person B has a large potential payout when they lose while person A does not.

It is hypothesized that sports betting markets have significantly lower tail risk ratios than options sellers at each probability of default level. This is because the potential payouts for sportsbooks are capped, which is not the case for options sellers, who could deal with a stock that increases by 1% or 100000%. A similar hypothesis was formed by Taleb (2012) when assessing the risk edge for buyers in speculative markets. Taleb (2012) noted that because of the fat-tailed distribution of stock market returns and the unbounded potential returns for buyers, options sellers face much higher risk.

Data

The data used in this paper was gathered from several separate sources. College basketball betting data was obtained from Sportsbook Reviews Online (2024). This data contains money line and score result data from 61,250 DI college basketball games from 2007-2019. This data was used to calculate the win percentage at each money line spread (ML_A - ML_B) level to find the probability of default for sportsbooks. One flaw with this data is that observations for games with very high positive and negative money line spread values are scarce, meaning the win percentage distribution could be skewed. A binning approach was used to increase the number of observations at each money line spread level, but the more extreme money line values still hold significantly less data than observations closer to the mean.

Historical betting data was obtained from Sports Betting Dime (2024) to find the percentage of bets placed on teams at each money line level. This was used to assess the favorite and underdog bet size proportion, necessary for computing the tail risk ratio and risk premium values for sportsbooks. One limitation of this data is that the bet proportions were not found for the sample data, but rather from 2023 and 2024 data. Because of this, a loess regression was used to predict the bet proportion for the sample data.

The historical options data was acquired from Options Metrics (2024). This data consists of 15,173 single-day options trades for 15 randomly picked stocks from February 2022 through February 2023. This data contained the ask price, bid price, and strike price of the contracts as well as whether it was a call or put option. One problem with this data is that it did not contain the spot price of the option, so data from Data Stream was used to find the opening prices of these stocks for each day in the sample size. This could be a problem because it assumes that the spot price of the contract is equal to the opening price of the stock, meaning the contract was bought at the beginning of the day. While this is a large generalization, there was no existing data that contained all the necessary variables including spot price. This limits the accuracy of the risk calculations, as the prices and necessary returns in the data may not coincide with the real-life values. This is one of the main limitations provided by the data.

Lastly, the stock market data was procured from Boris Marjanovic, Kaggle (2018). This data contains single-day high and low stock prices for over 200 randomly picked stocks from 1997-2017. The total number of observations in the dataset is 1,690,088 (845,044 lows and highs). This data was used to approximate the odds that a stock moves a certain percentage within one day. Using this, and the assumptions of random walk theory, the probability that an option reaches its necessary returns for breakeven can be calculated, giving the probability of default for each option. Similarly to the money line data above, extreme values of returns are scarce, meaning the large potential payouts involved with tail risk are over-specific to the data used in this study. While the large amount of data serves to combat this potential limitation, the extreme values could still be cherry-picked, resulting in over or underrepresented outlier data.

Table 1 shows the summary statistics of key variables from the above datasets. It can be seen that the mean absolute value of high and low returns is the same. The standard deviation for these two measures is slightly different, though. This makes sense because a stock price can only depreciate up to 100%, but can appreciate an unbounded amount. Another interesting feature of the dataset is that the average necessary returns on an option is positive, showing that there are more high-necessary-return call options than put options in the data. As far as the betting data, the average money line spread is 0, which makes sense given that the spread of the favorite and underdog in a given game should cancel out. Another feature is that the average money line is negative. This is intentionally done by sportsbooks, because arbitrage opportunities would occur if the absolute value of the favorite money line is less than the underdog money line.

Methodology

Now that the conceptual framework and available data are clear, the methodology for calculating and comparing risk in the options and sports betting markets can be better understood.

I. Sports Betting:

Going back to the risk premium and tail risk ratio calculations from the Conceptual Framework section, the necessary independent variables include the probability of winning, proportion of money bet, and money line for both the favorites and underdogs. Of these, the money line data was the easiest to obtain, as it was included as a variable in the Sportsbooks Reviews Online dataset.

The proportion of money bet on the favorite and underdog was not given in the initial set, and because of this, needed to be estimated. To do this, data which includes money line and betting proportions for November 2023 through February 2024 from sportsbettingdime.com was obtained. Using this data, a loess regression was implemented to estimate the betting proportions, with the square root of positive money line spreads (the money line of the underdog subtracted by the money line of the favorite) as the independent variable. This model resulted in a residual standard error of roughly 11.1%. A loess model was chosen over an ordinary least squares regression model for two main reasons: Firstly, the residual standard error was roughly 0.4 percentage points lower for the loess model. Secondly, the linear model was not equipped to handle outlier data from the original dataset, which had money line spreads far larger than in the

Sports Betting Dime (2024) set. The loess regression, meanwhile, accounted for the dampening trend as the money line spread increased. The square root of the independent variable was chosen over the raw value to create a stronger relationship between the two variables, in the end reducing the residual squared error by roughly 3 percentage points. Figure 1 shows the sample data with the fitted loess regression curve. The loess model was then used to predict the percentage of money bet on underdogs in the Sportsbook Reviews Online (2024) dataset based on the money line spread. The estimated percentage of money bet on the favorites was found by taking 1 subtracted from the estimated percentage for the underdog from the same game. All estimated values of 1 were rounded to 0.99999 and values of 0 were rounded to 0.00001. This is because there will never be a bet where 100% of the money is put on only 1 team.

The probability of a team winning a game is not something that has a known true quantity, so again, this variable had to be estimated. Using the Sportsbook Reviews Online dataset, each game was binned by its money line spread value. Then, the outcomes of these games were used to determine the probability of an underdog team winning at each given money line spread level. For example, there were roughly 588 instances in the dataset where a game had a money line spread of 500, and the underdog won in these games roughly 155 times, meaning the estimated probability that an underdog wins with a +500 money line spread is around 26.36%. Conversely, the favorite with a -500 money line will win roughly 73.64% of the time (1-0.2636). One problem with this approach is that high money line spreads are scarce compared to low money line spread values. This means the number of sample games for large underdogs and favorites is much smaller than the number of close games in the sample. This could lead to an inaccurate estimated win probability for outlier data, throwing off the value of the risk statistics. A binning approach was used to increase the number of sample games at each money line spread level, but there were still many bins with a very low frequency of data, creating outlier win probability points. For example, there was only 1 game with a money line spread of 8400, and in this instance the underdog won, creating an estimated win probability of 100%. To combat this type of extrapolation, a loess model was used to estimate the win percentage for underdogs at each money line spread level. A loess model was chosen over an ordinary least squares regression model because the relationship between the two variables was not linear. Figure 2 shows the sample data with the fitted loess regression curve. The model contained estimated win percentage as the dependent variable and money line spread as the sole independent variable. The model produced a residual standard error of 0.2% when excluding the two most prominent outliers, and the estimated win probability at the +8400 money line spread level changed from 100% to 3.03%.

II. Options Sellers

When looking at the necessary independent variables for the risk premium and tail risk ratio of options sellers, there are far more variables to account for. Equations 15 through 19 show that spot price, strike price, and ask price of the option are needed as well as the probability of each possible return level. The Options Metrics dataset contained the strike price and ask price. As mentioned in the Data section, the spot price was not included in this dataset but the daily start price of the stock, taken from Data Stream, was used to estimate this variable.

The probability that a stock reached a certain return point was the only parameter that needed to be estimated for the options sellers' risk calculations. Tying this idea back to the Conceptual Framework section, the distribution of possible returns is equal to the distribution of the error term in the random walk equation (the movement of stock prices from their original value). Typically, it is held that the error term is normally distributed with a mean of zero, however, real-world financial data tends to be fat-tailed as opposed to normal. Because of this, historical data was used to create a more accurately defined distribution of the error term. In addition, this term was used to estimate the high and low percent change of the option over a day, instead of the close price. The theoretical interpretation of this is shown below in equations 20 and 21. To create this estimate, single-day stock market data for 200 randomly picked stocks ranging from 1997-2017 was obtained from Boris Marjanovic, Kaggle (2018). The daily high and low change percentage was then taken for all the available data points. This data was again binned, meaning that all x% returns were grouped. For example, a +5.0% high return occurred 1,902 times out of 845,504 total instances. This would mean the estimated probability that a stock will reach exactly 5% returns over a day is around 0.23%. This estimated probability was obtained for every observed high and low value in the dataset, with the observed value rounded to 3 decimal places (i.e. 1.0%, 1.1%, 1.2%, etc.). One problem with this approach is that, once again, very high and low return values are scarce. Using the above method, the estimated probability of these scarce events occurring is most likely extrapolated from the existing dataset and not representative of the true distribution. The main way to mitigate this issue is to use as much data as possible, but even with extensively large datasets, this problem will remain to some end. The reason a model predicting the probability of returns was not implemented is that there is significantly more stock data available than college basketball betting data, meaning the estimate using available data is much closer to the true distribution of returns in this domain. In addition, the estimates of the model would be based on the available outlier data in the set, essentially negating the purpose of using it in the first place.

20. ***High Return Random Walk***

$$\max(y_t) = y_{t-1} + \varepsilon_{t,\max}$$

21. Low Return Random Walk

$$\min(y_t) = y_{t-1} + \varepsilon_{t,\min}$$

III. Comparing Risk Statistics:

Two methods were used for comparing the risk statistics between industries: a difference in means test (t-test), and a Mann-Whitney difference in medians test. The first test was implemented to see if there is a significant difference between the average risk values for sports betting and options trading. Mean values are influenced by outlier data, meaning that this difference accounts for individual instances which returned exceptionally high or low risk values. One problem with this test is that it assumes that the data is normally distributed. The Mann-Whitney difference in medians test was implemented to account for this problem. This test does not assume normally distributed data, and median values are not impacted by outliers. Because of this, the Mann-Whitney test will serve to compare the risk statistics in these two industries in a more robust manner.

Results

One important result of this study was that there was no significant relationship between the risk premium and tail risk ratio statistics. This affirms that the statistics analyzed different aspects of risk in the two observed industries. Figures 3 and 4 depict the relationship between risk premium and tail risk ratio in the sports betting and options trading markets respectively.

Table 2 displays the summary statistics of the risk measures for both sportsbooks and options sellers. The mean risk premium value for sportsbooks is equal to 0.065, implying that they make 6.5% profit on every dollar bet they receive. In contrast, the mean risk premium value for options sellers is equal to -3.47, meaning that they lose $3.47 for every dollar in premium price they receive. An important factor in this is the standard deviation. While the standard deviation for sportsbooks is roughly 0.098, it amounted to around 14.05 for options sellers, implying that there is a much larger spread of values in the options trading dataset. It is important to note that the maximum value for this statistic is 1, so most of the variation will come from outlier negative values. When looking at the minimum values of this risk measure for both industries, it is evident that the options trading data contained some massive outliers, with a risk premium value of -248.96. The sportsbook risk premium had a minimum of -0.42, providing a much tighter range of values. The spread of the data is important to account for when conducting a difference of means test, which assumes normally distributed data. Figures 5 and 6 show the spread of the risk premium data for sportsbooks and options sellers. While the distribution of risk premium values for sportsbooks looks somewhat normal and right skewed, the distribution for options sellers looks completely left skewed. This implies that the data is not normally distributed, and thus the results of the difference in means test will be distorted. The Mann-Whitney difference in medians test will provide stronger analysis for this data due to the skew. The median values of 0.045 for sportsbooks and -0.019 for options traders are much closer together and are likely more representative of the risk premium associated with an average sports bet or options trade.

The mean tail risk ratio for sportsbooks is 2,769. This implies that the tail risk taken on by sportsbooks is roughly 3,000 times higher than that taken by sports bettors. While this seems to be a large value, it is more than 1000 times smaller than the mean tail risk ratio value for options traders, which is around 3.19 million. The standard deviation for sportsbooks is around 35,000 versus roughly 23 million for options traders, showing that the spread of values is much larger in the options dataset. Again, these values are heavily influenced by outlier data and the range of possible values for the risk measure. Tail risk ratio is lower bounded by 0, but has no upper bound, thus most of the outlier data will be in the extreme positives. This can be seen when looking at the maximum tail risk ratio value for options traders, which is 455 million, while the maximum value for sportsbooks is around 1.1 million. To account for the large scale of values in the data, as well as the fact that the lower bound is 0, a log transformation was used to better interpret the tail risk ratio. Table 3 shows the summary statistics of the logged values and Figures 7 and 8 show the spread of these values. Again, the assumption of normally distributed data is broken, reducing the validity of the difference in means test. This means the Mann-Whitney difference in medians test will be more accurate in showing the true differences in the data.

Table 4 shows the results of the difference in means tests while Table 5 presents the outcomes of the difference in medians tests. The difference in means test comparing risk premiums returned a p-value of less than 0.05, meaning the null hypothesis that there is no difference between means can be rejected at the 5% significance level. The 95% confidence interval for the true difference in means was between 3.31 and 3.76. As previously mentioned, this test may be inaccurate due to the non-normal spread of the data. The Mann-Whitney test, which does not assume normally distributed data, also gave a p-value of less than 0.05, again rejecting the null hypothesis that there is no difference between the two sample medians. The 95% confidence interval for the difference in medians was between 0.060 and 0.091. While the Mann-Whitney test significantly reduced the expected difference in risk premium, it was found in both cases that options traders have significantly smaller risk premiums than sportsbooks. This counters the hypothesis that both industries would have a similar risk premium of roughly 4.6%.

Similar results were found when testing the difference between log tail risk ratio in the two industries. Both the difference in means and difference in medians tests returned a p-value far less than 0.05, meaning the tail risk ratio for options traders was significantly higher than sportsbooks'. Unlike the results for risk premium, the difference in means was roughly the same as the difference in medians, with both 95% confidence intervals containing a similar range. This supports the previously proposed hypothesis that the tail risk ratio for options sellers will be significantly larger than that of sportsbooks.

Discussion

This paper sought out to determine whether the risk taking strategies employed by options traders were different to those used by sportsbooks. While previous literature, including Galai & Masulis (1975), Grammaticos (1992), Pan (2002), Kappaguntula & Ruddy (2007), Bollerslev & Todorov (2011), Berkowitz et al. (2017), and Arscott (2022) looked at risk in these two fields individually, there was no side-by-side risk comparison. Moskowitz (2021) was the only study that looked at the betting and investing markets together, but made no conclusions on industry-level risk.

The implementation of tail risk ratio was another novel aspect of this paper. While similar metrics have been used to measure tail risk, such as Taleb's risk edge, tail risk ratio served to compare the tail risk between buyers and sellers to determine who faced a volatility-based risk edge. The concept of risk premium, on the other hand, had been used for similar purposes in previous literature, such as Arscott (2022) and Moskowitz (2021).

One of the main conclusions provided by this paper is that there was no significant relationship between risk premium and tail risk ratio. It was initially hypothesized that these two variables would have a positive relationship, as it was assumed that sellers would be willing to take on more tail risk at higher expected return levels. This proved to be untrue for both sportsbooks and options sellers. One reason for this could be that sellers do not adequately account for tail risk when assessing the value of trades. Another possibility is that the estimated variables were skewed, leading to inaccurate risk calculations. In either case, the lack of relationship between the two variables proves that they look at distinctly different areas of risk.

Using a difference in means and difference in medians test, it was found that the risk premium charged by options sellers on single-day options was significantly smaller than those charged by sportsbooks. This contradicts the hypothesis that both markets charge similar risk premiums due to similar consumer behavior and price-setting ability. Another key finding was that the median risk premium for sportsbooks was roughly 4.5% with a mean of around 6.5%. It appears that outlier data skewed the spread of risk premium data to the right, causing the average value to be higher than the approximately 4.6% found in Arscott (2022). The median value, though, appears to align with the previous literature.

In addition, the tail risk ratio of options sellers was found to be significantly higher than that of sportsbooks for both tests. This coincides with the hypothesis that options sellers have a larger tail risk ratio than sportsbooks due to uncapped potential payouts. This also corroborates with the ideas expressed by Taleb (2012), which stated that options sellers take on significantly more tail risk than options buyers.

These findings imply that options traders take on more risk than sportsbooks when looking at an expected-returns-based framework as well as a volatility-based framework. There could be several reasons for this. One is that the behavior of buyers could be different in the options trading setting. While it was found that speculative and gambling behaviors tend to be similar, it could be that options buyers are more likely to spot disadvantageous trading opportunities than sports bettors. Another potential explanation is that the potential payouts of options traders are unbounded, creating exponentially more risk at any level than sportsbooks, which have capped payouts. Yet another explanation could be that options sellers are charging prices that are too low. This would imply that buyers are willing to pay higher premium prices for options, but the competitiveness of the options selling industry is artificially lowering premium prices beyond a point of maximum profitability. In any case, the findings indicate that options sellers should either increase the premium price or stop engaging in certain trades to lower their risk.

While all of the above explanations possibly serve to explain the difference in risk taken on by sellers in the sports betting and options trading industries, another potential conclusion is that the observed difference is due to data limitations. One of the major limitations of the options data was that the spot price was not included, and thus had to be estimated using the start-price of the stock on the day it was traded. The spot price is especially important, not only because it appears in the risk equations, but because it also affects the necessary returns of the option. If both of these variables were altered in a significant way, the risk calculations would be completely thrown off. Another limitation is that the estimated probability that a stock will reach a certain level of returns is limited by the quantity of data and the extreme values in the data. Small changes to outlier data could significantly impact the findings of the risk calculations. Furthermore, there was no data pertaining to who was buying the option or placing the sports bet. Micro-level data would have allowed for a better comparison of risk preferences within and between the two industries.

Further research is needed to compare risk between these two industries. Risk premium and tail risk ratio both performed well to explain the average and volatility based risk in each industry. More comprehensive data is necessary to come to a better conclusion on the difference in risk for each market. In addition, these statistics could potentially be used to compare different speculative markets than the two covered in this paper.

References

Aguilar J-P, Korbel J, Luchko Y. *Applications of the Fractional Diffusion Equation to Option Pricing and Risk Calculations*. Mathematics. 2019; 7(9):796.

Arscott, R. (2022, December). *Risk Management in the Shadow Economy: Evidence from the Sport Betting Marke*t. Journal of Corporate Finance.

Arthur, J. N., Williams, R. J., & Delfabbro, P. H. (2016, November 24). *The conceptual and empirical relationship between gambling, investing, and speculation*. AKJournals.

Bengel, C., & McCarriston, S. (2023, November 17). *U.S. Sports Betting: Here is where all 50 states currently stand on Legalizing Online Sports Betting Sites*. CBSSports.com.

Berkowitz, J. P., Depken, C. A., & Gandar, J. M. (2018). *The Conversion of Money Lines Into Win Probabilities: Reconciliations and Simplifications*. Journal of Sports Economics, 19(7), 990-1015.

Bollerslev, T. & Todorov, V. (2011). *Tails, Fears, and Risk Premia*. Wiley Online Library, The Finance Journal.

Brad, H., Rodney, P., & Andrew, W. (2013, June 14). *Consumption benefits and gambling: Evidence from the NCAA Basketball Betting Market*. Journal of Economic Psychology.

Dan Galai, Ronald W. Masulis, *The Option Pricing Model and the Risk Factor of Stock*. Journal of Financial Economics, Volume 3, Issues 1–2, 1976, Pages 53-81

The Trustees of Princeton University. (2024). *Options - datastream* - research guides at Princeton University. Princeton University. https://libguides.princeton.edu/datastream/options

David S. Bates, 20 *Testing option pricing models*, Handbook of Statistics, Elsevier, Volume 14, 1996, Pages 567-611.

David S. Bates, *Empirical option pricing: a retrospection*, Journal of Econometrics, Volume 116, Issues 1–2, 2003, Pages 387-404

Deck, C., Lee, J., & Reyes, J. (2014). *Investing versus gambling:experimental evidence of multi-domain risk attitudes*. EBSCO.

Dixon, E. (2023, September 21). What's the state of the US Sports Betting Market in 2023? SportsPro.

ETD Volume - December 2022. FIA. (2023).

Fama, Eugene F. *Random Walks in Stock Market Prices.* Financial Analysts Journal, vol. 21, no. 5, 1965, pp. 55–59.

Fedenia, M., & Grammatikos, T. (1992). *Options Trading and the Bid-Ask Spread of the Underlying Stocks*. The Journal of Business, 65(3), 335–351.

Global Futures and Options Volume 2022. Statista. (2023, March 29).

Jacobsen, L. (2024, February 5). *Insurers such as State Farm and Allstate are leaving fire- and flood-prone areas. home values could take a hit*. CNBC.

Johnson, H. & Stutz, R. (1987). *The Pricing of Options with Default Risk*. Wiley Online Library, The Finance Journal.

Jun Pan, *The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study*, Journal of Financial Economics, Volume 63, Issue 1, 2002, Pages 3-50.

Kappagantula, R., & Reddy, S. (2007). *Estimating Risk and Hedging in Options Trading*. SSRN.

Linetsky, V. *The Path Integral Approach to Financial Modeling and Options Pricing*. Computational Economics 11, 129–163 (1997).

Marjanovic, Borris. *Huge Stock Market Dataset*. Kaggle (2018)

Mark M. Meerschaert, Enrico Scalas, *Coupled Continuous Time Random Walks in Finance, Physica A: Statistical Mechanics and its Applications*, Volume 370, Issue 1, 2006, Pages 114-118,ISSN 0378-4371

McClymont, M. (2024, February 15). *US Sports Betting Revenue Tracker - state by state handle*. Covers.com.

Merton, Robert C. *Theory of Rational Option Pricing.* The Bell Journal of Economics and Management Science, vol. 4, no. 1, 1973, pp. 141–83. JSTOR

Moskowitz, T. J. (2021). *Asset Pricing and Sports Betting*. Wiley Online Library, The FInancial Journal.

MURPHY, GOVERNOR OF NEW JERSEY, ET AL. v. NATIONAL COLLEGIATE ATHLETIC ASSN. ET AL. . Supreme Court. (2017, October).

Nicolae Gârleanu, Lasse Heje Pedersen, Allen M. Poteshman, *Demand-Based Option Pricing*. The Review of Financial Studies, 2009

Wharton Research Data Services. (2024). *Optionmetrics*. WRDS. https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/optionmetrics/

Parry, W. (2023, May 8). *Americans bet $220 billion on sports in 5 years since legalization*. PBS.

NCAAB public betting trends - Daily Bet & Money percentages. Sports Betting Dime. (2024). https://www.sportsbettingdime.com/college-basketball/public-betting-trends/

Historical NCAA Basketball Scores and Odds Archives. Sportsbook Reviews Online. (2024). https://www.sportsbookreviewsonline.com/scoresoddsarchives/ncaabasketball/ncaabasketballoddsarchives.htm

Taleb, N. N. (2012). *Antifragile*. Penguin.

Van Horne, James C., and George G. C. Parker. *The Random-Walk Theory: An Empirical Test*. Financial Analysts Journal, vol. 23, no. 6, 1967, pp. 87–92.

Appendix:

Table 1:

Var Name	Minimum	Median	Maximum	Mean	SD	n	Data Source
Money Line	-500,000	-110	499,980	-254.5	3,913.65	122,848	Sportsbook Reviews Online
Money Line Spread	-200,000	0	200,000	0	6,537.36	122,848	Sportsbook Reviews Online
Percentage of Money	0	0.5	1	0.5	0.337	7,314	Sportsbettingdime .com
Ask Price	0.01	0.4	439	8.93	19.97	15,173	Options Metrics
Strike Price	0.5	131	1,440	202.353	206.95	15,173	Options Metrics
Spot Price	0.786	138.93	774.97	202.137	206.947	15,173	Data Stream
Necessary Returns	-248.4%	0%	1,153.1%	2.1%	2.7%	15,173	Options Metrics
High Change	0%	1.0%	922.2%	1.6%	2.5%	845,044	Boris Marjanovic, Kaggle
Low Change	-90.4%	-1.0%	0%	-1.6%	2.0%	845,044	Boris Marjanovic, Kaggle

Table 2:

Var Name	Minimum	Median	Maximum	Mean	SD	n
Sports Betting Risk Premium	-41.5%	4.5%	99.8%	6.5%	9.8%	61,424
Options Selling Risk Premium	-24,896.2%	-1.9%	100%	-3.47%	1,400.5%	15,173
Sports Betting Tail Risk Ratio	0	1.408	1,138,716	2,769	35,520	61,424
Options Trading Tail	0	821.703	488,407,485	3,194,468	22,937,108	15,173

Risk Ratio						

Table 3:

Var Name	**Minimum**	**Median**	**Maximum**	**Mean**	**SD**	**n**
Sports Betting Log Tail Risk Ratio	-35.621	0.342	13.945	1.073	3.528	61,424
Options Trading Log Tail Risk Ratio	-23.825	6.711	20.007	5.976	6.875	15,173

Table 4:

	Sportsbooks Mean	**Options Sellers Mean**	**Estimated Difference**	**95% Confidence Interval**	**t-value**
Risk Premium	0.065 (0.098)	-3.472 (14.052)	3.551*** ($1*10^{-204}$)	(3.313 , 3.760)	30.998
Log Tail Risk Ratio	1.073 (3.528)	5.976 (6.875)	-4.903*** ($2*10^{-500}$)	(-5.016 , -4.790)	-85.124

Table 5:

	Sportsbooks Median	**Options Sellers Median**	**Estimated Difference**	**95% Confidence Interval**	**W-value**
Risk Premium	0.045	-0.019	0.086*** ($1*10^{-50}$)	(0.070 , 0.092)	429,530,277.5
Log Tail Risk Ratio	0.342	6.711	-4.802*** ($2*10^{-500}$)	(-4.892 , -4.594)	648,237,131

****** = significant at 0.01***
***** = Significant at 0.05***
**** = significant at 0.1***

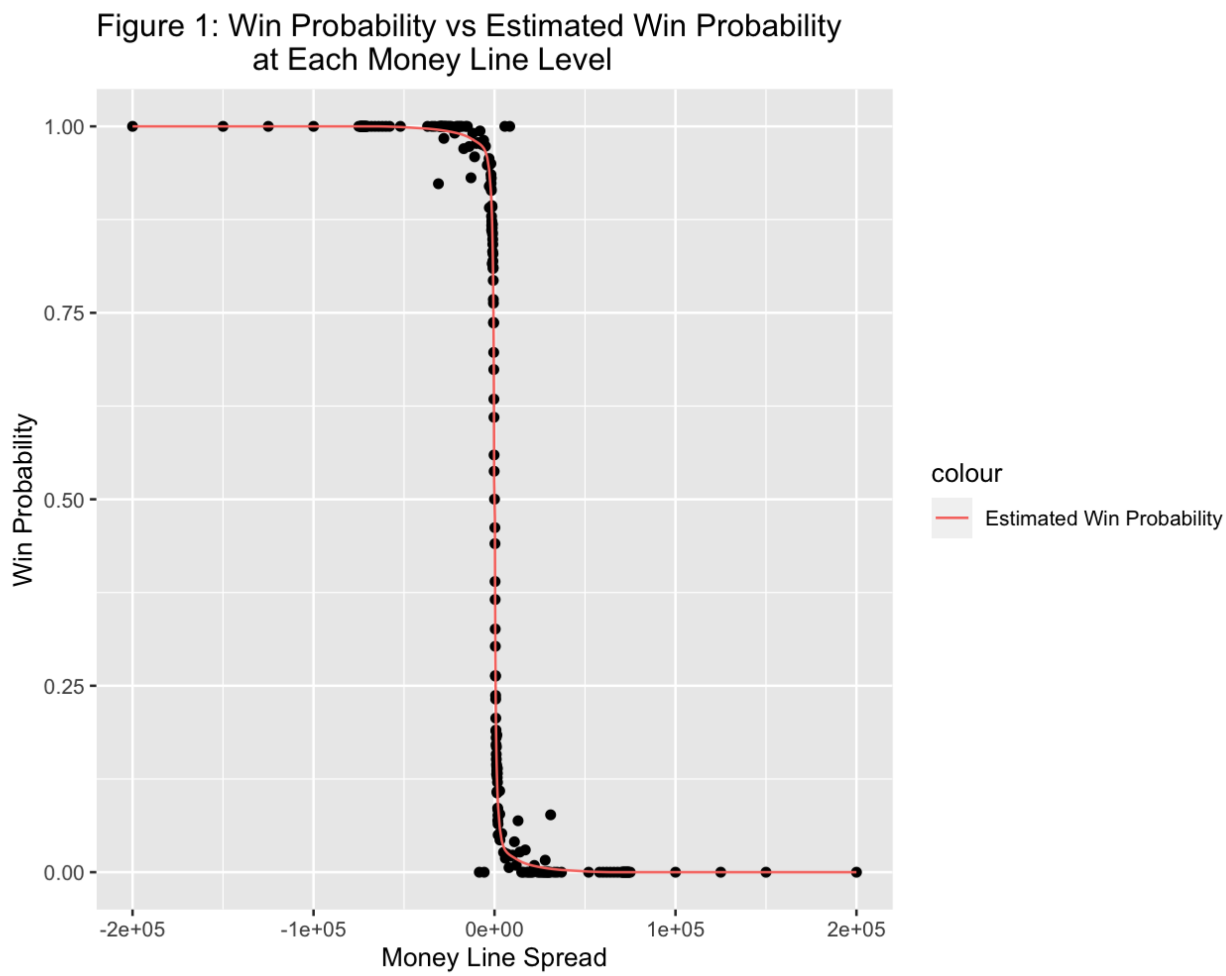

Figure 1: Win Probability vs Estimated Win Probability
at Each Money Line Level
Win Probability
1.00
0.75
0.50
0.25
0.00
-2e+05
-1e+05
0e+00
1e+05
2e+05
Money Line Spread
colour
Estimated Win Probability

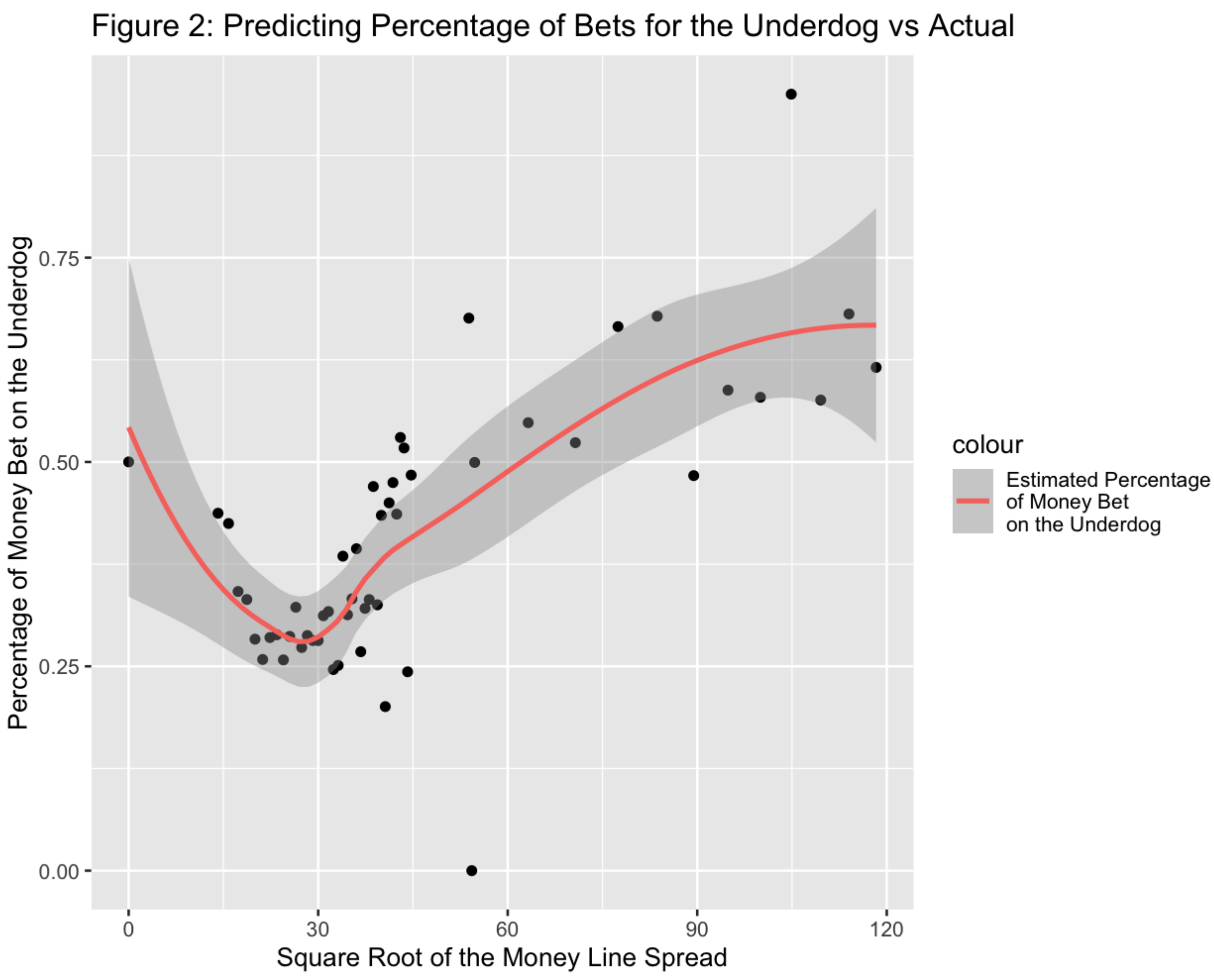
Figure 2: Predicting Percentage of Bets for the Underdog vs Actual
Percentage of Money Bet on the Underdog
0.00
0.25
0.50
0.75
0
30
60
90
120
Square Root of the Money Line Spread
colour
Estimated Percentage of Money Bet on the Underdog

Figure 3: Risk Premium vs Log Tail Risk Ratio

Sportsbooks

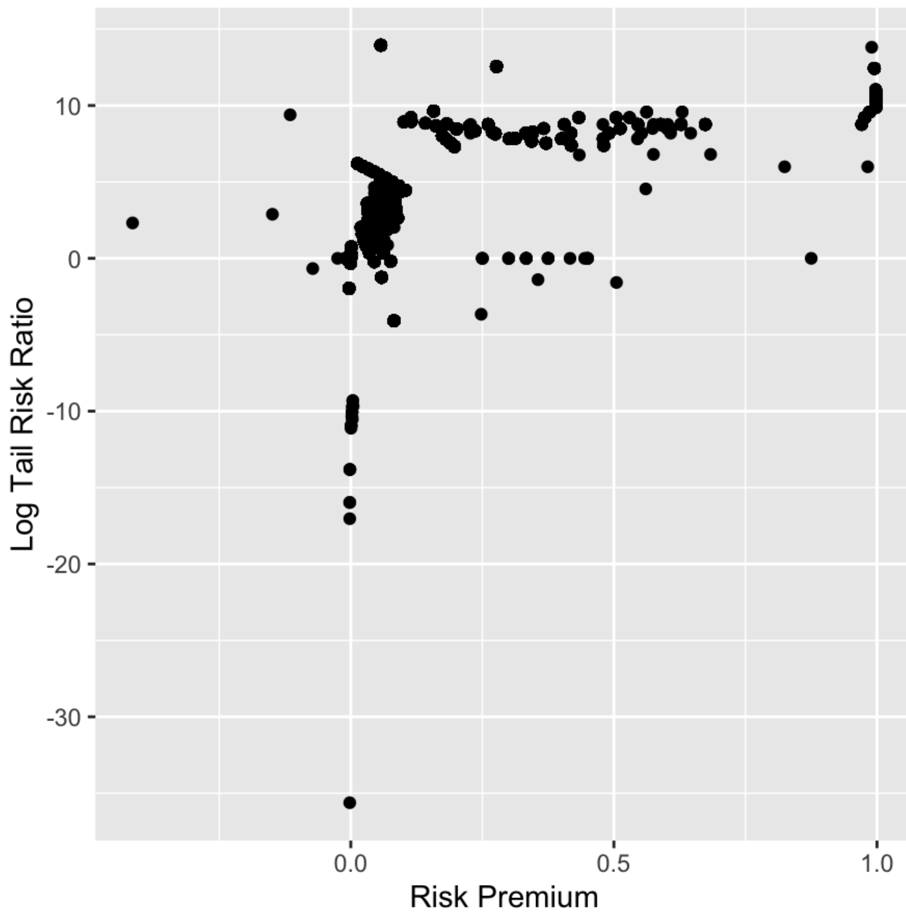

Figure 4: Risk Premium vs Log Tail Risk Ratio
Options Sellers

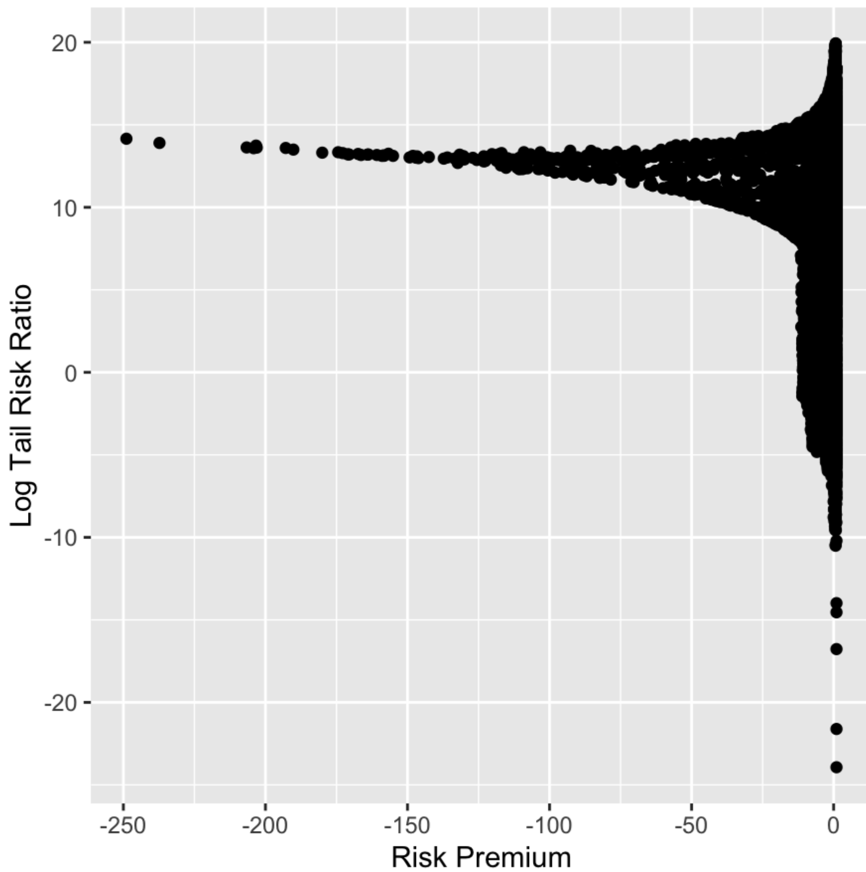

Figure 5: Histogram of Options Risk Premium

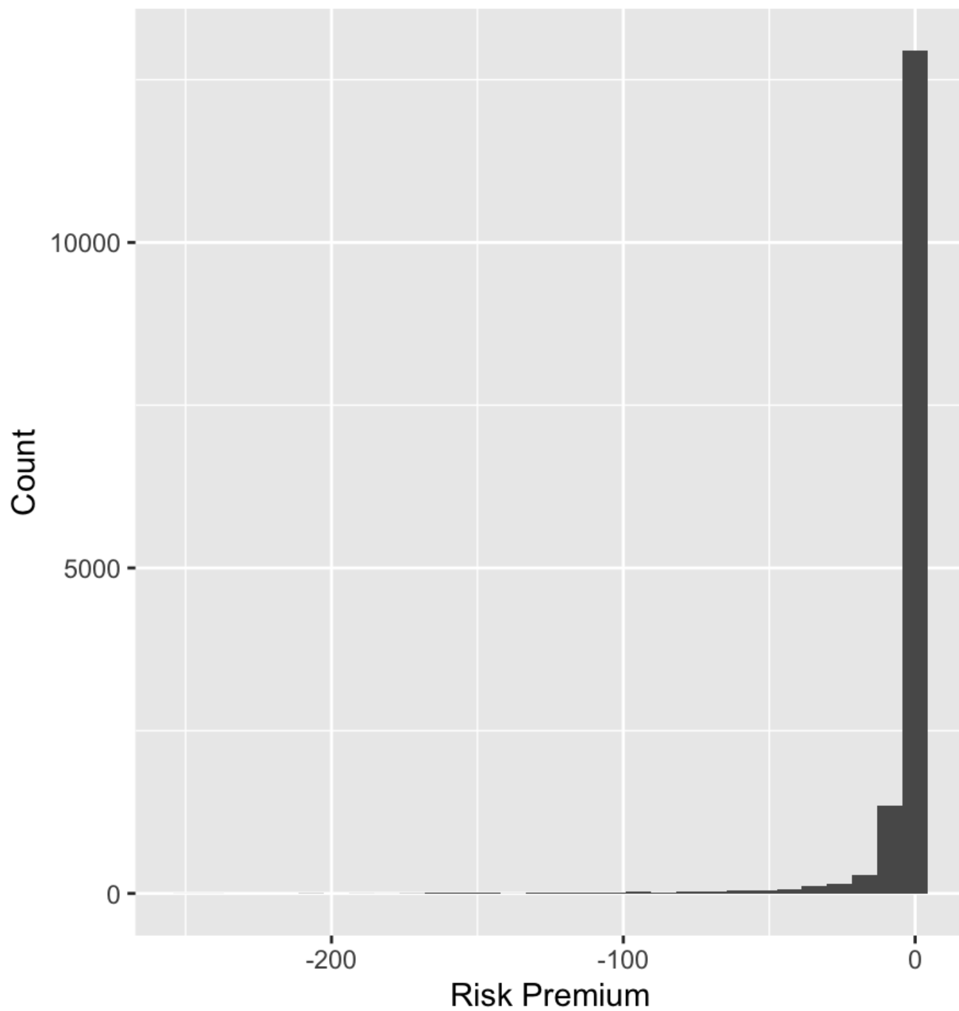

Figure 6: Histogram of Sportsbooks Risk Premiun

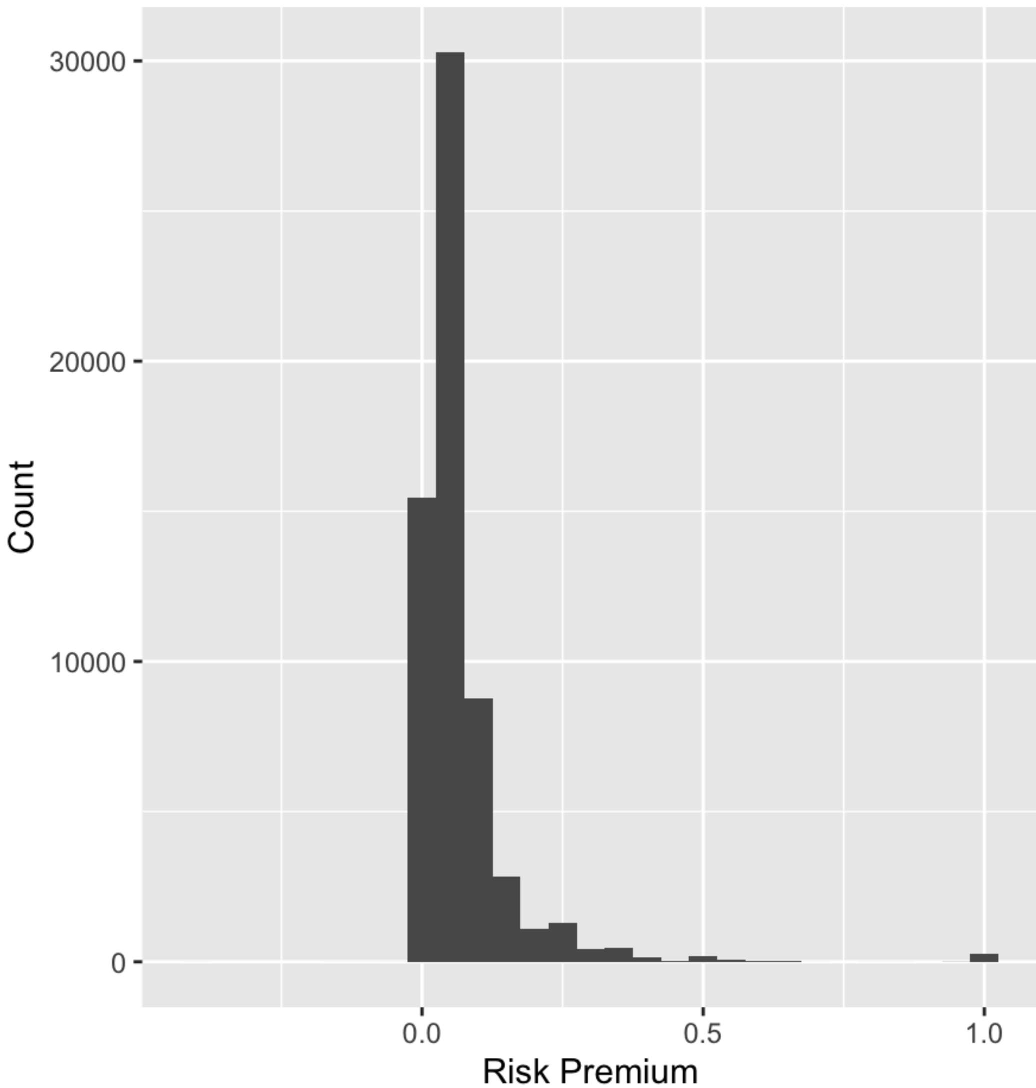

Figure 7: Histogram of Options Log Tail Risk Ratio

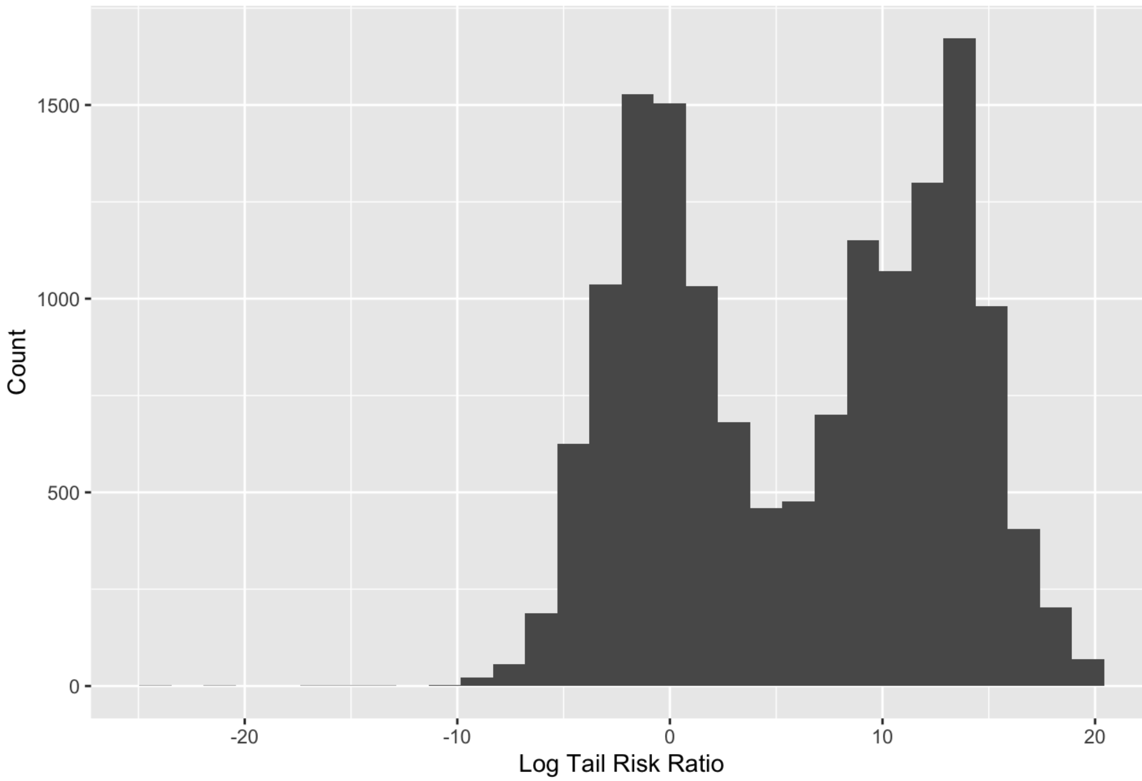

Figure 8: Histogram of Sportsbooks Log Tail Risk Ratio