Betting Without Beta

Tobias J. Moskowitz and Kaushik Vasudevan*

May 2, 2022

Abstract

Sports betting markets offer a novel test distinguishing the roles of preferences and beliefs in asset prices. Analyzing two contracts on the same outcome – one where payoff risk varies with expected outcome (Moneyline) and one where it does not (Spread) – we find that preferences for lottery-like payoffs, rather than incorrect beliefs, drive the lower returns to betting on risky underdogs versus safe favorites. Drawing parallels to low-risk anomalies in financial markets, we find the magnitude of pricing effects matches those in options and equity markets, with a model of lottery preferences providing a unifying explanation.

*Moskowitz is at the Yale School of Management, Yale University, NBER, and AQR Capital Management, email: tobias.moskowitz@yale.edu. Vasudevan is at the Yale School of Management, Yale University, email: kaushik.vasudevan@yale.edu. We thank Nick Barberis, Thomas Bonczek, Ted O'Donoghue (discussant), Stefano Giglio, Jon Ingersoll, Ed Kaplan, Bryan Kelly, Ben Matthies, Rufus Peabody, Nils Rudi, Kevin Zhao, seminar participants at Yale and AQR, and conference participants at the NBER Behavioral Finance meeting for helpful conversations, comments, and suggestions. AQR Capital is a global asset manager who may or may not use the insights and methods in this paper. The views expressed here are those of the authors and not necessarily those of AQR. A previous version of this paper was circulated under the title "What Can Betting Markets Tell Us About Investor Preferences and Beliefs? Implications for Low-Risk Anomalies."

1 Introduction

We use a novel setting, sports betting markets, to distinguish between bettor preferences and beliefs, and apply this analysis to better understand *low risk anomalies* in betting and financial markets. Low risk anomalies describe the phenomenon that low risk assets have higher (risk-adjusted) returns than high risk assets, a feature that is common to both betting and financial markets. In betting markets, this pattern is known as the Favorite-Longshot bias (FLB), and is one of the most well-studied betting phenomena. First documented by Griffith (1949), the FLB is an empirical regularity first found in horse racetrack betting, where the returns to betting on longshots (low probability of winning) are substantially lower than the returns to betting on favorites (high probability of winning). However, evidence varies across betting settings.[1] A central, unresolved debate in the literature is why the pattern emerges, with much of the debate focused on whether incorrect beliefs of outcomes or preferences for risk are driving it.

We offer a novel test to distinguish whether bettor beliefs or preferences drive the FLB. We also rectify why some betting settings find stronger evidence of a FLB than others. We relate our findings to low risk anomalies in financial markets, which share many of the same features. Sports betting markets provide an attractive research laboratory because they are particularly well suited for studying decision-making under uncertainty. As noted by Thaler and Ziemba (1988), betting contracts are contingent claims written on observable, idiosyncratic outcomes that have well-defined termination points, where outcomes are not affected by the beliefs and preferences of bettors. These features are not present in financial markets, which makes analysis of decision-making with financial asset price data challenging.

To understand our test, consider the way the FLB is typically studied − using fixed-odds contracts, where bets on the underdog are negatively skewed, offering high payoffs with low probabilities, and bets on favorites are positively skewed, offering low payoffs with high probability. An FLB in returns to these contracts can be consistent with incorrect bettor beliefs about outcomes, such as overoptimism about underdogs, or with a preference for highly skewed payoffs. Disentangling these two explanations requires either observing bettors' preferences, or measuring bettor expectations. We employ a novel approach to measuring bettor expectations that can disentangle these explanations. By examining two different contracts on the same outcome of each sporting contest – the first being the fixed-odds contract discussed above, and the second is a fixed-payoff contract that offers the same risk across contracts − we can difference out all other factors, including expectations, from risk preferences. Both contracts derive their payoffs from the *same* outcome of

[1] For example, Woodland and Woodland (1994, 2001) and Gil and Levitt (2007) find little FLB evidence in Major League Baseball (the MLB) and the National Hockey League (the NHL) bettng contracts. (Newall and Cortis (2021) review some of the empirical evidence.)

the *same* contest, but where one contract varies risk and potential payoffs, resulting in variation in the risk-return tradeoff across contests, the other holds them fixed. Assuming that both markets are integrated, which we find strong evidence of in the data, both contracts should be priced by the same expectations and beliefs among bettors. But only the fixed-odds contract can be affected by bettor preferences over higher moments of payoffs. Hence, by comparing the prices of the two contracts on the same outcome, we can tease out preference-based effects from expectations. Furthermore, because sports betting contracts reveal a terminal value at the end of each contest, we can assess the magnitude of any pricing distortions precisely.

The fixed-odds contract that varies risk and payoffs is known as the Moneyline. The fixed-payoff contract that holds risk and payoffs constant across contests is called a Spread contract. We analyze Moneyline and Spread contracts for a sample of 36,609 college and professional basketball and football games. Spread contracts are structured as bets on the expected outcome of the game.[2] These contracts are unique to betting markets and do not have a parallel in financial markets. A bet on the favorite pays off a fixed amount when the point difference exceeds the *expected* point difference, and offers zero payoff otherwise. A bet on the underdog offers the same potential payoff as betting on the favorite, except that the contract pays off when the point difference falls short of the expected point difference. Spread contract prices should reflect systematic biases about the performance of underdogs versus favorites if bettors are systematically optimistic (or pessimistic) about the performance of underdogs versus favorites. Hence, under belief-based explanations for the FLB, returns to betting on underdogs should be systematically lower (or higher) than the returns to betting on favorites for both Moneyline and Spread contracts. However, Spread contracts cannot speak to risk preference-based explanations for the FLB, because the payoffs and payoff probabilities (e.g., the amount of risk) are identical across all Spread contracts. Moneyline contracts, on the other hand, provide variation in risk across contracts, and can therefore reflect bettor preferences for risk. By comparing the pricing of Moneyline contracts to Spread contracts on the *exact same* contest, we can control for the effects of bettors' beliefs, and isolate any preference for risk.

For Moneyline contracts, we find evidence of a strong FLB, with positively skewed bets on underdogs earning much lower returns than the negatively skewed bets on favorites. However, for Spread contracts we find no evidence of such a relationship, indicating that bettors' expectations are reasonably well-calibrated for the games in our sample. The lack of pattern found in Spread contracts indicates that bettors largely have correct expectations about game outcomes, while the presence of a FLB in the Moneyline contracts indicates a preference for risk, where bettors

[2] Many studies have focused on Spread contracts in team sports to study the informational efficiency of betting markets, and how betting markets respond to the arrival of information (Zuber, Gandar and Bowers (1985), Sauer et al. (1988), Gandar et al. (1988), Camerer (1989), Brown and Sauer (1993), Golec and Tamarkin (1991), Gray and Gray (1997), Gandar et al. (1998), and Levitt (2004)).

accept lower returns for betting on highly skewed underdogs. The analysis provides a sharp test of expectations-based versus risk-based theories. Since the contracts are written on the *same* game outcomes, any non-payoff related characteristic of teams and games that may enter into bettors' decisions – sentiment, home team bias, beliefs – are common across Spread and Moneyline contracts and cannot explain these different return patterns. Risk is the only characteristic that differs across the two contract types and therefore has to be what is driving the pricing differences.

Our analysis provides a novel and powerful test of a central, unanswered question in the literature on the FLB – whether erroneous expectations or preferences for lottery-like payoffs drive the bias. We find that bettor preferences for risk play the dominant role, in contrast to the conclusions drawn from recent work on the FLB at the horse racetrack.[3] Moreover, our conclusion that lottery preferences drive the FLB may be consistent with the lack of evidence for it in other settings, such as professional baseball and hockey, where outcome probabilities are much closer to 50-50, and accordingly do not provide contracts with the extreme positive and negative skewness as in other sports (e.g., basketball) that elicit a large pricing effect under lottery preferences. We also evaluate alternative belief-based explanations. For example, bettors may correctly perceive the expected point difference but might misperceive higher moments of the point distribution. Or, bettors may have heterogeneous beliefs that naturally sort them into contracts with different skewness. We test these alternative explanations but find little evidence consistent with these other theories.

Can our results on the pricing of idiosyncratic risk in betting markets inform us about theories of risk and return in traditional financial markets? In financial markets, the low risk anomaly is a robust phenomenon that is similar to the FLB observed in betting markets. Lower risk assets earn higher risk-adjusted returns than high risk assets across a number of asset classes, including individual stocks (Black et al. (1972), Ang et al. (2006, 2009), Frazzini and Pedersen (2014), Asness et al. (2020)), options contracts (Bondarenko (2014), Ni (2008), Boyer and Vorkink (2014), Frazzini and Pedersen (2022), Baele et al. (2016)), US Treasuries, corporate bonds, equity indices, and commodities (Frazzini and Pedersen (2014)). The low risk anomaly is at odds with predictions of classical models of asset pricing, such as the Capital Asset Pricing Model, which assume that investors are risk-averse and have correct beliefs. A number of explanations have been proposed for the anomaly, including market frictions, such as constraints on borrowing and the use of leverage, omitted systematic risk factors, as well as 'behavioral' explanations that suggest investors have a

[3] For example, Snowberg and Wolfers (2010) use the prices of exotic bets on the exact order that horses will finish and find that the representative bettor incorrectly reduces compound lotteries. They interpret this result as evidence that the FLB is driven by informational inefficiency. In contrast, we find that market prices reflect the key piece of information required to determine win probabilities and price Moneyline contracts – the expected point difference – and do so in a manner inconsistent with informational inefficiencies driving the FLB. This interpretation does not speak to other potential inefficiencies in betting markets, such as the pricing of compound lotteries.

preference for assets with lottery-like payoffs or that they are too optimistic about riskier assets.[4]

Because betting markets are not exposed to systematic risks (Moskowitz (2021)), only 'behavioral' explanations can simultaneously explain the patterns in financial markets and betting markets. Assuming economic agents approach uncertainty similarly across decision contexts, our findings in betting markets may help distinguish among theories proposed for similar facts in financial markets. Specifically, our setting can help differentiate between belief-based and preference-based explanations for low risk anomalies in betting markets that may hint at similar explanations for the same pattern in financial markets. Our finding that people demonstrate non-traditional preferences for bets with lottery-like features may therefore suggest that such preferences possibly relate to low-risk anomalies found in financial markets.

To push this connection further, we offer lottery preferences as a unifying explanation for the facts in both betting and financial markets. We perform a set of exercises to quantitatively link the patterns in these two markets. First, we construct an implied volatility surface for sports betting contracts akin to the implied volatility surface used to analyze stock option prices. We estimate the "implied volatility" of a Moneyline contract as the standard deviation of the distribution of point differentials that would make the contract have the same expected return as the corresponding Spread contract. Strikingly, we observe an implied volatility smile that is qualitatively and quantitatively similar to the famous volatility smile in options markets, with contracts that are "deep in- and out-of-the-money" (very high and very low probabilities of paying off) having higher implied volatilities. We also observe that the risk-adjusted returns of call options display a quantitatively similar pattern to the pattern we find in Moneyline contracts. The similarity of results in a (much) simpler state space with no dynamics provides evidence in favor of a preference-based explanation for low risk anomalies in options markets.

Second, we calibrate a model with reference-dependent preferences that features rank-dependent probability weighting and diminishing sensitivity. We find that similar parameter values for diminishing sensitivity and probability weighting used to explain the facts in financial markets are also able to explain the FLB in betting. Probability weighting induces people to overweight tail events in their decision, and explains the low required return for betting on risky underdogs, as noted in prior work.[5] Diminishing sensitivity, where the magnitude of marginal utility is smaller for gains

[4]A prominent theory of market frictions from Black (1972) and Frazzini and Pedersen (2014), is that excess borrowing costs drive investors seeking higher returns to increase the weight of risky assets in their portfolio, rather than holding an optimally leveraged position in the unconstrained efficient portfolio. In equilibrium, this behavior lowers the return on systematically riskier assets. Ang et al. (2009) and Gormsen and Lazarus (2021) suggest that low-risk anomalies may be explained by exposure to non-diversifiable risk factors not captured by the Capital Asset Pricing Model. Brunnermeier, Gollier and Parker (2007) and Barberis and Huang (2008) are prominent examples of models where investors have a preference for idiosyncratically skewed assets.

[5]Previous work that emphasizes the potential role of probability weighting in explaining the FLB includes Jullien

and losses that are further from the investor's reference point, explains why agents choose to wager on negatively skewed, negative expected return contracts on favorites. The potential importance of diminishing sensitivity to explain bettor behavior is a new insight for the FLB, and a novel economic setting where diminishing sensitivity (a relatively less studied component of reference-dependent preferences) may be applicable.[6] Relative to previous work studying betting behavior, our specification also endogenizes the choice to bet and the amount that bettors choose to wager. Using unique data on betting volume, we are able to test these additional implications of theory and find that the calibrated model is also able to explain variation in betting volume across contracts.

While we hope to draw parallels between our findings in betting markets and similar risk-return patterns in financial markets, we recognize this link is speculative and relies on these markets not being totally detached from each other. There are reasons to be both cautious and optimistic in connecting these two markets. On the one hand, there are institutional differences between betting markets and financial markets. However, the two markets share a number of common features: large transaction volume, widely available information, market making activity, arbitrage activity (from professional bettors and hedge funds), and professional analysts. Also, preferences related to entertainment are secondary motives in betting and are not absent in financial markets either (where they are also secondary in the stock market, see e.g., Dorn and Sengmueller (2009) and Grinblatt and Keloharju (2009)).[7] A previous generation of work recognizes the connection between betting and financial markets, primarily focusing on the informational efficiency of betting markets and its implications for theories in financial markets (see e.g., Pankoff (1968), Durham, Hertzel and Martin (2005), and Moskowitz (2021)). Finally, although the strength of the link between the two markets can be debated, it is interesting that preferences calibrated to match the betting market patterns can also match the magnitudes of low risk anomalies in financial markets. Hence, lottery preferences may play a significant role in providing a unifying explanation for the facts across different settings.

The rest of the paper proceeds as follows. Section 2 describes the betting markets and our data in more detail. Section 3 studies patterns in Moneyline and Spread contract returns, and analyzes potential explanations for the risk-return relations. Section 4 presents a quantitative comparison of

---

and Salanié (2000) and Snowberg and Wolfers (2010). Studies of the potential role of probability weighting in financial markets include Kliger and Levy (2009) and Baele et al. (2016). Barberis, Huang and Santos (2001), Barberis and Huang (2008), Barberis, Mukherjee and Wang (2016), and Barberis, Jin and Wang (2021) focus on the ability of probability weighting, and Cumulative Prospect Theory more generally, to explain stock price behavior.

[6] In their review chapter on reference-dependent preferences, O'Donoghue and Sprenger (2018) note that "Most applications of reference-dependent preferences...ignore the possibility of diminishing sensitivity" and further note that "the literature still needs to develop a better sense of when diminishing sensitivity is important."

[7] See Peta (2014) for a discussion of the industry of professional gambling and the use of financial tools from Wall Street in the sports betting market.

the patterns in betting contract returns with patterns in equity and options returns from financial markets. Section 5 presents a calibration of preferences to explain the evidence in both betting and financial markets. Section 6 concludes.

2 Empirical Setting and Data

Sports betting markets are large, liquid, and active. According to Statista.com, global sports betting markets produced an aggregate gross gaming yield (notional bets taken by betting operators minus winnings/prizes paid out) of nearly $200 billion in 2017 and 50 percent of U.S. adults have made a sports bet (which is higher than stock market participation rates, see Vissing-Jørgensen (2002)). In the U.S., the American Gaming Association estimates that 4 to 5 billion U.S. dollars are wagered legally each year at Nevada sportsbooks, the only state where it was legal, but the amount bet illegally with local bookies, offshore operators, and other enterprises is roughly 30 times that figure.[8] With the recent U.S. Supreme Court decision overturning the Professional and Amateur Sports Protection Act of 1992 that prohibited state-sanctioned sports betting, the expectation is that this market will grow considerably.[9]

We study bets placed on four types of games: National Collegiate Athletic Association (NCAA) Football games, NCAA Basketball games, National Basketball Association (NBA) professional games, and National Football League (NFL) professional games. For the primary betting contracts we study, a bookmaker controls the "lines" (prices). At the start of betting, bookmakers set an opening line, which is determined by pre-betting market activity from large professional bettors, after which bettors may place bets until the start of the game. Betting volume flows can change the price of bets if bookmakers seek to balance the money being bet on each side. On average, bookmakers generally set lines such that they take no risk, though this may not be true in every game. Without taking risk, they profit from the 'vig,' which refers to the transaction cost implicitly embedded in contract prices that makes betting a negative expected return proposition for bettors.[10] Bettors receive the price at the time they make their bet, even if the line later changes. Betting closes before the start of the game. For some contracts (e.g., the NFL), the time between open and close can be six days, while for others (e.g., the NBA), it may only be a few hours.[11] See Moskowitz

[8] According to the 1999 Gambling Impact Study, an estimated $80 billion to $380 billion was illegally bet each year on sporting events in the U.S., dwarfing the $2.5 billion legally bet each year in Nevada (Weinberg (2003)).

[9] Hudson (2014) shows that in the UK, where sports betting is legal, betting has increased annually by about 7%, fueled by online and mobile betting.

[10] Bookmakers may occasionally set prices between the 'informationally efficient' price and the price that balances demand. See Levitt (2004).

[11] This style of betting is different from parimutuel pools, which are the common way in which betting at the horse racetrack is organized in North America (horse betting in the UK is organized with bookmakers, similar to our setting). In a parimutuel pool, prices and odds are not set at the beginning of betting. Rather, each bettor

(2021) for more details on line setting in these markets.

Our primary focus is on two types of betting contracts: Moneyline and Spread contracts. The payoffs to each contract are determined by the difference in points scored by each team in the game, $P_A - P_B$, where P_A and P_B are the number of points scored by teams A and B. Moneyline contracts are outright bets on which team will win the game (i.e., a bet on $P_A - P_B \lessgtr 0$). Spread contracts are bets on whether the point-differential between the two teams exceeds the "Spread Line" for the game (i.e., a bet on $P_A - P_B \lessgtr \bar{x}$, where $\bar{x}$ is the Spread). In an auxiliary test, we also analyze Over/Under contracts. Over/Under contracts are bets on whether the total number of points scored in a game will exceed the line total, T, set by the bookmaker (i.e., a bet on $P_A + P_B \lessgtr T$). We describe the mechanics and pricing of the three contract types in more detail below. For all of our analyses, we use the prices at the close of betting and assume that all bettors transact at the closing price. In the appendix, we verify that the same pricing patterns persist using opening prices.

2.1 Moneyline Contract

The Moneyline (ML, also known as American Odds) contract is a fixed-odds contract that is a bet on which team wins. The Moneyline contract offers different potential payoffs per dollar wagered on a team depending upon which team is bet in a game. Larger potential payoffs are offered for betting on underdogs (that are less likely to win) and smaller potential payoffs are offered for betting on favorites (that are more likely to win). For example, if a bet of \$100 on Chicago (the favored team) over New York is listed as −165, then the bettor risks \$165 to win \$100 if Chicago wins. Betting on New York (the underdog) the Moneyline might be +155, which means risking \$100 to win \$155 if New York wins. The \$10 difference is commission paid to the sportsbook. The payoffs for a \$100 bet on team A over team B on a Moneyline contract listed at $-\$M$ are as follows:

$$\text{Payoff}^{ML} = \begin{cases} |M| + 100, & \text{if } (P_A - P_B) > 0 \quad (\text{"win"}) \\ \text{Max}(M, 100), & \text{if } (P_A - P_B) = 0 \quad (\text{"tie"}) \\ 0, & \text{if } (P_A - P_B) < 0 \quad (\text{"lose"}) \end{cases} \tag{1}$$

where M is either > 100 or < -100 depending on whether team A or B is favored to win.

For readers familiar with fixed-odds contracts in other settings, Moneylines can be directly converted to both fractional odds, as quoted in the UK, and decimal odds, as quoted in continental

specifies the amount of money they wish to bet, without knowing prices or odds. The prices are only set at the close of betting, and are set such that winning bets are paid out in proportion to the stakes wagered from the losing bets. The literature has studied horse betting both in the parimutuel and the bookmaker context and found similar patterns in the FLB for both.

Europe. Positive Moneylines quote the money to be won for a \$100 wager, so for example, an ML of +400 would be quoted as 4/1 in fractional odds and as 5 in decimal odds. Negative Moneylines quote the amount of money to be wagered to win \$100, so for example, an ML of −400 would be quoted as 1/4 in fractional odds and as 1.25 in decimal odds.

2.2 Spread Contract

The Spread (S) contract is a bet on a team winning by at least a certain number of points known as the "spread." For example, if Chicago is a 3.5 point favorite over New York, the spread is quoted as −3.5, which means that Chicago must win by four points or more for a bet on Chicago to pay off. The spread for betting on New York would be quoted as +3.5, meaning that New York must either win or lose by less than four points in order for the bet to pay off. Unlike Moneyline contracts, Spread contracts offer the same potential payoff for betting on either team. In the next section, we show empirically that Spreads are set to make betting on either team roughly a 50-50 proposition or to balance the total amount bet on each team. The typical bet is \$110 to win \$100. So, the payoffs for a \$110 bet on team A over team B on a spread contract of $\bar{x}$ points are:

$$\text{Payoff}^S = \begin{cases} 210, & \text{if } (P_A - P_B) > \bar{x} \quad (\text{"cover"}) \\ 110, & \text{if } (P_A - P_B) = \bar{x} \quad (\text{"push"}) \\ 0, & \text{if } (P_A - P_B) < \bar{x} \quad (\text{"fail"}) \end{cases} \tag{2}$$

where "cover, push, and fail" are terms used to define winning the bet, a tie, and losing the bet, respectively. For half-point spreads, ties are impossible since teams can only score in full point increments.

2.3 Over/Under Contract

Finally, the Over/Under contract (O/U), is a contingent claim on the total number of points scored ($y = P_A + P_B$). Sportsbooks set a "total", T, which is the predicted total number of points the teams will score combined. Bets are placed on whether the actual outcome of the game will fall "over" or "under" T. The payoffs are similar to the Spread contract in that a bet is for \$110 to win \$100. For example, wagering on the over contract in a game earns the following payoffs:

$$\text{Payoff}^{OU} = \begin{cases} 210, & \text{if } (P_A + P_B) > T \quad (\text{"over"}) \\ 110, & \text{if } (P_A + P_B) = T \quad (\text{"push"}) \\ 0, & \text{if } (P_A + P_B) < T \quad (\text{"under"}) \end{cases} \tag{3}$$

Bookmakers set lines such that there is a 50-50 chance of either side of the contract paying off, or the dollar volume on both sides of the contract is approximately 50-50.

2.4 Betting Contract Risk

Betting on spread contracts (on either side) amounts to taking a gamble with (approximately) 50% probability of paying off; there is no cross-sectional dispersion in risk (measured as idiosyncratic variance or skewness) in Spread contracts. Betting on Moneyline contracts on the favorite amounts to taking a gamble with a greater than 50% probability of paying off, while betting on the underdog amounts to taking a gamble with less than 50% probability of paying off. Per dollar wagered, the lines for each contract are set such that the potential payoff for a winning bet is decreasing in the probability of the contract paying off. That is, bets on more extreme underdogs are riskier, and bets on more extreme favorites are less risky. There is substantial cross-sectional heterogeneity in risk across Moneyline contracts, based on the relative quality of the teams playing.

2.5 Data

The data come from SportsInsights.com beginning in 2005 and ending in May 2013 for all four sports we analyze. There are 36,609 total games in the sample: 21,982 NCAA Basketball games, 8,392 NBA Basketball games, 4,392 NCAA Football games and 1,843 NFL Football games. Each game contains two betting contracts – the Moneyline and Spread contract – based on the point difference between the two teams, and a third contract based on the total points scored by both teams combined (Over/Under contract).

The betting lines are drawn from the Las Vegas legalized sportsbooks and online betting sportsbooks, where all bookmakers offer nearly identical closing lines on a given game. The data include all games from the regular season and playoffs/post-season. The data for all games include the team names, start and end time of game, final score, and the opening and closing betting lines across all contracts on each game. The betting lines are taken from the Las Vegas legalized sportsbooks and online sportsbooks and are transactable quotes. For the majority of contests, all bookmakers offer nearly identical lines in the database. On the rare occasion when lines differ (less than 1% of the time), that contract is removed from the sample. Results are robust to using the highest or lowest line, or an average of the lines when there is a discrepancy. In addition, the data also include information on the proportion of bets placed on the two teams in each contract from three sportsbooks: Pinnacle, 5Dimes, and BetCRIS. These three sportsbooks are collectively considered the "market setting" sportsbooks that dictate pricing in the U.S. market.[12]

[12] In sports betting parlance, market setting means that if one of the three big sportsbooks moves their line, other sportsbooks will follow, "moving on air," even without taking any significant bets on the game.

The data offer some unique advantages relative to previous studies on sports betting. The data are more comprehensive, covering multiple sports over a decade (most studies cover a single sport over a few years), and uniquely provide multiple contracts on the same outcome of the same game – namely Moneyline and Spread contracts, which are key to our identification strategy to isolate the relevance of risk to betting prices. In addition, the data contain the actual betting lines/prices to compute realized returns to betting. Many studies (Gandar et al. (1988), Avery and Chevalier (1999), Levitt (2004)) use newspaper-sourced, composite, or even survey-based betting lines that are not real transaction prices. Using real returns from transactable prices allows us to quantify economic magnitudes and compare them to real returns in financial markets related to the same phenomenon.

3 Identifying Expectations versus Preferences

We begin with a simple conceptual framework to set up our tests for the presence of a FLB in Spread and Moneyline contracts. The basic insight is that Spread contracts allow us to extract bettor expectations of game outcomes. Patterns in Moneyline contract returns, however, may be driven by both bettor expectations of game outcomes and preferences for risk. Hence, by evaluating patterns in Moneyline and Spread contract returns simultaneously, we can disentangle preferences for risk from belief-based explanations. Using this framework, we examine the cross-section of Spread and Moneyline contract returns. Finally, we consider alternative explanations for the results outside of this framework.

3.1 Conceptual Framework

Consider the point differential of a contest between two teams A and B, $P_A - P_B$. Assume that the point differential follows a symmetric probability distribution, $f(\cdot;\mu,\sigma)$, where μ is expected point differential, and σ are the other parameters required to characterize the distribution (e.g., the standard deviation if f is ergodic), which are assumed to be known. Bettors perceive the expected point differential as $\hat{\mu}$.

Assuming bookmakers take no risk, the point spread is set at $\hat{\mu}$. From equation (2), bets on the favorite offer a payoff when $P_A - P_B > \hat{\mu}$, and bets on the underdog offer a payoff when $P_A - P_B < \hat{\mu}$. Patterns in Spread contract returns reveal the presence of any FLB driven by systematic errors in expectations (i.e., $\hat{\mu} \neq \mu$). When $\hat{\mu} = \mu$ (bettors' expectations correspond with the true expected outcome), then the returns of favorites and underdogs will be equal on average. If $\hat{\mu} > \mu$, then the line is biased towards the favorite, and betting on the favorite will offer lower average returns than betting on the underdog. The opposite is true if $\hat{\mu} < \mu$. Hence, evaluating patterns in Spread contract returns presents a test of *expectations*-based theories of any FLB. In addition, preferences for home team, or certain types of teams, etc. can show up as $\hat{\mu} \neq \mu$ and can be interpreted as biased beliefs. However, because risk is constant across Spread contracts, any preference for risk cannot affect $\hat{\mu}$. So, Spread contracts rule out any risk-based price distortions.

For the Moneyline, bets on the favorite pay off when $P_A - P_B > 0$, which occurs with probability $p \equiv \int_0^\infty f(x;\mu,\sigma)dx > 0.5$, which may be perceived by bettors as $\hat{p} \equiv \int_0^\infty f(x;\hat{\mu},\sigma)dx > 0.5$. Bets on the underdog payoff with probability $1-p$ (perceived as $1-\hat{p}$). Denoting the potential profits associated with a \$1 wager (the fractional odds) on the favorite and underdog as y_f and y_u, the expected returns for the favorite and underdog contracts under objective probabilities are $ER_f = py_f - (1-p)$ and $ER_u = (1-p)y_u - p$, while bettors' subjectively perceived expected returns are computed by replacing p with $\hat{p}$. If bettors correctly perceive the expected outcome ($\hat{\mu} = \mu$) then any pattern in returns must be driven by *preferences*. If $\hat{\mu} = \mu$, and bettors are (weakly) risk-averse, then we expect $ER_u \geq ER_f$, where risky bets on underdogs should earn returns at least as high as the returns on favorites. If, however, $ER_u < ER_f$ (a FLB in Moneyline contracts) then this would indicate that bettors are not risk-averse, and have a preference for risk (e.g., lottery-like payoffs) offered on underdog contracts that leads them to accept lower returns.

Because the one difference between the Moneyline contract and the Spread contract on the same games is that the Moneyline varies risk across games, which the Spread contract does not, any pricing pattern differences between the Moneyline and Spread must come from risk exposure. Other preferences or erroneous expectations should affect both the Moneyline and Spread contracts. By comparing the two contracts on the same games, we can difference out all other aspects and isolate the role of risk in setting prices, assuming the same set of investors participate in both markets, an assumption we test later in the paper that finds strong support in the data. If there is no pattern in Spread contract returns, then under the assumptions of our framework, any FLB pattern in Moneyline contract returns must be attributable to risk preferences and not expectations or other preferences.

3.2 Testing the Favorite-Longshot Bias

Following the literature, we identify favorites and longshots using the Moneyline contract payoffs, based on the simple assumption that higher potential payoffs are required to entice bettors to wager on lower probability events. Teams with higher payoffs from winning are longshots and teams with lower payoffs from winning are favorites.

Methodology

To test for the FLB, we sort betting contracts (both Spread and Moneyline contracts) into deciles based on the Moneyline posted for each game. The Moneyline is a function of the probability of a team winning *and* any pricing distortions due to erroneous beliefs or preferences. An alternative method for identifying favorites and longshots would be to use the Point Spread quoted on each game, which also reflects the probability of winning and any pricing distortions. Sorting contracts within each sport based on the point spread yields nearly identical results for reasons we discuss in more detail below.

Table 1 reports summary statistics of the deciles, with decile 1 containing the biggest underdogs and decile 10 the biggest favorites. The first row of Table 1 reports the estimated win probability of each decile based on fitted values from non-parametric estimates of team win probabilities using the Spread and Moneyline for each game. The details of these estimates are provided in Appendix A. As Table 1 reports, the teams in decile 1 (biggest underdogs) have an estimated 0.09 probability of winning, while the teams in decile 10 (biggest favorites) have an estimated 0.91 probability of winning, with the probability of winning steadily increasing across deciles. The second row of Table 1 reports the actual average win percentages of each decile (frequence of actual wins), which match the estimated probabilities almost perfectly, and increase from 0.09 (decile 1) to 0.91 (decile 10) in approximately 0.10 increments across the deciles.[13]

We examine the risks and returns of the contracts across the deciles to identify patterns associated with long shots versus favorites. The first panel of Table 1 reports results for the Moneyline contracts and the second panel for Spread contracts.

Moneyline Contracts

The first panel of Table 1 reports the average Moneyline across the deciles, along with the average return to each decile and two measures of risk for each decile: standard deviation and skewness of returns. For standard deviation, we compute the standard deviation of realized returns on the decile of bets, as well as an ex ante measure of standard deviation based on payoffs and payoff probabilities. We report the mean ex ante standard deviation of each decile. The realized standard deviation as well as the ex ante standard deviation show the same pattern: risk decreases monotonically across the deciles, ranging from 2.64% for decile 1 (underdogs) to 0.31% for decile 10 (favorites). In a similar fashion, the next two rows report the realized skewness of returns and the average ex ante skewness of returns within each decile. Skewness significantly decreases monotonically across

[13]The actual probabilities matching the estimated probabilities almost perfectly is examined further in the next subsection and is consistent with Moskowitz (2021).

the deciles, ranging from 4.51 for extreme underdogs to -2.84 for extreme favorites. These are big differences in risk across Moneyline contracts that are strongly tied to the probability of the team winning. F-tests easily reject equal standard deviation or skewness across the deciles. This pattern is by design and is how Moneyline contracts construct their payoffs based on fixed odds. It is this feature of fixed odds contracts we exploit in comparison to Spread contracts that allows us to isolate the role of risk versus expectations in pricing these contracts.

The last row of the first panel reports the average return to each decile, which is monotonically decreasing in risk. More (less) risky contracts earn significantly lower (higher) average returns than less (more) risky contracts. The biggest underdogs (riskiest contracts) in decile 1 have average returns of -22.60%. Contracts in decile 10 (safest contracts) earn an average return of -0.51%.[14] We obtain nearly identical results by sorting on the Spread line rather than Moneyline, since the lines have a nearly perfectly monotonic relationship within each sport. We discuss this fact further in the next subsection. An F-test easily rejects that the returns are equal. The results show a FLB in Moneyline contracts.

Spread Contracts

The second panel of Table 1 reports the results for Spread contracts (which are sorted by the Moneyline values). The Spread ranges from +14.64 for the 10% most extreme underdogs, and -14.93 for the most favored 10%. By construction, the Spread contracts have ex ante constant volatility across games and thus the deciles all have an ex ante standard deviation of 95%. The actual standard deviation of returns to each decile match this almost perfectly with a nearly constant 94% to 95% standard deviation. The ex ante skewness of Spread contracts is zero by construction, and the actual skewness of returns to the deciles is essentially zero across the deciles (ranging from -0.04 to +0.04 with no discernable pattern across the deciles). By all measures there are no differences in risk across the Spread contract deciles. Hence, the Spread contracts provide no variation in risk, but are written on the same games with the same underlying win probability across deciles. However, the average returns of the contracts do *not* appear to vary with the likelihood that the favorite wins the game. The returns in the deciles range from -2.77% to -6.32%, with no reliable pattern and the average returns are not statistically different across the deciles, with an F-stat of 0.97 and a p-value of 0.46 that fails to reject that the decile returns are all equal.[15]

Figure 1 summarizes these results by plotting the returns to the Moneyline and Spread contracts by decile, as well as the standard deviation and skewness by decile. The plot highlights how risk

[14] The vig or transaction cost in betting markets is 5% and is substantially punitive that average returns are negative.

[15] To the extent there is a pattern, it is in the opposite direction of the FLB, and can be resolved by comparing bets made on home teams across games and bets made on away teams across games. See Appendix Figure D.8.

changes across deciles for the Moneyline contract but not for the Spread contract, and shows how returns vary with risk as average returns are decreasing with risk as exhibited by the Moneyline contract, but are constant across Spread contracts that exhibit no risk differences. Since Moneyline and Spread contracts are written on the same game and outcome, the difference in return patterns between them must be driven by risk, since all else is equal between the two contract types.

3.3 Interpreting the Empirical Results

The empirical evidence indicates that favorites versus longshots are not 'mispriced' because the market has a biased assessment about game outcomes. Rather, our results show that market prices are accurate in forecasting the expected difference in points scored in the game, as evidenced by the lack of a FLB in Spread contract returns. Spread contract prices are not systematically optimistic about underdogs and pessimistic about favorites. Spreads seem to correctly predict the expected outcome of the game. Hence, the strong FLB exihibited for the Moneyline contract is unlikely driven by erroneous expectations, but seems to be driven by the riskiness of the payoffs, which is consistent with a preference for risk. High payoffs with low probability appear to be overpriced relative to low payoffs with high probability. This pattern looks like a FLB, but appears to be driven by a preference for lottery-like payoffs associated with long shots rather than false expectations about the expected return.

Similarly, the results also reject non-monetary preferences for certain characteristics of the betting contracts, such as a preference for the home team or any specific team (big market, recent performance, etc.) driving the FLB, since such preferences should also be reflected in Spread contracts, where we do not see any pattern. Because we compare the returns of two different contracts on the *same* outcome of the *same* game, everything about the contracts remains equal except for risk. Hence, the only explanation for the return differences across the two contracts has to be driven by the risk of payoffs.

The fact that our results point to risk is notable. A central, unresolved debate on the FLB is whether the bias is due to the market incorrectly assessing the outcomes of sporting events, or whether it is a reflection of preferences for lottery-like payoffs.[16] Recent work has come down on both sides of the debate. Snowberg and Wolfers (2010) use the prices of exotic bets on the order in which horses finish as evidence that the market does not properly reduce compound lotteries, representing a misperception of horse win probabilities. Taking a non-parametric approach, Chiappori et al. (2019) find evidence that preferences and probability misperceptions both play a role in the FLB in horse racing. Our approach – evaluating how the market evaluates game outcomes in the Spread market, where risk does not vary across contracts – presents a novel way to identify the

[16]See Thaler and Ziemba (1988) for a discussion of various explanations for the FLB.

market's expectations of the game outcome, and suggests that the FLB is likely driven by bettor preferences. We interpret this result as bettors having a taste for lottery-like payoffs. We turn next, however, to potential alternative explanations of the fact that risk differences explain returns.

3.4 Alternative Interpretations

We evaluate alternative explanations that may be able to explain the existence of a FLB in Moneyline contracts but the absence of such a pattern in Spread contracts. One alternative is that Spread and Moneyline contracts are informationally segmented. Second, bettors may have correct beliefs about the *expected* point difference of games but incorrect beliefs about higher moments of the point spread distribution. Finally, bettors may have heterogeneous beliefs that could result in a FLB in the Moneyline but not the Spread contract. We discuss these alternatives below, but find little support for them in the data and conclude that they are unlikely to be the primary drivers of the risk-return patterns we find.

3.4.1 Market Segmentation

One alternative explanation for our results is that the Moneyline and Spread markets are informationally segmented, in which case pricing in the Spread market says nothing about the Moneyline market. However, we show that Spread contract lines span nearly all relevant information regarding win probabilities in Moneyline contracts, suggesting these two markets are highly integrated. This result supports our claim that expected point differences are the key piece of information determining win probabilities. More formally, we run the following regression,

$$\mathbb{I}_{\text{home win},i} = \alpha_s + \beta_{\text{Spread,s}}\text{SpreadLine}_i + \beta_{\text{ML,s}} \log\left(\frac{1+y_{h,i}}{1+y_{a,i}}\right) + \epsilon_i \tag{4}$$

where for game i, $\mathbb{I}_{\text{home win},i}$ is an indicator equal to one if the home team wins the game (and zero otherwise), SpreadLine_i is the Spread line for game i, and $y_{h,i}$ and $y_{a,i}$ are the payoffs per dollar wagered associated with winning bets on the home and away teams in the game. The quantity $\log\left(\frac{1+y_{h,i}}{1+y_{a,i}}\right)$ is termed "the Moneyline ratio" and captures how favored the home team is versus the away team as implied by the Moneyline, with smaller values corresponding to the home team being more favored.[17] $\beta_{Spread,s}$ and $\beta_{ML,s}$ are separate regression coefficients for each sport, s. We standardize the independent variables within sports to have zero mean and unit standard deviation for ease of interpretation.[18]

[17]For readers familiar with decimal odds, the Moneyline ratio is simply the log of the ratio of the decimal odds on the two teams. The Moneyline ratio is 0.98 correlated with the Spread line across games in our sample.

[18]We estimate regression coefficients separately by sport because the point spread distributions differ across sports. Additionally, the transformation that we apply to Moneyline contracts is done to make the independent variable more

We estimate three versions of the regression: 1) including only the Spread line as an independent variable, 2) including only the Moneyline ratio as an independent variable, and 3) including both. Panel A of Table 3 reports the results from the regressions. The first two columns show that the Spread and Moneylines have extremely similar return predictability for wins, with nearly identical (and highly significant) regression coefficients (ranging from -0.18 for the NFL to -0.26 for NCAA Football, indicating that a one-standard deviation change in the independent variable corresponds with an 18 to 26 percent increase in win probability for the home team). The third column shows the results from the multivariate regression that includes both the Spread and Moneyline. The R^2 for the multivariate regression is 20.87%, indicating that the multivariate regression adds little explanatory power over the univariate regressions, and that the Money lines and Spread lines capture the same predictive information for wins. In fact, the Moneyline and Spread are so correlated that including both in the regression creates a multicollinearity problem, which is why the significance of the coefficients declines sharply. The F-statistic comparing the multivariate regression with the univariate Spread regression is 0.04, which fails to reject the univariate regression in favor of the multivariate regression, consistent with the Spread line and Moneyline capturing the same information for wins.

We also plot a binned scatterplot of the Moneyline ratio versus home win percentages in Appendix Figure D.1, both including and excluding a control for the Spread Line. The figure shows that win probabilities are monotonically decreasing in the Moneyline, but the second figure shows that the Spread Line captures almost all of the Moneyline's explanatory power. It is also apparent that alternative functional forms for the relationship between win probability and Moneyline-implied probability are unlikely to explain the results.

Second, we show that changes in the Spread line and the Moneyline ratio from the open to the close of betting are highly correlated. In Panel B of Table 3, we show regression results of the open-to-close *changes* in the Moneyline ratio on the open-to-close *changes* in the Spread line for each sport. The regression coefficients range from 0.53 (for NCAA Football) to 0.63 (for the NBA), with t-stats ranging from 34.08 to 114.49. Betting lines move because either volume demand on one side of the contract exceeds demand to take the other side, or information about game outcomes arrives between the open and close of betting (e.g., an injury to a key player). The results in Panel B of Table 3 show substantial commonality in Moneyline and Spread line responses to the arrival of information or volume demand, indicating that these markets are highly integrated and not segmented.

The near complete degree of overlap in the information captured by Spread lines and Moneylines regarding win probabilities, and the high correlation between changes in the Moneyline ratio and

normal. Alternative formulations have *worse* predictive power for wins.

changes in Spread line across games from open to close of betting, are inconsistent with these markets being segmented. It is also the case that most bookmakers set lines/prices in the Spread and Moneyline markets simultaneously, and there is an active arbitrage market between them.

3.4.2 Misperceptions of Higher Moments of the Point Difference Distribution

Another alternative explanation for our results is that while bettors do not seem to misperceive the *expected* point difference, perhaps they misperceive the higher moments of the point-difference distribution. Under our conceptual framework, perhaps bettors do not know the true distribution of points, f, or incorrectly perceive moments of the point distribution other than the mean. In particular, our results could be consistent with bettors believing that the point-difference distribution is more dispersed than it actually is. This misperception would lead bettors to erroneously believe that the probability of a favorite winning is lower, and the probability of an underdog winning is higher than reality. Bettors with such mistaken beliefs will find bets on underdogs more attractive than bets on favorites, in a manner consistent with the FLB.

To test this theory, we start by examining the distribution of point differences across games. Does the point-difference distribution differ for lopsided versus evenly-matched contests? Figure D.2 plots the standard deviation of the point-differential minus the Spread (expectation of point difference) across all games, sorted into deciles by the Moneyline ratio in each sport (from greatest underdogs to greatest favorites). The standard deviations are largely constant across deciles. There is no systematic pattern in standard deviations across the deciles. Point-differences are not more or less dispersed based on the probability of the team winning. For misperceptions about the point distribution to matter, bettors would have to get the largely constant distribution of point differences consistently incorrect. Moreover, this explanation seems especially unlikely, given that, based on the evidence in Table 3, Spread and Moneylines capture nearly completely overlapping information for win probabilities, implying bettors recognize the point spread distribution is constant across games and treat the expected point difference as sufficient to determine win probabilities. Hence, the alternative explanation that bettors misperceive higher moments of the point difference distribution seems less consistent with the data.[19]

But, to take this alternative explanation one step further, we also examine another contract written on the fundamental of each game – total points scored – using the Over/Under contract. Recall that the Over/Under contract is a contingent claim on the total points scored by both teams. We show that market prices in the Over/Under market also accurately reflect the total number of points scored by both teams in games. We sort games into twenty bins based on the Over/Under

[19] In contrast, the evidence in financial markets indicates that the standard deviations of asset returns vary substantially over time for the same assets, as well as across different assets.

line of the game, and plot a binned scatterplot of the Over/Under line versus the total number of points scored in Figure 2. All points lie on the 45 degree line, illustrating that the Over/Under contracts accurately capture point totals in games. The R^2 from a regression of the raw data (not binned) of actual points on expected point totals is 89.3%, suggesting a very good fit between expected points and actual points scored across all games in all sports.

The second panel in Figure 2 sorts games into deciles based on the Moneyline ratio, and plots the average Over/Under line and the average total score of games in each decile. The Over/Under lines are also accurate in capturing the total score of games regardless of how close or lopsided the game is expected to be. This evidence indicates that the betting market is very accurate at forecasting expected points scored, the key piece of fundamental information for all betting contracts. The fact that Spread and Over/Under contracts are not mispriced, and predict point differences (Spread) and point totals (Over/Under) extremely accurately, challenges the notion that the FLB in Moneyline contracts is due to erroneous understanding of the point distribution. For misperceptions of the point distribution to be driving the results, the market must consistently misperceive the relatively static point distribution across games, despite displaying extreme sophistication in forming expectations about the number of points each team will score in each game, which is a conditional random variable that varies substantially across games. It seems implausible that betting markets are so good at capturing a complex dynamic component of points scored while simultaneously and consistently misperceiving a very stable and static feature of points scored. The evidence suggests that while it may be possible for misperceptions of the point distribution to play a role, this explanation seems unlikely.

3.4.3 Heterogeneous Beliefs

Our framework assumes that bettors are homogeneous; but belief heterogeneity is potentially another explanation for the facts we document. Other studies have suggested belief heterogeneity as an explanation for the FLB.[20] The logic behind the heterogeneous beliefs explanation for the FLB is that in fixed-odds markets, the dollar volume wagered on a team is increasing in how favored the team is. In equilibrium, only a small proportion of dollar volume is wagered on extreme underdogs, meaning that the low returns to betting on underdogs can be theoretically explained by the fact that only a small proportion of bettors with extreme beliefs choose to wager on underdogs. If the

[20] Previous studies link belief heterogeneity with the FLB from a theoretical perspective (e.g., Ali (1977), Shin (1991, 1992), and Ottaviani and Sørensen (2009, 2010)). Gandhi and Serrano-Padial (2015) empirically estimate the degree of belief heterogeneity assuming a discrete choice environment, where the choice of which horse to bet on is isomorphic to a model of horizontally differentiated demand. Green, Lee and Rothschild (2020) suggest a slightly different, but related, form of belief heterogeneity as an explanation for the FLB. In particular, they suggest that racetracks post morning-line odds that reflect a FLB, which are believed by gullible traders, and create a wedge between the valuations of these gullible traders and risk-neutral arbitrageurs.

average belief is unbiased, then the underperformance of underdogs in Moneyline contracts can co-exist with the lack of any pattern in Spread contracts.

In the absence of data on betting volume, empirical work studying the role of heterogeneous beliefs in the FLB (Gandhi and Serrano-Padial (2015) and Green, Lee and Rothschild (2020)) follows the common assumption made in the literature that bettors wager an exogenously fixed amount, which is assumed to be the same across contracts. This assumption pins down the distribution of beliefs using the equilibrium relationship between dollars wagered and odds (Gandhi and Serrano-Padial (2015)). Our unique data, however, provide us with the proportion of the *number* of bets placed on each team, which allows us to test the logic behind a heterogeneous beliefs model as an explanation for the FLB without having to assume equal bet sizes across contracts.

Assuming the bookmaker takes no risk, the proportion of dollars bet on a team in a game (which we refer to as *dollar market share*) is pinned down by the odds offered on the two teams. More favored teams should have a higher proportion of dollars bet on them, while less favored teams should have a lower proportion of dollars bet on them. Assuming equal betting amounts across bettors, the FLB naturally emerges, with the marginal bettor who bets on extreme underdogs having more extreme beliefs.

We compare the estimated dollar proportion of bets with the proportion of the number of bets placed on a team in our data. We sort contracts into twenty bins based on the estimated dollar proportion of bets. Figure 3 plots a binned scatterplot of the average dollar market share of bets on the x-axis against the average proportion of bets on the y-axis for each bin. The proportion of bets is generally increasing in the dollar proportion of bets on a contract, though the opposite is seen in the tails (for extreme favorites and extreme underdogs.) The relationship between the proportion of bets and the dollar proportion of bets is flatter than the 45 degree line. For example, while the dollar proportion of bets for contracts in the top and bottom bins (extreme favorites and underdogs) is approximately 90% and 10%, the average proportion of bets placed on contracts in these bins are 56% and 44%, respectively. Notably, the average returns for contracts in these bins are -0.40% and -24%. Hence, 44% of all bets are being placed on contracts with the most negative returns, though only about 10% of the *dollar* volume is being placed on these contracts. While theoretically, the FLB may emerge with only a small proportion of bets on extreme underdogs (as noted in Gandhi and Serrano-Padial (2015)), our evidence suggests that many bettors bet on extreme underdogs. The evidence suggests that while in theory heterogeneous beliefs may help explain the FLB, the large proportion of bets on extreme underdogs with very negative returns is hard to rationalize in a heterogeneous beliefs model.[21]

[21] Appendix B presents a stylized model of risk-neutral bettors with heterogeneous beliefs about the expected point differential in each game, who partition into Moneyline contracts based on their beliefs. Assuming that the

3.5 Reconciliation with Other Betting Evidence

We find strong evidence of a FLB in football and basketball, but only in Moneyline and not Spread contracts, which we interpret as a preference for lottery-like payoffs. However, can this story also reconcile why some researchers fail to find a FLB in other sports, such as the National Hockey League (NHL) and Major League Baseball (MLB) (see Woodland and Woodland (1994, 2001))?

We obtain data on Moneyline contracts for all MLB and NHL games from SportsInsights.com from 2005 through May 2013.[22] Analyzing Moneyline contracts from MLB and NHL games, we find that the cross-sectional dispersion in expected outcomes and risk is simply much lower in NHL and MLB betting contracts than it is in football and basketball. The lack of a FLB found in the NHL and MLB betting contracts may therefore be partially attributed to the lack of extremely risky bets in these sports' betting contracts.

The first panel of Figure 6 plots the proportion of MLB and NHL betting contracts that fall within each of the decile bins formed by sorting the Moneyline contracts in our main sample of basketball and football games. The majority of MLB and NHL contracts lie within the fifth and sixth decile bins in terms of Moneyline, and more than 90% of contracts lie in the fourth through seventh deciles. Put differently, less than 10% of MLB and NHL Moneyline contracts lie within the most extreme 40% of football and basketball contracts, and looking back at Figure 1, most of the return action comes from these extreme risky bets, which are largely non-existent in MLB and the NHL.

The second panel of Figure 6 plots the returns of contracts in our main sample, as well as those in the sample of MLB and NHL contracts, that fall into decile bins four through seven from our main sample, which covers the majority of all MLB and NHL contracts. As the figure shows, there is no FLB for these deciles (e.g., the average return of decile four is -2.0% and decile seven is -2.5%).

The data present a way to partially reconcile the lack of a FLB found in MLB and NHL Moneyline contracts with the presence of a FLB that we find in our sample of basketball and football contracts: there simply are not enough contracts with extreme risk (deciles 1 and 2 in our main sample) that bettors have a preference for in MLB and the NHL. This evidence may also

distribution of beliefs of bettors is the same across games, the model cannot reconcile the increasing proportion of bettors choosing to bet on extreme underdogs. This result holds regardless of parametric assumptions about the shape of the distribution of beliefs. However, a caveat is that the analysis is conditional on the choice to bet on a game. In reality, bettors' choice sets involve all games offered on a date (as well as not betting), and bettors may sort across games based on the bets they find most attractive. In other words, the distribution of beliefs may not be the same across games. However, when we consider a wider choice set, we find that the choice sets of bettors are unlikely to explain our findings.

[22] Spread contracts are offered on MLB and NHL games. However, because of the low-scoring nature of the sports, the lines offered are uniformly identical -1.5 point spreads in MLB and -0.5 or -1.5 point spreads in the NHL.

further support a preference-based explanation over beliefs, since beliefs should still matter even if the distribution of risk is small. However, if bettors have a preference for tail events, there needs to be enough tail risk for those preferences to show up in prices. In other words, we conjecture that a preference-based explanation exhibits non-linear effects, consistent with models of tail risk preferences used in financial economics, while a belief-based explanation may be more linear, or at least does not naturally exhibit non-linear effects. The lack of FLB in betting prices for MLB and the NHL is therefore consistent with the lack of risk available in those contracts under a preference for skewness. This result is similar to what researchers have found in equity options markets, where there are large price distortions for deep out-of-the-money options, but no significant pricing differences for modest out-of-the-money or near-the-money contracts. The risk-return effects in equity options markets seem to be non-linear as well and are similar to what we uncover in sports betting markets. We now compare our results in betting markets to the finance literature on low risk anomalies and option pricing.

3.6 From Betting Markets to Financial Markets

What can our results tell us about similar patterns in financial markets? In financial markets, the low-risk anomaly refers to an empirical phenomenon where riskier assets earn lower risk-adjusted returns than less risky assets. Low-risk anomalies are a robust phenomenon found in a number of asset classes, including individual stocks, corporate bonds, options contracts, equity indices, U.S. Treasuries, and commodities. This evidence seems to contradict predictions from traditional asset pricing models, which assume investors are risk averse, have correct beliefs, and evaluate the risk of any asset in the context of their entire wealth portfolio. Under these assumptions, people diversify away idiosyncratic risk, and the only priced risks in financial markets should be systematic, non-diversifiable risks, where equilibrium returns compensate investors for bearing those risks. To explain low-risk anomalies, the literature has proposed two types of explanations: those based on capital market frictions (e.g., leverage constraints Black et al. (1972) and Frazzini and Pedersen (2014)), and those based on bettor preferences for lottery-like payoffs that are reflected in equilibrium prices (Brunnermeier, Gollier and Parker (2007) and Barberis and Huang (2008)).

Applying the common assumption that people's risk-preferences are stable across contexts, our results are able to shed light on the low-risk anomaly in financial markets. Since theories based on capital market frictions in financial markets are not applicable to betting markets, our finding that the FLB is likely driven by preferences suggests that preferences could play an important role in explaining the facts in financial markets as well.

It is difficult to assess the extent to which preferences may play a role in explaining low risk anomalies in financial markets, given other confounding factors. However, one way to make this assessment is to perform a *quantitative* comparison between the evidence in betting markets and the evidence in financial markets. If quantitative magnitudes of these patterns are similar in betting markets, where there are fewer confounding factors, then it may be more reasonable to conjecture that the same preference-based explanation can account for a substantial portion of the financial markets evidence.

In the next two sections, we perform two sets of quantitative comparisons. First, we compare the magnitude of patterns in betting data to patterns observed in equity options and equity markets. Riskier out-of-the-money call (and put) options are priced more expensively than less risky in- and at-the-money options. High beta and high idiosyncratic volatility stocks are priced more expensively than low beta and low idiosyncratic volatility stocks. We compare the magnitude of these pricing patterns with the magnitudes we observe in betting markets. Second, we calibrate preferences for skewness required to rationalize the betting market facts and seek to understand how those preferences align with preferences calibrated to match the financial markets evidence in other work. Across our different exercises, we find evidence that the FLB is quantitatively comparable to low-risk anomalies in financial markets, suggesting that a preference for skewness might be a unifying explanation for these facts across both settings.

4 Betting Contracts and Equity Options/Contracts

We compare the returns of betting contracts with the returns of equity options contracts in two ways. First, we show that betting contract prices can be converted to *implied volatilities*, based on the implied volatility of the underlying point-difference distribution required to deliver the same contract expected return as a Spread contract. With this conversion, we analyze how implied volatilities vary with contract payoff probabilities in sports betting contracts, and then compare this with implied volatility surfaces in options contracts. Second, we directly analyze option excess returns, and compare the patterns observed across options of different payoff probabilities with the patterns in betting contract returns. Finally, we analyze low risk trading strategies in financial markets and compare their returns to those in sports betting markets. For all tests, we find evidence that the magnitude of pricing patterns observed in betting markets mirrors the magnitude of pricing patterns in equity and options markets.

4.1 The Betting Implied Volatility Surface

Each of the betting contracts we study is a bet on the difference between the number of points scored by each team, with the bet paying off if the point-differential exceeds a certain threshold, and offering zero payoff otherwise. Here, we use data on the underlying point-differentials to synthesize and quantify our results in terms of an *implied volatility surface*, which is analogous to the implied volatility surface studied in options markets. Framing our results in this way, we can quantify the return differences we observe in terms of observable fundamentals, and compare the results in betting markets with evidence from options contracts in equity markets.

To construct the implied volatility for each betting contract, we use two ingredients. The first is an estimate of the objective probability that a contract pays off, p. The second is the market "implied probability" of a contract paying off, $\hat{p}$, embedded in the contract price. Using p and $\hat{p}$, we calculate the implied volatility of a contract, $\hat{\sigma}$, as the value that satisfies the following,

$$F\left(F^{-1}(p;\sigma);\hat{\sigma}\right) = \hat{p} \tag{5}$$

where F is an assumed cumulative distribution function of the point differential minus the spread line, and σ is the standard deviation of F. For each sport, we use the Normal CDF for F, and we estimate σ using Maximum likelihood. To compute the objective win probability, p, we run a kernel regression of Moneyline contract returns on the Moneyline and Spread line of the contract, and use this regression to estimate the expected return of each Moneyline contract. We back out p from the expected returns and the price of each contract. We find that this procedure produces reliable estimates of win probabilities. We further discuss our methodology and present evidence of the accuracy of the estimated win probabilities in Appendix A.[23] We calculate the implied probability, $\hat{p}$, directly from contract prices as the probability of a contract payoff that sets the contract's expected return equal to the expected return of a Spread contract with 50% chance of paying off. Although our approach uses the Normal CDF to calculate implied volatilities, this is done purely to capture the magnitudes in an interpretable way. Our calculations do not rely on the assumption that the point-differential minus spread line follows an identical normal distribution across games. Rather, our methodology imposes no parametric assumption on the true distribution of the point-differential minus spread line, but simply uses the Normal distribution as a convenient tool to express the magnitude of the pricing patterns we observe in terms of standard deviation.[24]

For this exercise, we exclude contracts where p is between 0.45 and 0.55, because our methodol-

[23]Our approach is broadly similar to the Lowess smoothing approach that Snowberg and Wolfers (2010) take to map from horse betting odds to probabilities. Additionally, we omit non-payoff contract characteristics in these regressions. While non-payoff characteristics have some explanatory power, they do not appear related to our main quantity of interest here, the relative performance of riskier versus less risky contracts.

[24]An alternative way to construct implied volatilities is to calibrate the point-spread distribution using the assumption that the point spread minus spread line is *iid* Normal across games, and use the calibrated distribution to find the implied volatility that matches the implied win probability for a game. This alternate method heavily relies upon the assumption of normality, whereas the method we use only uses the normal distribution to quantify the magnitude of implied probability distortions in terms of the standard deviation of a Normal distribution. We empirically analyze how well the Normal distribution fits the data in Appendix D.1. The fit is reasonable, though the non-parametric estimates do a substantially better job of capturing win probabilities.

ogy is not particularly well-suited to handle win probabilities very close to 50%. Note that Equation (5) has an undefined solution at $p = 0.5$. This restriction excludes 31% of contracts. However, our interest is primarily in studying contracts with low and high values of p.

For each sport, we sort contracts into one of 90 equally spaced bins based on their estimated win probability (i.e., one bin for contracts with win probability between 0% and 1%, a bin for win probability between 1% and 2%, etc.). For betting contracts on underdogs (payoff probabilities less than 50%), higher implied volatilities correspond with more expensively priced contracts. For contracts on favorites (payoff probabilities greater than 50%), higher implied volatilities are associated with cheaply priced contracts. This pattern occurs because a more volatile point distribution corresponds with favorites being less favored and underdogs being more likely to win, ceteris paribus.

Figure 5 plots a scatterplot of the average estimated win probability versus the average implied volatility of each bin for Moneyline contracts in each sport. We refer to these plots as "implied volatility surfaces," using the options market terminology, as they relate the moneyness of the contract (the probability of the contract paying off) with the implied volatility in the contract. For each sport, the figure reveals a striking volatility *smile* or *smirk*, that is reminiscent of the smile and smirk observed in options implied volatility surfaces. Implied volatilities are higher on average for contracts that have a high probability of paying off and for contracts that have a very low probability of paying off. There also appears to be some asymmetry in the plot around 50% in the implied volatilities for favorites versus underdogs, suggesting that bettors treat gains and losses differently.

To provide context for the magnitude of the implied volatility surface smile, we compare it to the implied volatility smile in options markets. Panel A of Table 2 displays the average implied volatilities for contracts expected to pay off approximately 5-15% of the time and 85-95% of the time. We compare these implied volatilities with the sample volatility of the point-spread distribution to calculate a measure of the implied volatility "premium" for the contracts. Contracts expected to pay off 5% to 15% of the time have implied volatilities that are 1.4% (for NBA contracts), 12.4% (NFL), 11.3% (NCAAB), and 12.8% (NCAAF) higher than the true sample volatility, with an average premium of 8.6% across sports. Contracts expected to pay off 85% to 90% of the time have implied volatilities that are 12.4% (NBA), 17.4% (NCAAB), 23.7% (NCAAF), and 24.8% (NFL) higher than the true sample volatility, with an average premium of 19.6% across sports.

For comparison, Panel B of Table 2 presents the simple average implied volatility for "standardized" 10 (and -10) delta and 90 (and -90) delta call (and put) options with one month maturity for a set of 13 equity indices. These correspond roughly to payout probabilities of 10% and 90%, respectively. Panel C presents the same quantities averaged across all individual equity options from the

OptionMetrics IvyDB Implied Volatility Surface file.[25] The options at these delta values are deep out-of-the-money (low probability of paying off) and deep in-the-money (high probability of paying off). Because the implied volatility of options embed a variance risk premium, we capture the magnitude of the options smile by comparing the implied volatilities of the out-of-the-money options with implied volatilities of 50 (and -50) delta call (put) options, which have payoff probability of approximately 50%. The results presented in Panel B show that index options have a pronounced smirk, which is consistent with the intuition that index options are often used to hedge against the possibility of a market crash. Deep out-of-the-money put options and deep in-the-money call index options have a substantial implied volatility premium (on average 43.6% and 48% above the implied volatility of at-the-money options). Deep in-the-money put options and out-of-the-money call index options have implied volatilities that are slightly greater than the implied volatility of the at-the-money options (on average 4.5% and -0.9% below at-the-money options). The results presented in Panel C suggest that equity options have a more symmetric smile. Deep out-of-the-money call and put options have implied vols that are, on average, 17.0% and 26.7% higher than less-risky in-the-money calls and puts. Deep in-the-money options have implied vols that are, on average, 21.1% and 12.4% higher than at-the-money calls and puts.

4.2 Call Options Returns

We can also directly compare the patterns in options and betting contracts in terms of average returns rather than implied volatilities. We compare options returns at different values of delta ($\approx$ probability of option paying off) with betting contract returns with similar win probabilities.

We group Moneyline contracts into five bins: contracts with win probabilities in the ranges [0, 0.2], (0.2, 0.4], (0.4, 0.6], (0.6, 0.8], and (0.8, 1]. We compare the returns of the betting contracts binned with the delta-hedged excess returns of one-month maturity call options, which are sorted into five bins based on option deltas with the same interval ranges. The options contract returns are drawn from Frazzini and Pedersen (2022), and correspond with the returns of value-weighted portfolios that are delta-hedged daily. We include returns separately for both index options and for all single name equity options.

Figure 6 plots the average returns of the betting contracts and the delta-hedged excess returns of the options contracts in each of the bins. Across each contract type, the returns are increasing by bin, with contracts in higher payoff probability bins earning higher returns. The average excess returns of index options ranges from -26.9% in the lowest delta bin to -0.6% in the highest delta

[25] We follow Frazzini and Pedersen (2022) in selecting our indices. OptionMetrics constructs implied volatilities for "standardized" delta values and maturities by interpolating the prices of traded options. The values produced, accordingly, do not correspond with any single traded option. See the appendix for more details.

bin and the average excess returns of single-name options ranges from -16.1% in the lowest delta bin to -2.8% in the highest delta bin. For betting contracts, average returns range from -21.0% in the lowest payoff probability bin to 0.2% in the highest payoff probability bin.

While put option returns display similar patterns as call option returns, we specifically focus on call option returns because they pose a challenge to traditional theories of asset pricing. Out-of-the-money call options are generally *more* positively exposed to market and macroeconomic risks, and hence under traditional asset pricing theories, should command a risk-premium that earns higher average returns. However, we observe the opposite in the data.[26] By focusing on call options, we find that the magnitude of the patterns of returns is similar to those that we observe in betting contracts, which suggests a possible unifying explanation for these facts in both settings.

4.3 Interpreting the Evidence

Comparing betting contracts with options contracts from financial markets raises some issues. For instance, there are many reasons that volatility smiles and smirks might exist in options markets. The difficulty in assessing the true probability distribution of underlying financial asset returns has led to several potential explanations. For example, implied volatility and option delta calculations assume that stock prices are log-normally distributed. However, if the true distribution of asset prices is fatter tailed, possibly coming from stochastic volatility or jumps in volatility, then this could lead to an implied volatility smile. Exposure to aggregate risk may also play a role, where deep-out-of-the money put options provide a hedge against market crashes, and the implied volatility smirk in index options may be driven by risk premia associated with market crashes.

The volatility smile can also come from other features, such as non-traditional preferences or belief heterogeneity. These features are consistent with the delta-hedged under-performance of risky, out-of-the-money options, as argued by Boyer and Vorkink (2014). The similar magnitude of the implied volatility smile in options on single name stocks, and the similar magnitude of return patterns, provides more quantitative evidence of the similarity in the pricing of out-of-the-money betting contracts and options. Our evidence in betting markets is particularly useful because the simple state space and idiosyncratic nature of contracts makes the probability distribution easy to estimate, so misspecification of the underlying distribution is unlikely to drive the results. Moreover, aggregate crash risk cannot be embedded in these contracts. Hence, our results more likely point to a preference-based explanation as a unifying concept for both markets. We investigate a preference-based model in the next section to see if it can match the magnitudes in both markets.

[26] Of course, in contrast, put options may offer insurance against macroeconomic downstates. The traditional paradigm predicts that such assets earn lower returns. However, the literature has generally found that the returns of out-of-the-money puts are even lower than standard risk-aversion would predict. For example, see Bondarenko (2014).

4.4 Betting Market and Financial Market Low-Risk Trading Strategies

Table 4 reports low-risk trading strategy returns for betting contracts and for financial securities. Panel A reports returns for betting markets. The first column reports the mean, t-stat, and annualized Sharpe ratio of a betting strategy that goes long the favorite in the Moneyline and short the underdog in the Moneyline contract, using the decile cutoffs from Figure 1 and Table 1. The trading strategy takes an equal-weighted position in every game within the decile, scaled to be one dollar long and short. We rescale the trading strategy returns to 10% ex post annualized volatility and report annualized means and Sharpe ratios by using the average nunber of games per year. We rescale to 10% annual volatility to compare to financial markets low-risk trading strategies below. The first column of Panel A reports that going long the favorite and short the underdog in the Moneyline produces a 22.7% average return with a t-stat of 5.85 and an annualized Sharpe ratio of 2.27. This premium is consistent with low-risk premia documented in equity option markets and equity markets, which we highlight below.

The second column of Panel A reports the returns to a trading strategy that is long the favorite and short the underdog in the Spread contract. Consistent with Figure 1 and Table 1, there is no return difference between betting on the favorite versus the underdog in the Spread contract, resulting in a -2.33% return with a t-stat of -0.60 and Sharpe ratio of -0.23. The Spread contract, which holds risk constant across all bets, exhibits no FLB and no return premium. Put differently, we find no low-risk premium when we examine a trading strategy that does not provide different risk exposure to investors. This placebo test rules out other potential explanations for the return difference found in the Moneyline contract (column 1) since risk exposure is the only different between the two contract types.

Columns 3 and 4 report the returns to being long the favorite in the Spread contract and short the underdog in the Moneyline, and long the underdog in the Spread and short the underdog in the Moneyline. Both strategies produce highly significant returns (11.53% and 21.51% with t-stats of 2.97 and 5.55, respectively) driven by the poor returns of extreme underdogs in the Moneyline. The 21.5% difference in returns between the underdog in the Spread versus the Moneyline contract for the *same* games highlight the premium bettors are willing to pay for the lottery-like payoffs provided by the Moneyline for the same underdogs. Likewise, columns 5 and 6 show that the Moneyline favorite is discounted heavily by bettors due to its low-risk payoffs. Column 5 reports returns for going long the favorite in the Moneyline and short the favorite in the Spread, which yields an impressive 35.34% return (t-stat = 9.11) with a Sharpe ratio of 3.53. Column 6 reports results for the strategy that is long favorites in the Moneyline and short underdogs in the Spread, which produces a 13.79% return (t-stat=3.65) and a 1.38 Sharpe ratio.

Taken together, Panel A of Table 4 shows clear evidence of poorer returns for Moneyline underdogs and abnormally positive returns for Moneyline favorites, but no return differences in Spread contracts based on winning probability, consistent with lottery-like preferences affecting prices.

Panel B of Table 4 reports statistics for financial market strategies. The first two columns of the table correspond with Options Betting-Against-Beta (BAB) strategies, using the numbers reported in Table VI of Frazzini and Pedersen (2022) for equity and index options trading strategies. The third and fourth columns of the table correspond with the excess returns of US and Global BAB strategies in equities from Frazzini and Pedersen (2014), using strategy returns from the AQR data library updated through February 2022. The last two columns of the table correspond with the idiosyncratic volatility trading strategy from Ang et al. (2006), using US and Developed Market data from the Jensen, Kelly and Pedersen (2021) data library through December 2021. For the idiosyncratic volatility strategies, the reported statistics correspond with residual alpha from a regression of the trading strategy returns on the Fama and French (1993) three factors. The premium documented in financial markets for other low-risk strategies are similar in magnitude to what we find in sports betting markets. Sharpe ratios range from 0.71 to 1.78. These Sharpe ratios are slightly smaller than those in betting markets, but the differences may be rationalized given the large transaction costs in betting markets relative to financial markets.

5 Preferences and the Favorite-Longshot Bias

In this section, we focus on the preferences of a bettor who has correct beliefs and is indifferent between wagering the optimal amount (given her preferences) on all betting contracts available. While there are a wide-range of bettors with potentially different motivations for betting (for example, recreational bettors versus professional 'sharps'), our focus is on a 'typical' or average bettor, who understands that betting is a negative expected return proposition and endogenously chooses to participate.

5.1 Preference Specification

Consider a wager of b on a simple binary outcome. The bet pays off with probability $(1-p)$, in which case the bettor receives yb dollars. With probability p, the bet does not pay off, and the bettor loses the b she wagered. We consider a bettor with reference-dependent preferences, who evaluates this bet by computing

$$V = (1-\eta)\underbrace{\left[p(-b) + (1-p)yb\right]}_{\text{Expected Utility}} + \eta\underbrace{\left[w(p)v(-b) + (1-w(p))v(yb)\right]}_{\text{Non-Expected Utility}} \tag{6}$$

where $w(\cdot)$ is a *probability weighting* function and $v(\cdot)$ is the *value* function. As advocated by Kőszegi and Rabin (2006), the formulation in Equation (6) has a "traditional" Expected Utility term, in addition to a non-Expected Utility term, with $\eta \in (0,1)$ governing the relative weight of the two terms.

The Expected Utility term in Equation (6) corresponds to a risk-neutral utility function. Given that bettors are unlikely to wager substantial portions of their wealth on any given bet, and that risk aversion is thought to be small for lotteries with stakes that are small relative to total wealth, risk-neutrality is a natural candidate for the Expected Utility component of preferences in our setting.

The non-Expected Utility term in Equation (6) corresponds with rank-dependent utility (Quiggin (1982)), and captures the reference-dependent features of the bettor's preferences. We use the functional forms proposed by Tversky and Kahneman (1992) for the value function, $v(\cdot)$, and the probability weighting function, $w(\cdot)$.

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \tag{7}$$

and

$$v(x) = \begin{cases} x^\alpha & x \geq 0 \\ -\lambda(-x)^\alpha & x < 0 \end{cases} \tag{8}$$

where $\alpha \in (0,1), \gamma \in (0.28, 1)$ and $\lambda \geq 1$.[27]

The value function is computed over gains and losses relative to a reference point. Here, we assume that the reference point is zero, which is the bettor's payoff if they do not take the bet.[28] $v(\cdot)$ is also concave over gains and convex over losses, with α governing the concavity and convexity of the value function over gains and losses. These characteristics of the value function lead the bettor to exhibit *diminishing sensitivity*, where for both gains and losses, the marginal effect of an additional gain or loss is smaller the further gains and losses are from the reference point. The

[27] γ is sometimes assumed to be in the range $(0,1)$, but it is only increasing everywhere for $\gamma \geq 0.279$. Values less than that can admit negative decision weights. See Ingersoll (2008).

[28] This is often referred to as "status-quo" reference dependence, where potential gains and losses are measured relative to the agent's wealth in the status quo, as specified in Tversky and Kahneman (1979). Other work assumes different reference points. For example, Kőszegi and Rabin (2006, 2007) propose that the reference point is an individual's *expected* outcome. In a more dynamic environment than assumed here, the reference point is sometimes assumed to be an individual's wealth in some past period and combined with an assumption about framing (e.g., an individual may compute potential gains and losses for an asset with respect to the time at which the individual purchased the asset). See Grinblatt and Han (2005), Barberis and Xiong (2009), Imas (2016), and Barberis, Jin and Wang (2021) for analysis of this idea as it relates to stock price behavior, and Andrikogiannopoulou and Papakonstantinou (2019) for evidence from individual bettors. See O'Donoghue and Sprenger (2018) for a broader discussion of reference dependence.

value function also allows the bettor to display *loss-aversion* ($\lambda > 1$), where she may be more sensitive to small losses than small gains; however, as we discuss further, loss-aversion appears to be at odds with the decision to bet. The features of the value function can be drawn in contrast with the general properties of utility functions under Expected Utility, where people are thought to evaluate decisions based on their terminal wealth (rather than based on gains or losses) and the utility function is assumed to be concave everywhere (rather than concave over gains and convex over losses).

The bettor also uses transformed probabilities (captured by $w(\cdot)$), rather than objective probabilities, in the non-Expected Utility term of Equation (6). Probability weighting is motivated by evidence from psychology in the laboratory and field that people tend to systematically overweight low-probability events and underweight high-probability events (Fehr-Duda and Epper (2012), and Barberis (2013) summarize the evidence). The probability weighting function, and specifically, the parameter γ, govern how much the bettor transforms probabilities, with smaller values of γ corresponding with more over-weighting of tails of the distribution. The standard interpretation of probability weighting is not that the bettor does not know objective probabilities but rather that the transformed probabilities represent the bettor's "decision weights" that give additional weight to tail events in her utility function.

We highlight two points regarding the non-Expected Utility preferences in Equation (6). First, the form of non-Expected Utility in Equation (6) can be considered a special case of Cumulative Prospect Theory (Tversky and Kahneman (1992); CPT). Under CPT, bettors probability weight for gains is expressed by the function $w^+(\cdot)$ and for losses with the function $w^-(\cdot)$. Our implementation corresponds with rank-dependent probability weighting of the form $w^+(1-p) = 1 - w^-(p)$. Our implementation of probability weighting is commonly used in other work (e.g., see Barseghyan et al. (2013) and Barseghyan et al. (2018)), but does have certain properties that substantially differ from those of other commonly used formulations of probability weighting (such as assuming $w^+(\cdot) = w^-(\cdot)$), particularly in how losses are treated. Second, and related, our implementation of probability weighting relies on rank-dependence, which faces recent empirical debate (Bernheim and Sprenger (2020)). We discuss these points, and how they relate to our results, in more detail in Appendix C.

5.2 Betting Decision and Equilibrium Prices

We consider the betting decision of the bettor introduced in the previous subsection. She chooses the optimal amount to wager, b^*, to satisfy her first- and second-order conditions. Her first order condition is given by

$$0 = (1-\eta)\left(-p + (1-p)(y)\right) + \eta\left(-w(p)\lambda\alpha b^{*\alpha-1} + (1-w(p))\alpha y^{\alpha} b^{*\alpha-1}\right) \tag{9}$$

which yields an expression for the optimal wager:

$$b^{*\alpha-1} = \frac{1-\eta}{\alpha\eta}\left(-\frac{-p+(1-p)y}{-w(p)\lambda + (1-w(p))y^{\alpha}}\right) \tag{10}$$

where the right hand side of the expression must be positive for the optimal wager to exist. The bettor's second-order condition is given by

$$0 > \eta(w(p)\alpha(1-\alpha)\lambda b^{*\alpha-2} - w(1-p)\alpha(1-\alpha)y^{\alpha} b^{*\alpha-2}). \tag{11}$$

Since $b^* > 0$ (bettors can't short), and $\eta, \alpha > 0$, the second-order condition reduces to $w(p)\lambda - (1 - w(p))y^{\alpha} < 0$. Jointly, the first- and second-order conditions imply that, under the assumptions of the model, a finite, positive betting amount for a contract exists when the returns to betting on the contract are negative, but where the bettor receives non-Expected Utility value from wagering on the contract.[29]

Labeling each contract $i = 1, \ldots, n$, we denote V_i as the utility that the bettor receives from wagering the optimal amount on contract i, b_i, which offers a payoff of y_i with probability $1 - p_i$. Following the literature on estimating preferences from aggregate betting data, we assert the equilibrium condition that the bettor is indifferent between all contracts offered at the equilibrium prices.[30] We implement this approach by asserting that each $V_i = V_s$, where V_s is the utility a bettor derives from wagering the optimal amount b_s on a spread contract that offers a potential payoff of $y_s = \frac{100}{110}$ per dollar wagered and has a 50% probability of paying off. This condition can

[29]Including risk-aversion in the Expected Utility component of preferences, as done in financial market applications, relaxes the requirement of negative expected returns.

[30]This condition can be motivated, for example, by following Jullien and Salanié (2000) and assuming that the bettor is part of a group of homogeneous bettors (who can co-exist with other bettor types), and that there is a bet placed on each contract offered on each game by at least one member of the group.

be expressed as,

$$
\begin{aligned}
0 &= V_i - V_s \\
&= (1-\eta)\left[p_i(-b_i) + (1-p_i)y_i b_i\right] + \eta\left[w(p_i)v(-b_i) + (1-w(p_i))v(y_i b_i)\right] \\
&- ((1-\eta)\left[0.5(-b_s) + 0.5 y_s b_s\right] + \eta\left[w(0.5)v(-b_s) + (1-w(0.5))v(y_s b_s)\right]) \\
&=^* \frac{\eta^{\frac{1}{1-\alpha}}}{(1-\eta)^{\frac{\alpha}{1-\alpha}}}\left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}\right)\left(\frac{((1-w(p_i))y_i^{\alpha} - w(p_i)\lambda)^{\frac{1}{1-\alpha}}}{(p_i - (1-p_i)y_i)^{\frac{\alpha}{1-\alpha}}} - \frac{((1-w(0.5))y_s^{\alpha} - w(0.5)\lambda)^{\frac{1}{1-\alpha}}}{(0.5-0.5y_s)^{\frac{\alpha}{1-\alpha}}}\right)
\end{aligned}
\tag{12}
$$

where the * step comes from substituting the optimal betting amounts for contracts into the expression. We also impose the additional condition, which is *not* imposed in the literature on the FLB, that V_i must be positive, implying that the bettor must derive positive utility from betting in the first place.

Equation (12) shows that in the interior of the range $(0,1)$, η, the relative weight the bettor places on the non-Expected Utility component of her preferences, does not matter for satisfying the equilibrium condition. It only enters into the equation (and utility at the optimum) via a multiplicative scaling factor $\frac{\eta^{\frac{1}{1-\alpha}}}{(1-\eta)^{\frac{\alpha}{1-\alpha}}}$. η plays a role in the amount of betting that a bettor chooses and the level of utility she derives from betting, but it *does not* affect her decision to bet on the extensive margin, and is not pinned down by her indifference between different contracts.[31]

5.3 Evaluating the Parameters of Interest

Our general approach to study the preference parameters of interest, $\Theta = (\alpha, \gamma, \lambda)$, is as follows. For a given choice of Θ, we construct the set of contract payoffs, denoted $\{y_{i,\Theta}\}_{i=1}^n$, implied by Θ and the loss probabilities across games, $\{p_i\}_{i=1}^n$. We evaluate a given Θ by observing how the expected returns constructed using the implied payoffs compare with the expected returns observed in the data. We treat the loss probabilities as known, and use the estimated probabilities from Section 4.1 (the estimation methodology for these probabilities is discussed in more detail in Appendix A).[32]

We rewrite Equation (12) to provide a fixed point expression for the implied payoff that satisfies

[31] One way to further study the value of η would be with additional data on the wealth of bettors and the dollar amounts that they choose to bet relative to their wealth.

[32] Our approach conceptually differs from the approach pioneered by Jullien and Salanié (2000) and developed further on the FLB (such as Gandhi and Serrano-Padial (2015)). In that approach, the probability of the winning horse is written as a function of the observed payoffs and the parameters of interest, and the objective is to find the parameter values that best match the empirically observed win probabilities, which is done via maximum likelihood.

the equilibrium condition, $y_{i,\theta}$, for each contract i,

$$y_{i,\Theta} = \frac{p_i - A_s^{\frac{\alpha-1}{\alpha}} \left((1-w(p_i))y_{i,\Theta}^{\alpha} - w(p_i)\lambda\right)^{\frac{1}{\alpha}}}{1-p_i}, \tag{13}$$

where $A_s \equiv \frac{((1-w(0.5))y_s^{\alpha}-w(0.5)\lambda)^{\frac{1}{1-\alpha}}}{(0.5-0.5y_s)^{\frac{\alpha}{1-\alpha}}}$. Given parameters Θ, and the loss probability p_i, we use Equation (13) to calculate the implied value of $y_{i,\Theta}$ for each contract (if it exists) by iterative methods. A solution need not exist for $y_{i,\Theta}$ for all assumed parameter values and loss probabilities.

We evaluate a given choice of Θ by constructing the mean squared error of the expected returns implied by Θ versus the expected returns estimated from the data,

$$\text{MSE}_{\Theta} = \frac{1}{n}\sum_{i=1}^{n}\left(\underbrace{(1-p_i)y_{i,\Theta} - p_i}_{\text{Implied Expected Return}} - \underbrace{((1-p_i)y_i - p_i)}_{\text{Realized Average Return}}\right)^2. \tag{14}$$

Note that this expression corresponds with evaluating the weighted mean squared errors of payoffs y_i, where the weight on observation i is $(1-p_i)^2$. This type of weighting of errors is desirable, as we expect low payoff probability contracts with high y_i to also have larger magnitude errors (and vice-versa for high payoff probability contracts with low y_i). When a given value $y_{i,\Theta}$ does not exist, we substitute a squared error of 1. Contracts must have negative expected returns in our model, and the expected return is bounded below at -100% (a bettor loses their wager with probability 1). Hence, this treatment has the economic interpretation of Θ yielding a "maximally incorrect" estimate of the expected return when it does not provide a feasible solution for $y_{i,\Theta}$.

5.4 How Well Does the Model Capture the Favorite-Longshot Bias?

To more closely study the ability of the model to explain the betting markets evidence with a reasonable parametrization, and in a manner consistent with the evidence from financial markets, we fix the probability weighting parameter, using values of $\gamma = 0.65$ and $\gamma = 0.5$, and calibrate the corresponding values of α and λ that provide the lowest mean squared error of expected returns.[33] Probability weighting is well-studied, with substantial experimental and field evidence. Tversky and Kahneman (1992) report a value of $\gamma = 0.65$, and subsequent evidence confirms that similar values are able to explain the facts in financial markets, so the choice of $\gamma = 0.65$ seems like a natural starting point. Evaluating $\gamma = 0.5$ also provides another point of comparison with more extreme probability weighting (that still lies within reasonable values reported in experimental evidence, for

[33] α, γ, and λ are tightly linked and not well-identified from one another without additional data, which is why we choose a calibration approach rather than an unrestricted estimation. We discuss the economic relationship between these features of preferences in more detail in the appendix.

example see Booij, Van Praag and Van De Kuilen (2010)), that also allows us to evaluate how α varies with γ.

Table 5 reports the results from the estimation. Panel A reports estimates of α, the mean squared error (MSE) of expected returns, and the number of contracts for which the model is able to calculate valid payoffs (positive payoffs that satisfy the bettor's optimality conditions). Standard errors constructed from 2,000 bootstrap samples are reported in parentheses. The lower bound for loss-aversion, $\lambda = 1$, tightly binds in both cases. The estimated values of α are 0.650 and 0.502, very close to the assumed values of γ. The model does a very similar job in explaining the data for $\gamma = 0.65$ and $\gamma = 0.5$, with similar mean squared errors (0.012 for both), and a similar number of contracts with valid payoffs (73,188 versus 73,187, out of a total of 73,218).

Panel B reports more details from the estimation, splitting contracts into deciles based on their Moneyline, with decile 1 corresponding to extreme underdogs and decile 10 extreme favorites. The panel also reports the fitted payoffs, expected returns, and proportion of contracts with valid payoffs across the deciles. For both values of γ, the model is able to capture the optimal decision to bet a positive and finite amount, calculating valid payoffs for more than 99.9% of betting observations in deciles 1 through 9, and 99.6% of observations in decile 10.

The model does reasonably well qualitatively and quantitatively at capturing contract prices and the FLB in expected returns. Focusing on $\gamma = 0.65$, the model slightly over-estimates the payoff a bettor demands to bet on extreme longshots in decile 1, resulting in slightly higher expected returns for extreme longshots implied by the model (-16.2%) versus the data (-22.5%). The model also underestimates the returns for more modest underdogs, but more closely captures the returns of moderate to heavy favorites. The patterns of implied returns are similar for $\gamma = 0.5$, which does a slightly better job capturing the returns of larger underdogs at the cost of doing a slightly worse job of capturing the returns of more moderate underdogs. We plot the model-implied expected returns (with $\gamma = 0.65$) versus empirically estimated average returns by decile in Figure 7.

5.5 Comparison with Evidence from Financial Markets

Returning to the initial motivation of our calibration exercise: are the set of preferences that capture the FLB in sports betting similar to those that are also able to explain the financial market low risk anomalies? Our evidence suggests that probability weighting and diminishing sensitivity may similarly feature in explaining the financial markets and betting markets evidence.

Given the tight relationship between γ and α, we fix the value of γ at 0.65 (and 0.5), and study the corresponding value of α. Studies find that a probability weighting parameter of $\gamma = 0.65$ is able to explain the pattern of risk and return in the cross-section of options (Baele et al. (2016)) and the patterns of risk-and return in the cross-section of stocks (Barberis, Jin and Wang (2021)), and match experimental values of γ reported by Booij, Van Praag and Van De Kuilen (2010). We estimate the value of the diminishing sensitivity parameter, $\alpha = 0.65$. Booij, Van Praag and Van De Kuilen (2010) conduct a meta-study of experimental estimates of α. Most of the estimates lie between 0.5 and 0.95, and the average value they report is approximately 0.7. Barberis, Jin and Wang (2021) use a value of $\alpha = 0.7$ in a dynamic model where bettors evaluate gains and losses in stocks relative to the purchase price that explains various stock market anomalies, including the low risk anomaly.

Although we find similar values for α to those from other financial market applications, our model is static, and in our application of diminishing sensitivity, gains and losses are evaluated relative to the status quo.[34] The differences in the exact application of diminishing sensitivity in betting versus financial markets are natural, however, given the differences between the two settings. Bettors make decisions to bet hours (at most, days) before the outcome is revealed, making their wealth in the status quo a natural reference point against which to compute potential gains and losses. This is in contrast to financial markets, where individuals make investments that they may hold onto for months or years, and which they are able to continually reevaluate over that investment period.

A notable point is that while loss-aversion is one of the most well-known and well-studied components of CPT, including loss-aversion in our setting reduces the fit of the model. For the particular application we are concerned with, the relative pricing of risk in the cross-section of risky assets, loss-aversion may not play an important role in explaining the facts in financial markets.[35] Additionally, recent evidence from a more representative sample suggests that there is substantial heterogeneity in loss aversion, and that perhaps half of the population does not demonstrate loss

[34] Financial market applications of CPT are often applied in dynamic contexts, where diminishing sensitivity delivers the disposition effect that investors sell winners too quickly and hold on to losers too long as a result of them treating gains and losses relative to purchase price and caring about realized rather than "paper" gains and losses. In these applications, diminishing sensitivity, which leads to the disposition effect, can contribute to the lower returns of riskier stocks (Barberis, Jin and Wang (2021)). This is different than the role diminishing sensitivity plays in our framework for static betting markets, where we find it leads bettors to demand higher returns for riskier bets. The different implications come from different reference points used to define gains and losses – in betting markets the reference point is zero and in financial markets it is often taken as the purchase price – *and* from using a static (betting markets) versus dynamic (financial markets) model.

[35] Baele et al. (2016) find that loss-aversion has little explanatory power for the cross-section of options returns and that probability weighting is the important component for generating the low returns of out-of-the money options. Similarly, decomposing the contribution of the components of CPT to the underperformance of idiosyncratically risky stocks, Barberis, Jin and Wang (2021) report that probability weighting, along with diminishing sensitivity when evaluating gains and losses from the time an asset was purchased, are the primary contributors. Loss aversion, if anything, slightly detracts from CPT's ability to explain the low-risk anomaly. More generally, loss-aversion may potentially play a role in explaining the size of the equity risk premium (Benartzi and Thaler (1995) and Barberis, Huang and Santos (2001)) and low stock-market participation (Barberis, Huang and Thaler (2006)), but appears less important for the relative pricing of risky assets.

aversion (Chapman et al. (2018)). It is also possible that individuals with little loss aversion select into betting markets, or alternatively, that people behave in a less loss-averse way when gambling than when investing.

Overall, our calibration suggests similarities in explaining the low-risk anomaly across financial and betting markets. In particular, our results indicate that probability weighting and diminishing sensitivity can reconcile the FLB simultaneously with low risk anomalies in financial markets.

5.6 Bet Sizes

The model also delivers implications for the amount bettors choose to wager on different contracts. Equation (10) derives the *optimal* amount a bettor chooses to wager, which is novel to previous models of the FLB that typically assume bettors wager a fixed amount across contracts.[36] For the majority of our sample (32,685 games), we have data on the proportion of bets placed on the two teams in a game, where we can compare the ratio of the average wager placed on each of the two teams against the model-implied values. Since we do not explicitly target bet size in our estimation, this comparison provides an independent assessment of the model by focusing on a feature of the data we do not use in our calibration.

For a given game, assuming that the bookmaker takes no risk, market clearing for the Moneyline contract requires that the payoffs to winning bettors come from the pool of wagers. This can be formally expressed as two market clearing conditions:

$$(1 + y_h)\, n_h b_h = n_h b_h + n_a b_a \tag{15}$$

$$(1 + y_a) n_a b_a = n_h b_h + n_a b_a \tag{16}$$

where n_j is the number of bets placed on team j, b_h and b_a are the *average* amount per bet wagered on the home and away teams, and y_h and y_a are the payoffs offered for winning bets on the home and away team in a game. We can rewrite these conditions to derive the expression,

$$\frac{b_h}{b_a} = \frac{(1 + y_a)\, n_a}{(1 + y_h) n_h}. \tag{17}$$

The proportion of bets in our data directly allows us to compute $\frac{n_h}{n_a}$, which in turn allows us to

[36] Weitzman (1965) estimates the preferences of a "Mr. Avmart" (average man at the race track), whose "wagers are allocated among the entrants in a race exactly in the same proportion as the entire crowd apportions its money among the various horses." The most recent generation of work on the FLB (Jullien and Salanié (2000), Snowberg and Wolfers (2010), Gandhi and Serrano-Padial (2015), and Chiappori et al. (2019)) assumes that bettors decide between wagering a fixed dollar amount on each betting contract available. Under a risk-loving Expected Utility function or a risk-neutral value function with probability weighting (e.g., as considered by Snowberg and Wolfers (2010)), or under the CPT specification studied by Jullien and Salanié (2000), there is no finite, positive optimal wager. Bradley (2003) makes this point with reference to Jullien and Salanié (2000).

estimate the ratio $\frac{b_h}{b_a}$ implied by Equation (17).

The first panel in Figure 9 plots a binned scatterplot of the log Moneyline ratio $(\frac{1+y_h}{1+y_a})$ against $\log\left(\frac{b_h}{b_a}\right)$ (the "log bet size ratio") where games are grouped into twenty equally spaced bins based on the log Moneyline ratio. The points in dark blue correspond with the Moneyline ratio estimated using observed Moneyline payoffs and the bet size ratio implied by Equation (17). The plot shows that the bet size ratio is decreasing in the Moneyline ratio, where bettors tend to wager more on the home team the more favored the home team is (as higher Moneyline ratios correspond with less favored home teams). The points in orange are the corresponding model implied values, using the parameter values $(\alpha, \gamma, \lambda) = (0.65, 0.65, 1)$. The model-implied values capture the decreasing relationship between the Moneyline ratio and the bet size ratio. In the region of games where the home team is favored (the log Moneyline ratio is negative), the model-implied values also closely match the slope of the relationship between the bet size ratio and the Moneyline ratio in the data. However, when the home team is less favored, the model-implied values display a steeper slope than in the data. The model-implied values for the Moneyline payoffs and bet size ratio also cover a wider range of values than those observed in the data.

The second panel in Figure 9 plots a binned scatterplot of the log bet size ratio estimated using the betting volume data against the model-implied bet size ratio for each game. The plot reveals that the model-implied bet sizes are monotonically increasing with the bet size ratio, and correspond reasonably well with the bet size ratio estimated from the volume data, with points lying very close to the 45 degree line.

One limitation of our calculation is that it requires the assumption that bookmakers take no risk in betting markets. This assumption is based on the idea that bookmakers primarily seek to profit from the "vig", or implicit transaction cost embedded in the prices of bets. This assumption appears to be true on average in practice (Moskowitz (2021)), though it need not hold in each game. To test whether this assumption matters and provide a robustness test of our results using a completely out-of-sample dataset, we obtain betting contracts from another betting market – Betfair – that is an exchange and has no bookmaker. Using betting data for a sample of soccer matches from 2006 to 2011 from Betfair, one of the world's largest betting exchanges, we measure the average wager size for a sample of games without having to make any assumptions about bookmaker behavior.[37] This test not only rules out the influence of bookmakers or assumptions about their behavior, but also provides a completely out-of-sample test on independent data from a completely different sport from another venue.

The first panel of Figure 9 adds a plot of the binned scatterplot of the log Moneyline ratio versus

[37]Details on the sample of games in the Betfair sample are described in Appendix D.3. We thank Angie Andrikogiannopoulou for graciously providing these data.

the log bet size ratio in the Betfair sample in light blue. The plot shows that the same decreasing relationship between the bet size ratio and Moneyline ratio persists in the Betfair sample. Moreover, the slope of the relationship between bet sizes and Moneylines in the Betfair sample mirrors the slope of the relationship implied by the model even more closely than our data. Hence, to the extent bookmakers may take some limited risk to exploit bettor preferences, the Betfair sample, which does not have a bookmaker, may more accurately capture the average size of wagers made by bettors. The Betfair results provide strong out-of-sample evidence consistent with the model's predictions.

Figure 9 suggests that the model qualitatively and quantitatively captures the fact that bettors optimally make larger wagers on more favored teams. This finding is notable given that accurately capturing the bet size is not explicitly part of the objective function of the model. The ability of our model to capture patterns in the size of the average wager across games, in addition to pricing patterns of the contracts, provides additional evidence in favor of our model's ability to explain the data.

6 Conclusion

We use unique data in betting markets on two contingent claims written on the same game outcome to infer bettor preferences and beliefs. The Favorite-Longshot bias in betting returns disappears when comparing equal-risk contracts across the same games. When cross-sectional differences in risk are removed, we find no difference in returns, highlighting that risk, and not any other characteristic of the game or teams, is the chief attribute driving the FLB. The results suggest that bettor preferences and not erroneous beliefs, are likely the primary driver of the well-known FLB.

We use these inferences to connect the FLB in betting markets with low-risk anomalies in financial markets. Quantitatively comparing the returns in betting markets with the low risk anomalies in financial markets, we find very similar magnitudes.

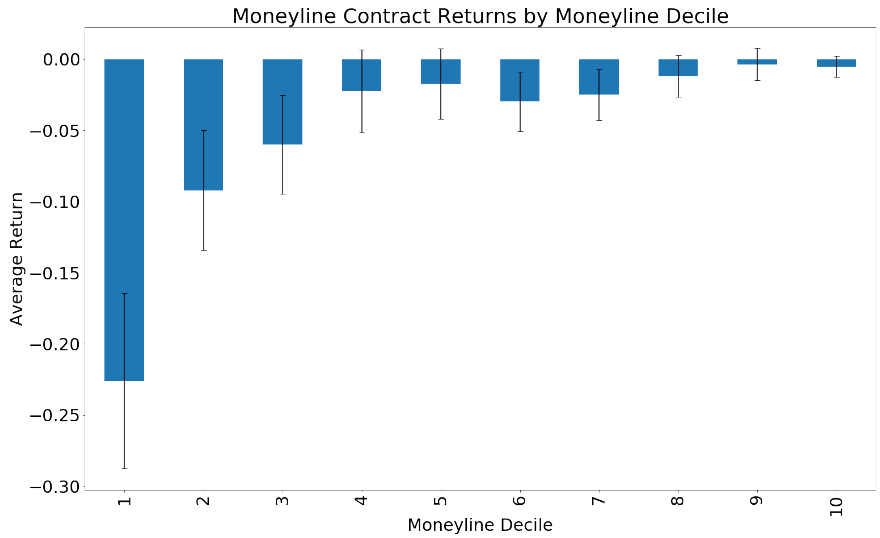
Synthesizing these results informs which theories are able to simultaneously explain the facts from both markets. Rational theories, whether they rely on capital market frictions or the presence of undiversifiable risk, predict no relationship between idiosyncratic risk and returns, and cannot explain our betting market results. Calibrating a model with reference-dependent preferences to match the data, we find that similar parameter values for diminishing sensitivity and probability weighting used to match the low risk anomaly in equity and options markets are simultaneously able to rationalize the betting market evidence. Alternative models, such as heterogeneous beliefs or misperceptions of probabilities, have more difficulty explaining the facts, particularly the unique betting volume data we examine. Overall, we find that non-expected utility preferences likely play an important role in the FLB and potentially low risk anomalies more generally.

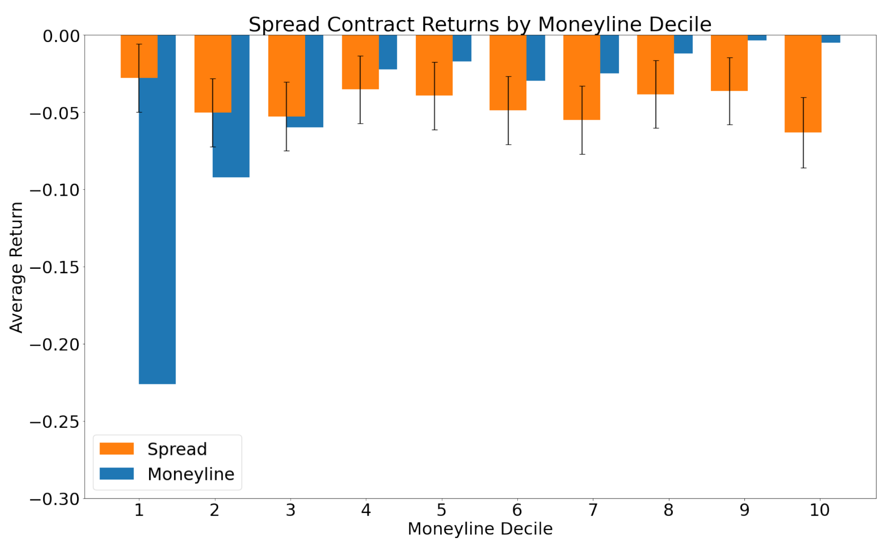
Using betting market data in connection with financial market evidence is helpful to identify explanations that can fit both markets when seeking a unifying explanation for investor decision making. While these markets differ along several dimensions, their commonality and, in particular, the commonality of return patterns, suggests that we can learn something interesting about decision making under uncertainty from examining both markets simultaneously.

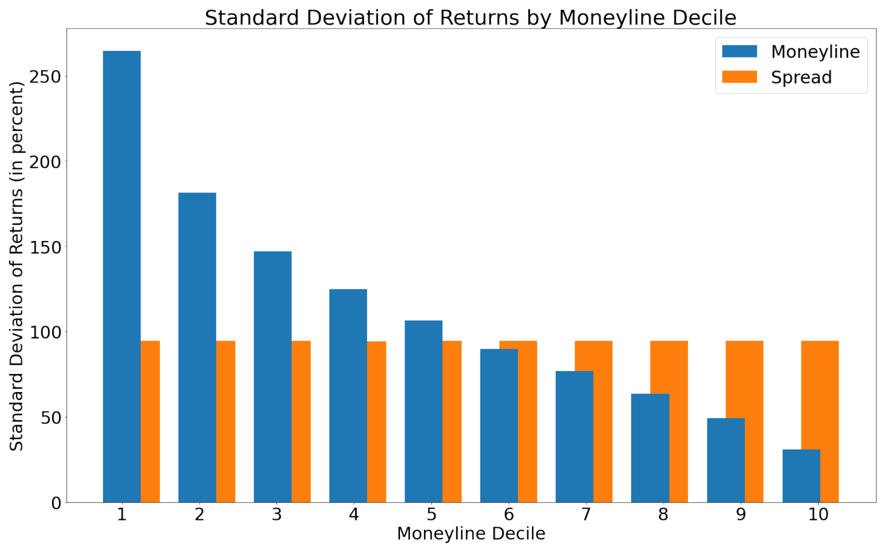
Tables and Figures

Figure 1: The Favorite-Longshot Bias in Betting Contracts

The figure presents statistics on the returns to betting on Favorites and Underdogs in Moneyline and Spread contracts. Both panels plot the average returns of contracts sorted into deciles based on the Moneyline. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. The first panel plots Moneyline contract returns. In the second panel, the darker bars in the foreground correspond with Spread contract returns, and the lighter bars correspond with Moneyline contract returns (as plotted in the top panel). The error bars correspond with 95% confidence intervals for the average return within a given decile. The third panel plots the realized standard deviation of betting contract returns and the fourth panel plots the realized skewness of betting contract returns within each decile; in both panels, the darker bars in the foreground correspond with Moneyline contract returns and the lighter bars in the background correspond with Spread contract returns.






Standard Deviation of Returns by Moneyline Decile
Moneyline
Spread
Standard Deviation of Returns (in percent)
0
50
100
150
200
250
1
2
3
4
5
6
7
8
9
10
Moneyline Decile

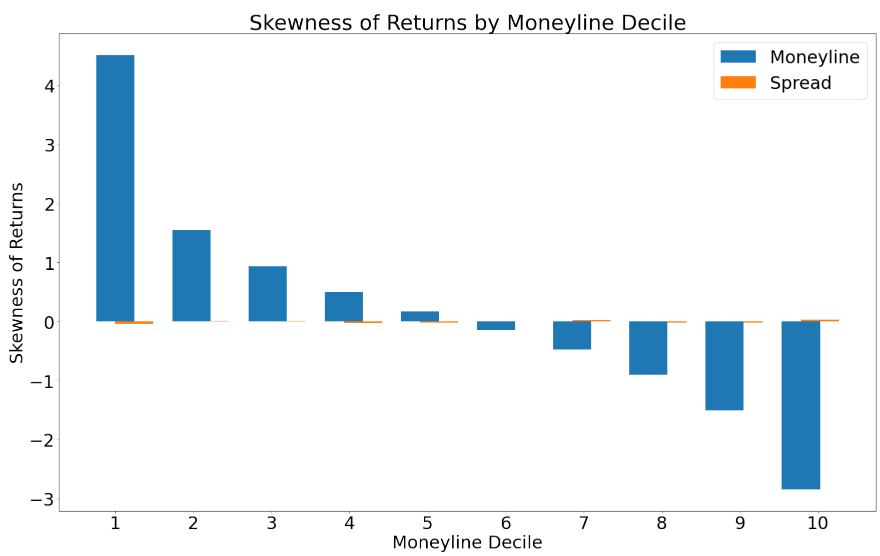

Skewness of Returns by Moneyline Decile
Moneyline
Spread
Skewness of Returns
4
3
2
1
0
−1
−2
−3
1
2
3
4
5
6
7
8
9
10
Moneyline Decile

Figure 2: Accuracy of Over/Under Contracts

The first panel in the figure is a binned scatter plot of the Over/Under line versus the total number of points scored in a game. Each game in our sample is sorted into 20 equal sized bins based on the Over/Under line of the game. Each point on the plot corresponds with the average Over/Under line and the average point total of each game in one of the bins. The 45 degree line is also plotted on the graph in red. The second panel in the figure sorts each game into deciles based on the quantity $\log\left(\frac{1+y_h}{1+y_a}\right)$ ("the Moneyline Ratio"), where y_h and y_a are the payoffs associated with winning Moneyline bets on the home and away teams, and plots the average total number of points scored and Over/Under line in each decile. The error bars correspond with plus/minus two standard deviations relative to the mean.

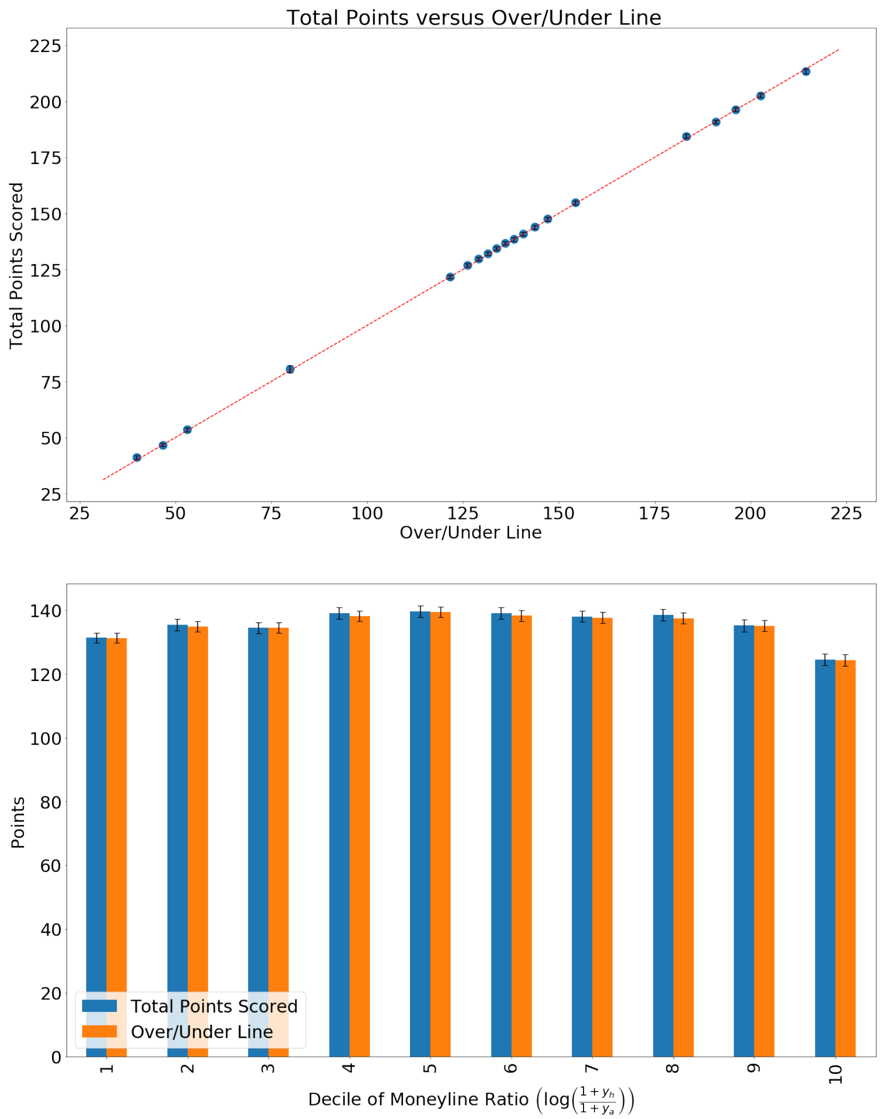


Figure 3: Comparison of Market Share Measures

The figure plots a binned scatterplot of the proportion of bets placed on a Moneyline contract in a game versus the estimated dollar proportion of bets placed on the contract. The estimated dollar proportion of bets is constructed by comparing the payoffs offered on the two Moneyline contracts on a game, using the market clearing conditions in Equations (B.1) and (B.2) which assume that bookmakers take no risk on a game. All observations are grouped into twenty bins based on the estimated dollar proportion of bets. The scatterplot plots the average proportion of bets versus the average estimated dollar proportion of bets for each bin. The figure also plots the 45 degree line in red.

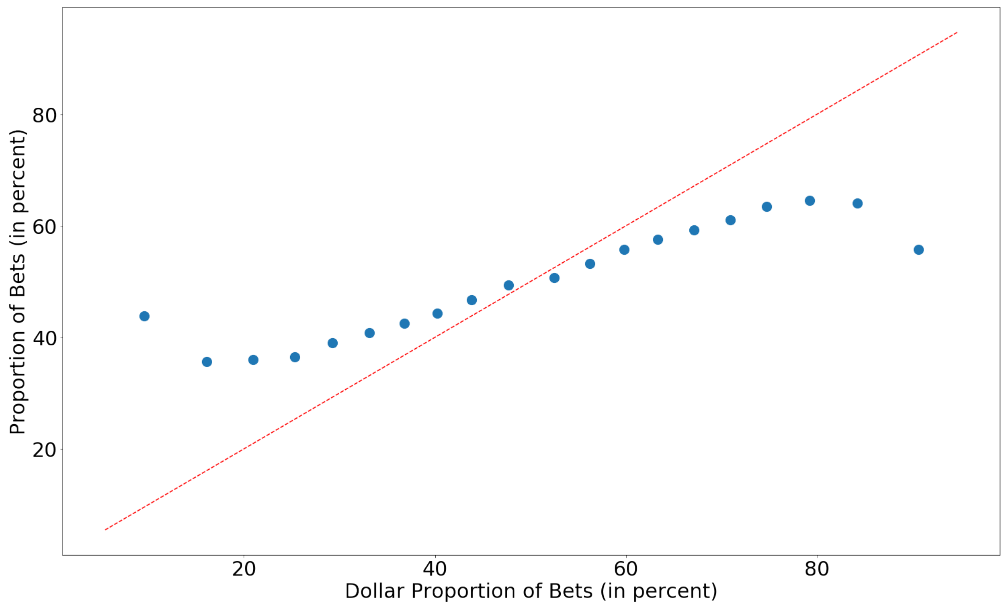


Figure 4: Comparison with MLB and NHL Evidence, Moneyline Contracts

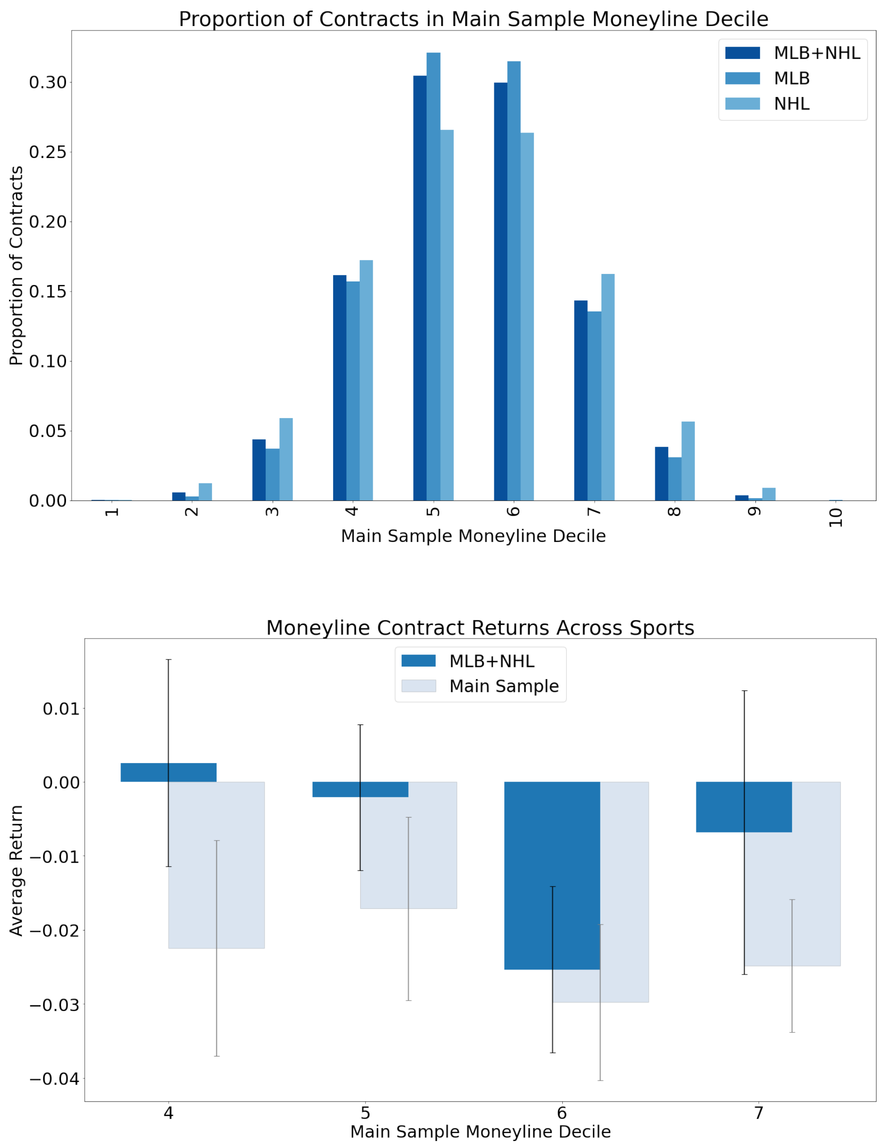


FIGURE 5: IMPLIED VOLATILITY SURFACE BY SPORT

The figure plots the Implied Volatility Surface for betting contracts, with one panel for each sport. For each game, we calculate the the win probabilities non-parametrically. The “implied win probability” for a contract is calculated as the probability that makes the contract’s expected return equal to a Spread contract with a 50% change of paying off. The implied volatility for a contract is the value σhat that satisfies Equation (5), using the win probability and implied probability of the contract. All contracts are placed into one of 90 equally spaced bins based on the win probability. Each plot plots a scatterplot of the average implied volatility against the average win probability in each bin. Each plot also has a dotted red horizontal line, which corresponds with the sample Maximum Likelihood Estimate of the standard deviation of the point-differential minus Spread line within each sample.

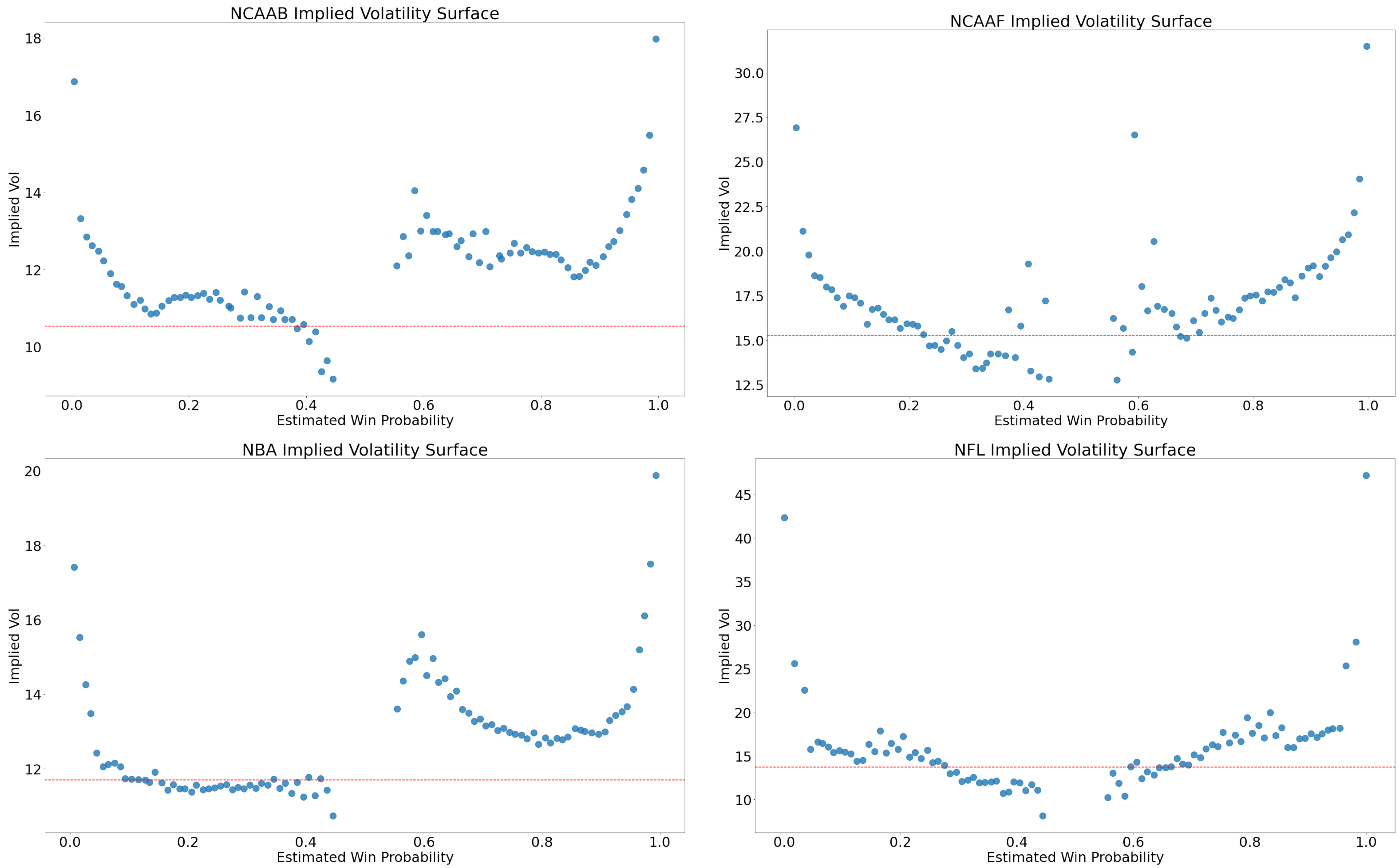


Figure 6: Options Returns and Betting Contract Returns

The figure plots average returns for Moneyline and call option contracts, sorted into five bins. For betting contracts, the contracts are sorted into bins based on estimated payoff probability. For options, the contracts are sorted into bins based on option delta. Options contracts are one-month maturity contracts held until expiration and delta hedged daily. The numbers are taken from Frazzini and Pedersen (2022). For options contracts, average returns are reported separately for index options, as well as for all single-name equity options in the OptionMetrics database.

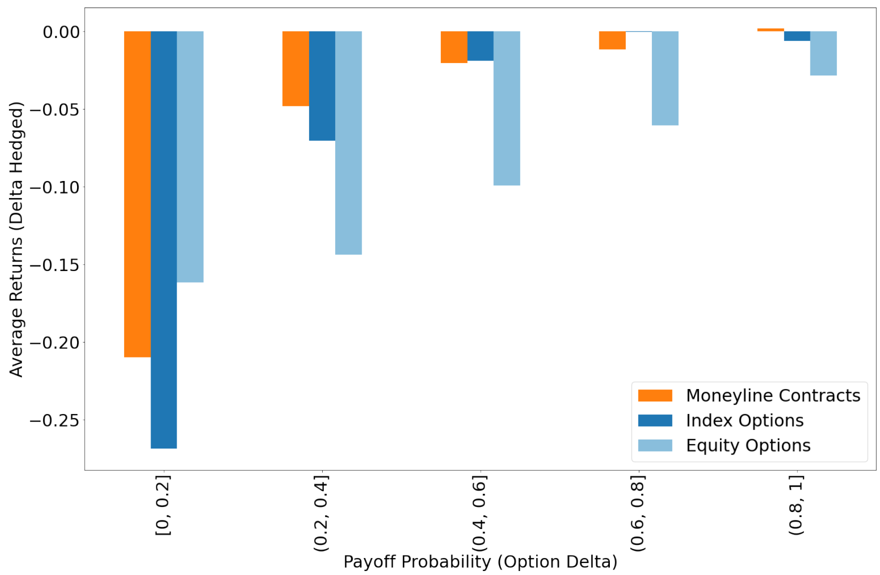


Figure 7: Preference-Implied Returns and Expected Returns

The figure plots the average returns and preference-implied expected returns of Moneyline contracts sorted into deciles based on the Moneyline. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. Model-implied expected returns are estimated by assuming a Cumulative Prospect Theory bettor that is indifferent between betting on each contract offered and a spread contract. The parameter values used are $(\alpha, \gamma, \lambda) = (0.65, 0.65, 1)$.

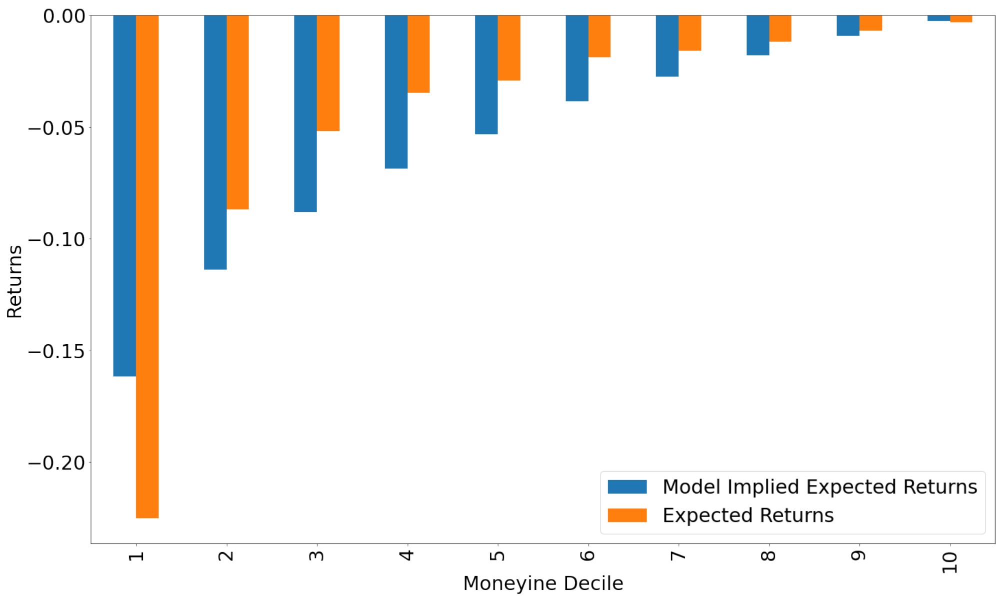


Figure 8: The Probability Weighting Function

The figure plots objective probabilities p, on the x-axis, versus weighted probabilities, $w(p)$ on the y-axis. The lines plotted correspond with $w(p) = p$ and the Tversky and Kahneman (1992) probability weighting function, $w(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{1/\gamma}}$, for $\gamma = 0.65$ and $\gamma = 0.5$.

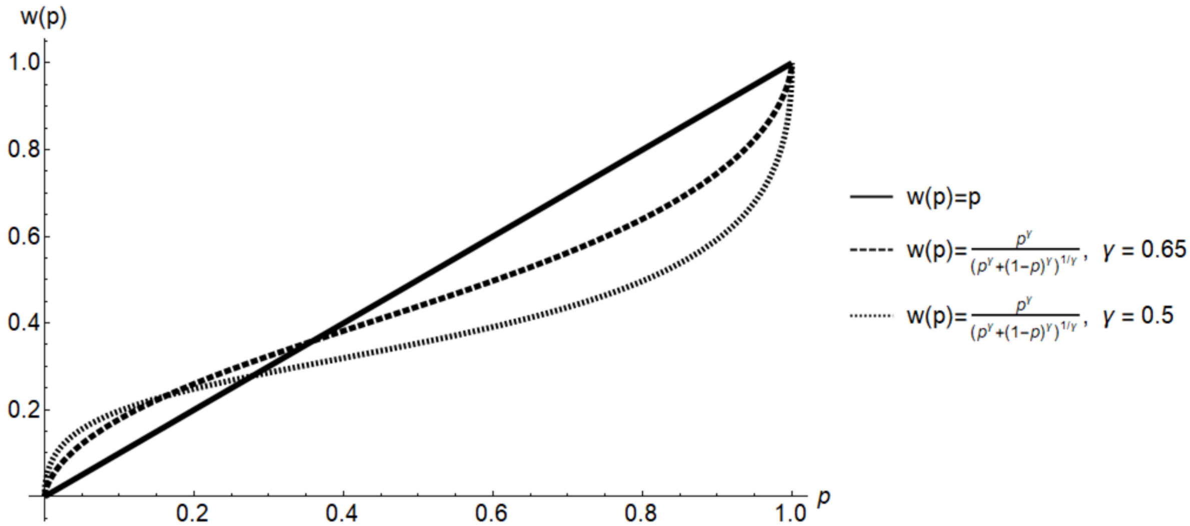


Figure 9: Average Bet Sizes

The first panel in the figure is a binned scatter plot of the log ratio of $\log\left(\frac{1+y_h}{1+y_a}\right)$ (the "Moneyline Ratio") versus the estimated average bet size for bets placed on the home team relative to the average bet size for bets placed on the away team (the "bet size ratio"), where y_h and y_a are the payoffs associated with winning Moneyline bets on the home and away teams at close. Each game is sorted into one of twenty equally sized bins based on the bet size ratio of the game. Each point corresponds with the average log bet size ratio and average Moneyline ratio within a bin. The dots in blue correspond with values calculated using the proportion of bets (estimated using Equation (17)) and empirically observed payoffs offered on games, the dots in orange correspond with values implied by the Cumulative Prospect Theory model estimated in Section 5, and the dots in light blue correspond with values calculated using the observed average bet sizes and payoffs for a sample of soccer games from Betfair (described in Appendix D.3). The second panel is a binned scatterplot of the bet size ratio from the data versus the bet size ratio implied by the model, where the bins are once again constructed by sorting games into one of twenty equally sized bins.

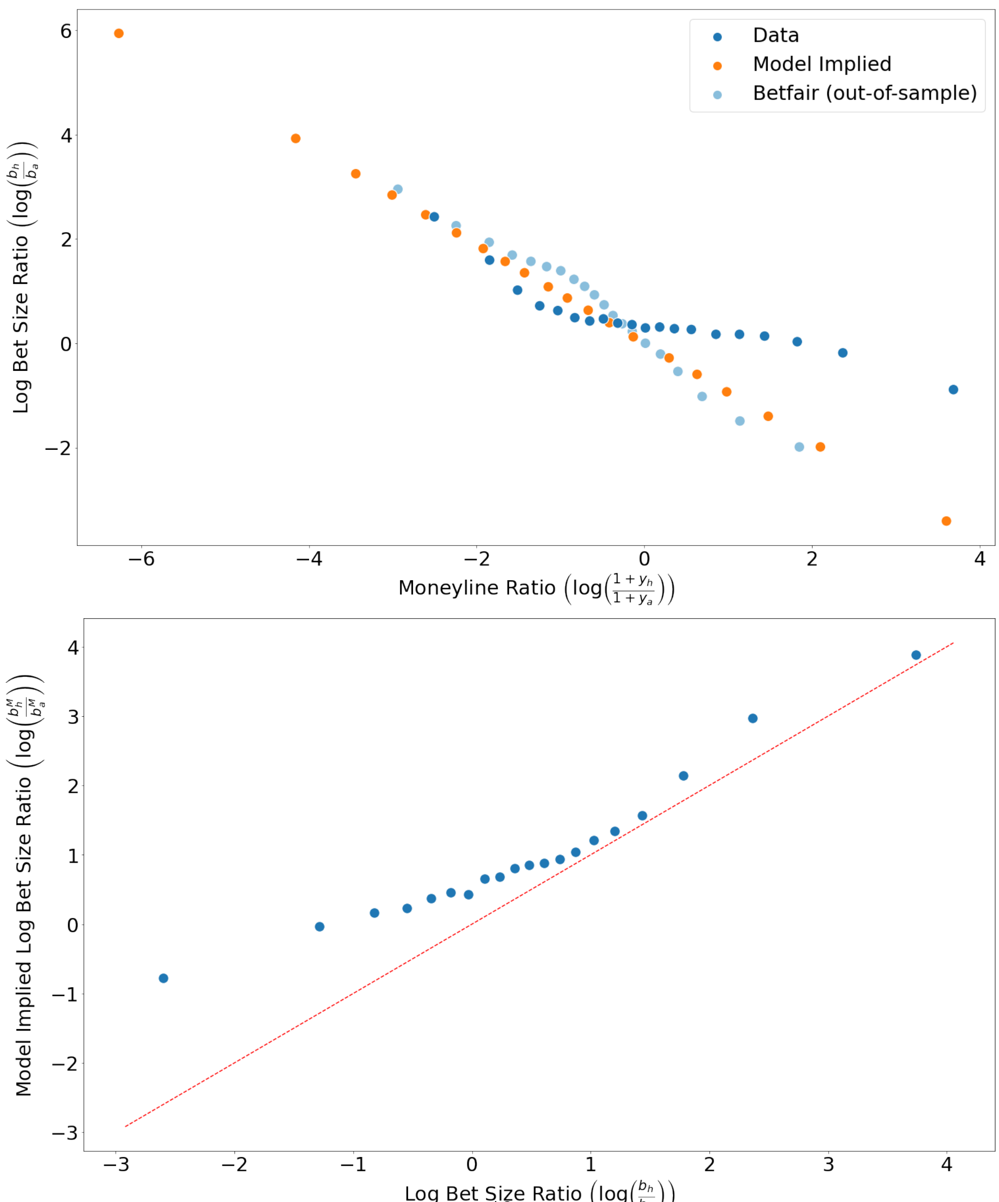


Table 1: Summary Statistics for Betting Contracts

The table reports summary statistics on the betting contracts in the sample for contracts that are sorted into decile based on the Moneyline, with statistics reported separately for Moneyline and Spread contracts. The table reports the betting line, the average return, the standard deviation of returns, and the skewness of betting contracts for each decile, and the $F-$statistic from a test of equality of average returns across the deciles. The table also reports the average ex-ante return standard deviation and ex-ante skewness of contract returns. For each moneyline contract, these statistics are computed for using the payoffs of the contracts and the win probabilities for the contract estimated following the methodology described in Appendix A. For spread contracts, these statistics are computed for each contract assuming a 50% win probability for the contract and a potential payoff of $100 for a $110 wager. The table also presents the average estimated win probability and the realized win percentage for teams in each decile, along with the number of contracts in the decile.

		Decile of Moneyline Contract									
		1	2	3	4	5	6	7	8	9	10
Game Characteristics	Estimated Win Probability	9%	20%	29%	38%	46%	55%	62%	71%	80%	91%
	Actual Win Percentage	9%	20%	29%	38%	46%	54%	62%	71%	80%	91%
	N	7,349	7,439	7,178	7,346	7,386	7,256	7,317	7,504	7,591	6,852
Moneyline Contracts	Line	1003.68	355.96	224.25	157.06	90.85	-125.91	-173.93	-255.09	-430.70	-1570.88
	Realized Return Std	264%	182%	147%	125%	106%	90%	77%	64%	49%	31%
	Average Ex-Ante Return Std	257%	181%	147%	124%	106%	90%	76%	63%	49%	29%
	Realized Return Skew	4.51	1.55	0.94	0.50	0.17	-0.15	-0.48	-0.90	-1.51	-2.84
	Average Ex-Ante Return Skew	4.47	1.50	0.94	0.51	0.18	-0.18	-0.51	-0.95	-1.55	-4.65
	Average Return	-22.60%	-9.21%	-5.98%	-2.25%	-1.71%	-2.98%	-2.48%	-1.18%	-0.35%	-0.51%
	F-Statistic: 19.48 (p $<$ 0.00)										
Spread Contracts	Line	14.64	8.87	5.98	3.77	1.69	-1.74	-3.77	-6.04	-9.07	-14.93
	Realized Return Std	95%	95%	94%	94%	94%	94%	94%	94%	95%	95%
	Average Ex-Ante Return Std	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%
	Realized Return Skew	-0.04	0.01	0.01	-0.02	-0.01	0.01	0.02	-0.02	-0.02	0.04
	Average Ex-Ante Return Skew	0	0	0	0	0	0	0	0	0	0
	Average Return	-2.77%	-5.03%	-5.28%	-3.54%	-3.93%	-4.88%	-5.52%	-3.84%	-3.62%	-6.32%
	F-Statistic: 0.97 ($p = 0.46$)										

Table 2: Implied Volatility Premium: Sports Betting Contracts and Options

The table presents data on the implied volatilities in sports betting contracts and options contracts. The first panel of the table presents the sample standard deviation of the point-spread minus the spread line, and the average implied volatility of Moneyline contracts expected to pay off 5%-15% and 85%-95% of the time within each sport. The payoff probabilities are estimated nonparametrically, as described in Appendix A. The last two rows of the panel correspond with the Implied Volatility premia, as the percent different between the average implied volatilities and the sample volatility. Panel B presents the average implied volatility for standardized, one month maturity 10, 50, and 90 delta call options and -10, -50, -90 delta put options across a set of 13 indices. The data are from the OptionMetrics Implied Volatility surface file, and are calculated by interpolating options of different maturities and deltas. Panel C presents the average implied volatility for standardized, one month maturity 10, 50, and 90 delta call and -10, -50, -90 delta put options for all equities for which data is available in the OptionMetrics Implied Volatility surface file. The last two rows of Panels B and C present the "IV Premium" for call and put options, which we define as the percent difference between an option and the corresponding 50 delta option.

Panel A: Sports Betting Contracts

	NCAAB	NCAAF	NBA	NFL	Average
Sample Volatility	10.5	15.3	11.7	13.7	-
IV for 5%-15% WP Contract	11.3	17.2	11.9	15.4	-
IV for 85%-95% WP Contract	12.4	18.9	13.1	17.1	-
IV Premium for 5%-15% WP	7.6%	12.8%	1.4%	12.4%	8.6%
IV Premium for 85%-95% WP	17.4%	23.7%	12.4%	24.8%	19.6%

Panel B: Index Options

	Call Options	Put Options	Average
IV for 50 Delta	0.22	0.21	0.21
IV for 10 Delta	0.19	0.28	0.24
IV for 90 Delta	0.30	0.19	0.24
IV Premium for 10 Delta	-4.5%	43.6%	19.6%
IV Premium for 90 Delta	48.0%	-0.9%	23.6%

Panel C: Equity Options

	Call Options	Put Options	Average
IV for 50 Delta	0.46	0.47	0.46
IV for 10 Delta	0.54	0.59	0.56
IV for 90 Delta	0.55	0.52	0.54
IV Premium for 10 Delta	17.0%	26.7%	21.9%
IV Premium for 90 Delta	21.1%	12.4%	16.7%

Table 3: Spread lines, Moneylines, and Predicting Wins

Panel A of the table presents regression results from OLS regressions where the independent variable is a 0/1 indicator variable capturing whether the home team won the game and the independent variables are the closing Spread line of the game and $\log\left(\frac{1+y_h}{1+y_a}\right)$ (the "Moneyline Ratio"), where y_h and y_a are the payoffs associated with winning Moneyline bets on the home and away teams at close. Panel B reports results from regressions of changes in the Moneyline ratio from open to close of betting on changes in the Spread line from open to close of betting for each sport. Independent variables are standardized to have zero mean and unit standard deviation within each sport, and separate regression coefficients are estimated by sport, with t-statistics are reported in parentheses.

Panel A: Wins, Spreads, and Moneylines

	Dependent variable = $\mathbb{1}_{\text{home team wins}}$		
	(1)	(2)	(3)
$\beta_{\text{ML,NCAAB}}$	-0.22		0.01
	(-76.49)		(0.36)
$\beta_{\text{ML,NCAAF}}$	-0.26		-0.14
	(-40.01)		(-2.40)
$\beta_{\text{ML,NBA}}$	-0.19		-0.04
	(-41.12)		(-1.33)
$\beta_{\text{ML,NFL}}$	-0.18		-0.11
	(-17.68)		(-1.27)
$\beta_{\text{Spread,NCAAB}}$		-0.22	-0.24
		(-76.98)	(-7.78)
$\beta_{\text{Spread,NCAAF}}$		-0.26	-0.12
		(-40.04)	(-2.06)
$\beta_{\text{Spread,NBA}}$		-0.20	-0.16
		(-41.57)	(-5.91)
$\beta_{\text{Spread,NFL}}$		-0.18	-0.07
		(-17.68)	(-0.80)
Sport FE	Yes	Yes	Yes
N	36,609	36,609	36,609
R^2	20.66%	20.85%	20.87%
F-statistic	1361.37	1377.75	877.72

Panel B: Changes in Spread Lines and Moneylines

	Dependent variable = ΔMoneyline Ratio$_{\text{open-to-close}}$			
	NCAAB	NCAAF	NBA	NFL
ΔSpread Line$_{\text{open-to-close}}$	0.61	0.53	0.63	0.62
	(114.49)	(41.31)	(73.49)	(34.08)
N	21,742	4,363	8,390	1,843
R^2	0.38	0.28	0.39	0.39

Table 4: Betting Market and Financial Market Low-Risk Trading Strategies

The table reports the average annualized return, t-statistic, and annualized Sharpe ratio for low-risk trading strategies in betting markets and financial markets. All trading strategies are scaled to realize 10% annualized volatility ex-post, using the in-sample return volatility of the strategy. Panel A of the table reports statistics for betting market trading strategies. Each column in the table corresponds with a trading strategy that takes a long position in one betting contract and an equal notional short position in another betting contract on the game, investing an equal amount across all the games in the sample. Columns are labeled as $A - B$, where A is the long leg of the trading strategy and B is the short leg of the trading strategy; ML^F and ML^U are the Moneyline contracts on the favorite and underdog, and S^F and S^U are the Spread contracts on the favorite and underdog. Trading strategy statistics are annualized by using the average number of games per year for years with complete data (5504 games). Panel B of the table reports statistics for financial market strategies. The first two columns of the table correspond with Options Betting-Against-Beta (BAB) strategies, using the numbers reported in Table VI of Frazzini and Pedersen (2022) for equity and index options trading strategies. The third and fourth columns of the table correspond with the excess returns of US and Global BAB strategies in equities from Frazzini and Pedersen (2014), using strategy returns from the AQR data library updated through February 2022. The last two columns of the table correspond with the idiosyncratic volatility trading strategy from Ang et al. (2006), using US and Developed Market data from the Jensen, Kelly and Pedersen (2021) data library through December 2021. For the idiosyncratic volatility strategies, the reported statistics correspond with residual alpha from a regression of the trading strategy returns on the Fama and French (1993) three factors.

Panel A: Betting Market Strategies

	ML^F-ML^U	S^F-S^U	S^F-ML^U	S^U-ML^U	ML^F-S^F	ML^F-S^U
Average annualized return	22.70	-2.33	11.53	21.51	35.34	13.79
t-statistic	5.85	-0.60	2.97	5.55	9.11	3.56
Annualized Sharpe ratio	2.27	-0.23	1.15	2.15	3.53	1.38

Panel B: Financial Market Strategies

	Options BAB		Equities BAB		Idiosyncratic Vol	
	Single Name	Index	US	Global	US	Global
Average annualized return	17.80	11.40	7.28	9.13	7.12	9.40
t-statistic	8.51	5.47	6.96	5.41	5.86	5.19
Annualized Sharpe ratio	1.78	1.14	0.73	0.91	0.71	0.94

Table 5: Model Estimates

The table presents details of the estimation of the main preference specification, calculated by fixing a value of γ and minimizing Equation (14). Panel A presents the estimate of α, the number of contracts (out of 73218) for which the model is able to calculate payoffs that satisfy the equilibrium and optimality conditions, and the mean squared error of expected returns from the model, with standard errors from 2000 bootstrap samples reported in parentheses. Panel B groups observations into deciles based on the Moneyline, and presents the average log payoff ($\log y_i$) and average expected returns of contracts in the decile, as well as the corresponding values output from the model (super-scripted by M) for the assumed value of γ. The panel also includes the proportion of contracts for which the model is able to calculate implied payoffs in each decile.

Panel A: Model Estimate Details

γ=0.65			γ=0.5		
α	Valid Bets	MSE	α	Valid Bets	MSE
0.650	73,188	0.012	0.502	73,187	0.012
(0.0004)	(49.4)	(0.005)	(0.002)	(53.9)	(0.005)

Panel B: Model Estimates By Decile

γ=0.65							γ=0.5						
Decile	ML	$\log y_i$	$\log y_i^M$	$E(r_i)$	$E(r_i)^M$	Valid	Decile	ML	$\log y_i$	$\log y_i^M$	$E(r_i)$	$E(r_i)^M$	Valid
1	1003.68	2.15	2.36	-22.5%	-16.2%	100.0%	1	1003.68	2.15	2.33	-22.5%	-18.4%	100.0%
2	355.96	1.26	1.22	-8.7%	-11.4%	100.0%	2	355.96	1.26	1.21	-8.7%	-12.1%	100.0%
3	224.25	0.80	0.75	-5.2%	-8.8%	100.0%	3	224.25	0.80	0.74	-5.2%	-9.1%	100.0%
4	157.06	0.45	0.39	-3.5%	-6.9%	100.0%	4	157.06	0.45	0.39	-3.5%	-7.0%	100.0%
5	90.85	0.12	0.08	-2.9%	-5.3%	100.0%	5	90.85	0.12	0.08	-2.9%	-5.4%	100.0%
6	-125.91	-0.23	-0.27	-1.9%	-3.8%	100.0%	6	-125.91	-0.23	-0.27	-1.9%	-3.8%	100.0%
7	-173.93	-0.55	-0.58	-1.6%	-2.8%	100.0%	7	-173.93	-0.55	-0.58	-1.6%	-2.7%	100.0%
8	-255.09	-0.93	-0.95	-1.2%	-1.8%	100.0%	8	-255.09	-0.93	-0.95	-1.2%	-1.8%	100.0%
9	-430.70	-1.44	-1.46	-0.7%	-0.9%	100.0%	9	-430.70	-1.44	-1.46	-0.7%	-0.9%	100.0%
10	-1570.88	-2.52	-2.61	-0.3%	-0.3%	99.6%	10	-1570.88	-2.52	-2.61	-0.3%	-0.3%	99.6%

References

Ali, Mukhtar M. 1977. "Probability and utility estimates for racetrack bettors." *Journal of Political Economy*, 85(4): 803–815.

Andrikogiannopoulou, Angie, and Filippos Papakonstantinou. 2016. "Heterogeneity in risk preferences: Evidence from a real-world betting market." *Swiss Finance Institute Research Paper*, , (13-53).

Andrikogiannopoulou, Angie, and Filippos Papakonstantinou. 2019. "History-dependent risk preferences: Evidence from individual choices and implications for the disposition effect." *Review of Financial Studies, forthcoming.*

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2006. "The cross-section of volatility and expected returns." *Journal of Finance*, 61(1): 259–299.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2009. "High idiosyncratic volatility and low returns: International and further US evidence." *Journal of Financial Economics*, 91(1): 1–23.

Asness, Cliff, Andrea Frazzini, Niels Joachim Gormsen, and Lasse Heje Pedersen. 2020. "Betting against correlation: Testing theories of the low-risk effect." *Journal of Financial Economics*, 135(3): 629–652.

Avery, Christopher, and Judith Chevalier. 1999. "Identifying investor sentiment from price paths: The case of football betting." *Journal of Business*, 72(4): 493–521.

Baele, Lieven, Joost Driessen, Sebastian Ebert, Juan M Londono, and Oliver Spalt. 2016. "Cumulative prospect theory, option prices, and the variance premium." *Review of Financial Studies.*

Barberis, Nicholas. 2013. "The psychology of tail events: Progress and challenges." *American Economic Review*, 103(3): 611–16.

Barberis, Nicholas, Abhiroop Mukherjee, and Baolian Wang. 2016. "Prospect theory and stock returns: An empirical test." *Review of Financial Studies*, 29(11): 3068–3107.

Barberis, Nicholas, and Ming Huang. 2008. "Stocks as lotteries: The implications of probability weighting for security prices." *American Economic Review*, 98(5): 2066–2100.

Barberis, Nicholas, and Wei Xiong. 2009. "What drives the disposition effect? An analysis of a long-standing preference-based explanation." *Journal of Finance*, 64(2): 751–784.

Barberis, Nicholas, Lawrence J Jin, and Baolian Wang. 2021. "Prospect theory and stock market anomalies." *The Journal of Finance*, 76(5): 2639–2687.

Barberis, Nicholas, Ming Huang, and Richard H Thaler. 2006. “Individual preferences, monetary gambles, and stock market participation: A case for narrow framing.” *American Economic Review*, 96(4): 1069–1090.

Barberis, Nicholas, Ming Huang, and Tano Santos. 2001. “Prospect theory and asset prices.” *Quarterly Journal of Economics*, 116(1): 1–53.

Barseghyan, Levon, Francesca Molinari, Ted O’Donoghue, and Joshua C Teitelbaum. 2013. “The nature of risk preferences: Evidence from insurance choices.” *American Economic Review*, 103(6): 2499–2529.

Barseghyan, Levon, Francesca Molinari, Ted O’Donoghue, and Joshua C Teitelbaum. 2018. “Estimating risk preferences in the field.” *Journal of Economic Literature*, 56(2): 501–64.

Benartzi, Shlomo, and Richard H Thaler. 1995. “Myopic loss aversion and the equity premium puzzle.” *Quarterly Journal of Economics*, 110(1): 73–92.

Bernheim, B Douglas, and Charles Sprenger. 2020. “On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting.” *Econometrica*, 88(4): 1363–1409.

Black, Fischer. 1972. “Capital market equilibrium with restricted borrowing.” *Journal of Business*, 45(3): 444–455.

Black, Fischer, Michael C Jensen, Myron Scholes, et al. 1972. “The capital asset pricing model: Some empirical tests.” *Studies in the Theory of Capital Markets*, 81(3): 79–121.

Bondarenko, Oleg. 2014. “Why are put options so expensive?” *Quarterly Journal of Finance*, 4(03): 1450015.

Booij, Adam S, Bernard MS Van Praag, and Gijs Van De Kuilen. 2010. “A parametric analysis of prospect theory’s functionals for the general population.” *Theory and Decision*, 68(1-2): 115–148.

Boyer, Brian H, and Keith Vorkink. 2014. “Stock options as lotteries.” *Journal of Finance*, 69(4): 1485–1527.

Bradley, Ian. 2003. “The representative bettor, bet size, and prospect theory.” *Economics Letters*, 78(3): 409–413.

Brown, William O, and Raymond D Sauer. 1993. “Does the basketball market believe in the hot hand? Comment.” *American Economic Review*, 83(5): 1377–1386.

Brunnermeier, Markus K, Christian Gollier, and Jonathan A Parker. 2007. “Optimal beliefs, asset prices, and the preference for skewed returns.” *American Economic Review*, 97(2): 159–165.

Camerer, Colin F. 1989. "Does the basketball market believe in the hot hand?" *American Economic Review*, 79(5): 1257–1261.

Chapman, Jonathan, Erik Snowberg, Stephanie Wang, and Colin Camerer. 2018. "Loss attitudes in the US population: Evidence from dynamically optimized sequential experimentation (DOSE)." National Bureau of Economic Research.

Chiappori, Pierre-André, Bernard Salanié, François Salanié, and Amit Gandhi. 2019. "From aggregate betting data to individual risk preferences." *Econometrica*, 87(1): 1–36.

Conlisk, John. 1993. "The utility of gambling." *Journal of Risk and Uncertainty*, 6(3): 255–275.

Cox, John C, Stephen A Ross, and Mark Rubinstein. 1979. "Option pricing: A simplified approach." *Journal of Financial Economics*, 7(3): 229–263.

Dorn, Daniel, and Paul Sengmueller. 2009. "Trading as entertainment?" *Management Science*, 55(4): 591–603.

Durham, Gregory R, Michael G Hertzel, and J Spencer Martin. 2005. "The market impact of trends and sequences in performance: New evidence." *Journal of Finance*, 60(5): 2551–2569.

Fama, Eugene F, and Kenneth R French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33(1): 3–56.

Fehr-Duda, Helga, and Thomas Epper. 2012. "Probability and risk: Foundations and Economic implications of probability-dependent risk preferences." *Annu. Rev. Econ.*, 4(1): 567–593.

Frazzini, Andrea, and Lasse Heje Pedersen. 2014. "Betting against beta." *Journal of Financial Economics*, 111(1): 1–25.

Frazzini, Andrea, and Lasse Heje Pedersen. 2022. "Embedded leverage." *The Review of Asset Pricing Studies*, 12(1): 1–52.

Gandar, John M, William H Dare, Craig R Brown, and Richard A Zuber. 1998. "Informed traders and price variations in the betting market for professional basketball games." *Journal of Finance*, 53(1): 385–401.

Gandar, John, Richard Zuber, Thomas O'brien, and Ben Russo. 1988. "Testing rationality in the point spread betting market." *Journal of Finance*, 43(4): 995–1008.

Gandhi, Amit, and Ricardo Serrano-Padial. 2015. "Does belief heterogeneity explain asset prices: The case of the longshot bias." *Review of Economic Studies*, 82(1): 156–186.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus. 2021. "Five facts about beliefs and portfolios." *American Economic Review*, 111(5): 1481–1522.

Gil, Ricard Gil Ricard, and Stephen D Levitt. 2007. "Testing the efficiency of markets in the 2002 World Cup." *The Journal of Prediction Markets*, 1(3): 255–270.

Golec, Joseph, and Maurry Tamarkin. 1991. "The degree of inefficiency in the football betting market: Statistical tests." *Journal of Financial Economics*, 30(2): 311–323.

Gormsen, Niels Joachim, and Eben Lazarus. 2021. "Duration-driven returns."

Gray, Philip K, and Stephen F Gray. 1997. "Testing market efficiency: Evidence from the NFL sports betting market." *Journal of Finance*, 52(4): 1725–1737.

Green, Etan, Haksoo Lee, and David M Rothschild. 2020. "The favorite-longshot midas."

Griffith, Richard M. 1949. "Odds adjustments by American horse-race bettors." *American Journal of Psychology*, 62(2): 290–294.

Grinblatt, Mark, and Bing Han. 2005. "Prospect theory, mental accounting, and momentum." *Journal of Financial Economics*, 78(2): 311–339.

Grinblatt, Mark, and Matti Keloharju. 2009. "Sensation seeking, overconfidence, and trading activity." *Journal of Finance*, 64(2): 549–578.

Hudson, James. 2014. "UK Sports Betting Market Growth and Mobile Betting." *Favourit.*

Imas, Alex. 2016. "The realization effect: Risk-taking after realized versus paper losses." *American Economic Review*, 106(8): 2086–2109.

Ingersoll, Jonathan. 2008. "Non-Monotonicity of the Tversky-Kahneman Probability-Weighting Function: A Cautionary Note." *European Financial Management*, 14(3): 385–390.

Jensen, Theis Ingerslev, Bryan T Kelly, and Lasse Heje Pedersen. 2021. "Is there a replication crisis in finance?" *Journal of Finance, Forthcoming.*

Jullien, Bruno, and Bernard Salanié. 2000. "Estimating preferences under risk: The case of racetrack bettors." *Journal of Political Economy*, 108(3): 503–530.

Kliger, Doron, and Ori Levy. 2009. "Theories of choice under risk: Insights from Financial markets." *Journal of Economic Behavior & Organization*, 71(2): 330–346.

Kőszegi, Botond, and Matthew Rabin. 2006. "A model of reference-dependent preferences." *Quarterly Journal of Economics*, 121(4): 1133–1165.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-dependent risk attitudes." *American Economic Review*, 97(4): 1047–1073.

Levitt, Steven D. 2004. "Why are gambling markets organised so differently from Financial markets?" *Economic Journal*, 114(495): 223–246.

Moskowitz, Tobias J. 2021. "Asset pricing and sports betting." *The Journal of Finance*, 76(6): 3153–3209.

Newall, Philip WS, and Dominic Cortis. 2021. “Are Sports Bettors Biased toward Longshots, Favorites, or Both? A Literature Review.” *Risks*, 9(1): 22.

Ni, Sophie Xiaoyan. 2008. “Stock option returns: A puzzle.”

O’Donoghue, Ted, and Charles Sprenger. 2018. “Reference-dependent preferences.” In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 1, 1–77. Elsevier.

Ottaviani, Marco, and Peter Norman Sørensen. 2009. “Surprised by the parimutuel odds?” *American Economic Review*, 99(5): 2129–34.

Ottaviani, Marco, and Peter Norman Sørensen. 2010. “Noise, information, and the favorite-longshot bias in parimutuel predictions.” *American Economic Journal: Microeconomics*, 2(1): 58–85.

Pankoff, Lyn D. 1968. “Market efficiency and football betting.” *Journal of Business*, 41(2): 203–214.

Peta, Joe. 2014. *Trading Bases: How a Wall Street Trader Made a Fortune Betting on Baseball.* Penguin.

Prelec, Drazen. 1998. “The probability weighting function.” *Econometrica*, 497–527.

Quiggin, John. 1982. “A theory of anticipated utility.” *Journal of Economic Behavior & Organization*, 3(4): 323–343.

Sauer, Raymond D, Vic Brajer, Stephen P Ferris, and M Wayne Marr. 1988. “Hold your bets: Another look at the efficiency of the gambling market for National Football League games.” *Journal of Political Economy*, 96(1): 206–213.

Shin, Hyun Song. 1991. “Optimal betting odds against insider traders.” *Economic Journal*, 101(408): 1179–1185.

Shin, Hyun Song. 1992. “Prices of state contingent claims with insider traders, and the favourite-longshot bias.” *Economic Journal*, 102(411): 426–435.

Snowberg, Erik, and Justin Wolfers. 2010. “Explaining the favorite–long shot bias: Is it risk-love or misperceptions?” *Journal of Political Economy*, 118(4): 723–746.

Thaler, Richard H, and William T Ziemba. 1988. “Anomalies: Parimutuel betting markets: Racetracks and lotteries.” *Journal of Economic Perspectives*, 2(2): 161–174.

Tversky, Amos, and Daniel Kahneman. 1979. “Prospect theory: An analysis of decision under risk.” *Econometrica*, 47(2): 263–291.

Tversky, Amos, and Daniel Kahneman. 1992. “Advances in prospect theory: Cumulative representation of uncertainty.” *Journal of Risk and Uncertainty*, 5(4): 297–323.

Vissing-Jørgensen, Annette. 2002. "Limited asset market participation and the elasticity of intertemporal substitution." *Journal of Political Economy*, 110(4): 825–853.

Weinberg, Ari. 2003. "The case for legal sports gambling." *Forbes. com.*

Weitzman, M. 1965. "Utility analysis and group behavior: an experimental study." *Journal of Political Economy*, 73: 18–26.

Woodland, Linda M, and Bill M Woodland. 1994. "Market efficiency and the favorite-longshot bias: the baseball betting market." *Journal of Finance*, 49(1): 269–279.

Woodland, Linda M, and Bill M Woodland. 2001. "Market efficiency and profitable wagering in the national hockey league: Can bettors score on longshots?" *Southern Economic Journal*, 983–995.

Zuber, Richard A, John M Gandar, and Benny D Bowers. 1985. "Beating the spread: Testing the efficiency of the gambling market for National Football League games." *Journal of Political Economy*, 93(4): 800–806.

Internet Appendix for

Betting Without Beta

Tobias J. Moskowitz, Kaushik Vasudevan

May 2, 2022

A Non-Parametric Estimation of Win Probabilities

We describe the procedure to estimate win probabilities for Moneyline contracts used in Section 4.1 and Section 5.

Within each sport, we run a non-parametric regression of the home team contract returns on the Spread line of the game, $\log(1 + y_{h,i})$ and $\log(1 + y_{a,i})$, where $y_{j,i}$ is the payoff associated with a winning dollar bet on team j in game i, h is the home team, and a is the away team. Each regression is run using the `KernelReg` function in the `statsmodels` package in Python. Each regression is a local linear regression using a Gaussian kernel. The bandwidth of the kernel is selected via least-squares cross-validation.

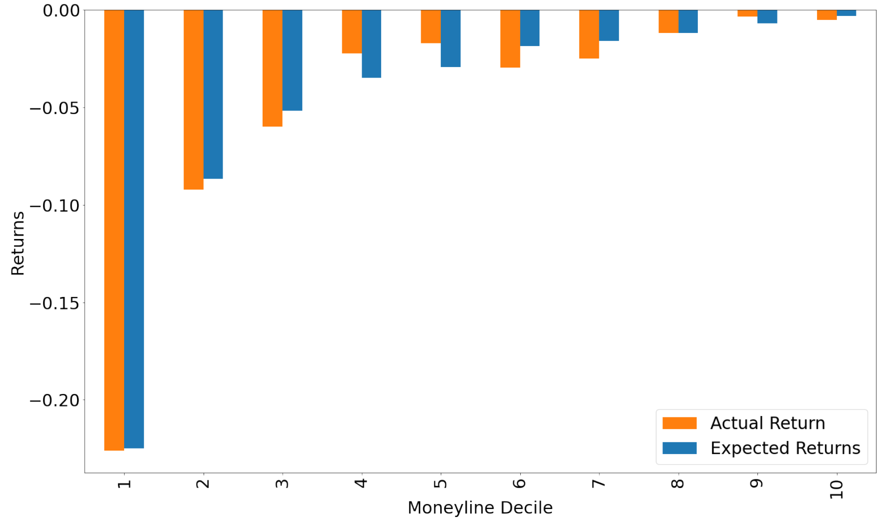
We use the fitted values from the regressions as estimates of the expected returns of the home team. We can write the expected return in terms of p_i, the probability of the home team winning game i, and $y_{h,i}$, where p_i is the value we are interested in estimating.

$$E(r_{h,i}) = p_i y_{h,i} + (1 - p_i)(-1) \tag{A.1}$$

Re-writing Equation (A.1) in terms of p_i, we get $p_i = \frac{E(r_{h,i})+1}{y_{h,i}+1}$. The win probability of the away team is $1 - p_i$. The non-parametric regression used to produce $E(r_{h,i})$ does not have any restrictions that the implied p must be less than 0 or greater than 1, and there are some instances where the implied win probabilities lie outside this range (588 games out of 36,609). We truncate any values of p_i less than 0 to $\underline{p}$ and any values greater than 1 to $\bar{p}$, where $\underline{p}$ and $\bar{p}$ are the minimum and maximum implied values of p within the range $(0, 1)$. We use these truncated values of p_i to re-estimate the expected returns for the home contract, and to form expected returns for the away contract in each game.[38]

[38] An alternative way to proceed is to use the same methodology except that the away team contract returns are the independent variable in the non-parametric regressions, where p_i is estimated as the away team's win probability. This approach yields similar results.

Figure A.1: Moneyline Contract Expected Returns v. Realized Returns



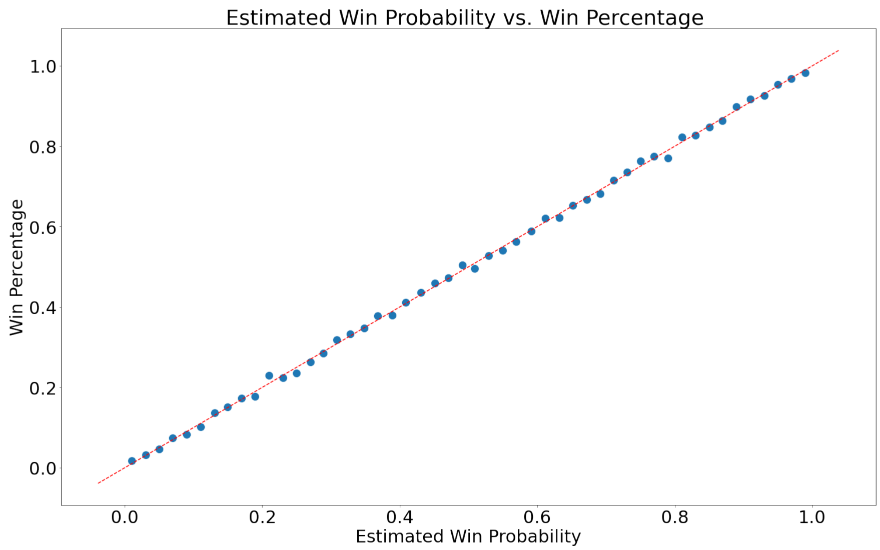
We next validate the accuracy of the estimated expected returns and win probabilities. We form expected returns for the away contracts, and sort contracts into deciles based on the Moneyline (as in Figure 1), and plot the average estimated expected returns and average realized returns of each contract in each decile in Figure A.1. The expected returns match up with the average realized returns of contracts within each decile and capture the patterns observed in the data.

To assess the estimated win probabilities, we categorize each contract into 50 equally-spaced bins based on the estimated win probability (i.e., one bin for contracts with estimated win probabilities of 0 to 2%, one bin for estimated win probabilities of 2 to 4%, etc.), and plot the average win probabilities against the win percentage of each bin. The plots are presented in Figure A.2. We find that the estimated and realized win probabilities match up along a 45 degree line using both methodologies.

Lastly, one point to note is that we also run alternate versions of this procedure, including non-payoff related characteristics in the regressions. The inclusion of these characteristics generally improves the explanatory power for contract returns, but does not do a substantially better job of explaining the Favorite-Longshot bias, the primary pattern of interest in our study. Additionally, the estimated win probabilities when including non-payoff related characteristics introduce noise into the estimation of the model in Section 5, where it is difficult to accommodate preferences for non-monetary characteristics into the preferences of bettors. Accordingly, we proceed using the estimated win probabilities that do not incorporate non-payoff related characteristics.

Figure A.2: Estimated Win Probabilities vs. Actual Win Percentages

The figure plots a scatterplot of estimated win probabilities against win percentages. The methodology to estimate win probabilities is described in Appendix A. For each sport, all contracts are split into 50 equally spaced bins. The average estimated win probability for each bin is plotted on the x-axis and the actual percentage of games won on the y-axis. A 45 degree line is plotted for reference (dotted red line).



B Matching the Favorite-Longshot Bias with Heterogeneous Beliefs

Here we present the setup for a stylized model of risk-neutral bettors that have heterogeneous beliefs about the expected point-difference in each game. Similar to Gandhi and Serrano-Padial (2015), bettors in the model partition themselves into one of the contracts based on their beliefs about the game's outcome. This type of model can theoretically rationalize the Favorite-Longshot Bias, if the marginal bettor in games where the favorite is more heavily favored than the underdog has more extreme beliefs (in favor of the underdog). We present details on estimating the belief distribution from the model using data on the actual proportion of bets on both teams in the game.

B.1 Setup

We denote the team that is favored to win the game as team 1 and the underdog as team 2. The probability distribution function and cumulative distribution function for the difference in points scored by team 1 and team 2 are f and F. We assume f is normal, with a mean that varies across games and standard deviation σ, which is common across all games in a sport. For each game, there are a continuum of risk-neutral bettors, each indexed by their probability distortion, b. Bettor b perceives the probability distribution as $\hat{f}(x) = f(x - b\sigma)$. A value of $b = 0$ indicates that a bettor has the correct beliefs about the point-spread distribution. More positive and negative values of b correspond with bettors believing the mean point-spread will be higher or lower than the truth (e.g., $b = 1$ indicates the bettor believes the expected point-spread will be one standard deviation higher than the true expected point spread). Bettors are dogmatic in their beliefs, in the sense that viewing market prices does not change their beliefs (they agree to disagree). We denote the probability distribution function of belief distortions b as $h(b)$, and assume that $h(b)$ is atomless and continuous, and that beliefs are drawn from the same distribution for each game. The goal of this section is to estimate $h(b)$, and to understand what the implied probability distribution of beliefs looks like to match prices and volume.

For each game, the bookmaker fixes prices for the Moneyline contracts exogenously. Bettors choose between betting on team 1 and team 2 in the Moneyline (and do not have the option not to bet, meaning we only capture the distribution of beliefs conditional on the choice to bet).[39] As before, betting on team 1 offers a potential payoff of y_1 per dollar wagered and betting on team 2

[39] Theoretically, we could expand the choice set of contracts to include Spread contracts as well. However, our empirical exercise relies upon being able to calculate market shares in each betting contract for each game. Unfortunately, we do not have reliable data to calculate the relative share of betting in Spread versus Moneyline contracts. Anecdotally, it also appears that the choice between betting in the Spread and Moneyline contracts might be driven by factors other than beliefs about the game's outcome.

offers a potential payoff of y_2 per dollar wagered. Because bettors are risk-neutral, they purchase the contract that provides them with the highest subjective expected return. When both contracts trade in equilibrium, this means there is a $\bar{b}$ such that $b > \bar{b}$ bets on team 1 and $b < \bar{b}$ bets on team 2 in the Moneyline, since subjective expected returns for betting on team 1 (team 2) are increasing (decreasing) in b.[40]

Bettor $\bar{b}$ is the marginal bettor in each game that prices contracts in equilibrium, in the sense that his belief is reflected in the price of the contracts offered. More negative values of $\bar{b}$ mean that the marginal bettor is more (incorrectly) optimistic about the underdog winning the game, and accordingly corresponds with more negative returns for betting on the underdog. In order to reproduce the Favorite-Longshot Bias in the data, we expect $\bar{b}$ to be more negative for games in which the outcome is more extreme.

Given the assumption that bettors sort into contracts based on their valuations, we can re-write the market shares for each contract (the proportion of betting on each contract) in terms of a cutoff bettor, $\bar{b}$, and the belief distribution $h(b)$.

$$s_1 = \int_{\bar{b}}^{\infty} h(b) \qquad \text{(Favorite Market Share)}$$

$$s_2 = \int_{-\infty}^{\bar{b}} h(b) \qquad \text{(Underdog Market Share)}$$

We measure market shares in two ways. First, we measure the market share of a contract as the proportion of the number of bets placed on the contract in a game (which is directly provided to us in the data). Second, we estimate the market share of a contract as the proportion of dollars bet on the contract assuming the bookmaker takes no risk. To estimate market shares in this way, we re-write the market clearing conditions in Equations (15) and (16) in terms of market shares (below). We then directly solve for market shares for each game by equating the market clearing conditions, where $s_1 + s_2 = 1$. Both measures of market shares have different interpretations for

[40]Partitioning by beliefs also occurs in a context where each bettor is risk-averse and makes a discrete choice to either bet the same amount on team 1 or team 2. This is because the subjective expected utility of a risk-averse bettor wagering a fixed amount on team 1 (team 2) is increasing (decreasing) in b. The discrete choice assumption with equal betting amounts is made elsewhere in related studies (Gandhi and Serrano-Padial (2015) and Chiappori et al. (2019)). Without imposing discrete choice, a risk-averse bettor may choose to bet on multiple contracts, because of a hedging motive. Similarly, while partitioning by beliefs would occur for risk-averse bettors that decide between betting the same amount on either team, it does not necessarily occur when bettors have the choice to bet different amounts on the two contracts.

our results, which we discuss in more depth below.

$$s_1(1+y_1) = s_1 + s_2 \tag{B.1}$$

$$s_2(1+y_2) = s_1 + s_2 \tag{B.2}$$

The likelihood function of bettor b choosing contract j can be expressed in terms of the cutoff agent and the distribution of beliefs

$$\text{likelihood}(j;b) = \begin{cases} \int_{-\infty}^{\bar{b}} h(b) & \text{if } j=1 \\ \int_{\bar{b}}^{\infty} h(b) & \text{if } j=2 \end{cases} \tag{B.3}$$

Using the market shares, we can express the likelihood of a sample of games $k = 1, \ldots, K$ as

$$\mathcal{L} = \prod_{k=1}^{K} \prod_{j_k \in J_k} \text{likelihood}(j_k; b_k)^{s_k^j} \tag{B.4}$$

and the corresponding log-likelihood of the sample as

$$\log \mathcal{L} = \sum_{k=1}^{K} \sum_{j_k \in J_k} s_k^j \log(\text{likelihood}(j_k; b_k)). \tag{B.5}$$

B.2 Empirical Implementation

To estimate belief heterogeneity in the data, we need to express the likelihood function in Equation (B.3) in terms of parameters we can estimate. Because bettors are risk-neutral, $\bar{b}$'s indifference condition is expressed as

$$\bar{p}y_1 - (1-\bar{p}) = (1-\bar{p})y_2 - \bar{p} \tag{B.6}$$

Re-arranging terms yields

$$\bar{p} = \frac{y_2+1}{y_1+y_2+2} \tag{B.7}$$

We then convert the corresponding value of $\bar{p}$ for each contract to $\bar{b}$ by computing

$$\bar{b} = F^{-1}(\bar{p}) - F^{-1}(p) \tag{B.8}$$

where F^{-1} is the inverse CDF of the standard normal distribution, and p is the objective probability that team 1 wins the game. As in Section 4.1 (and discussed in Appendix A), we use the nonparametrically estimated win probabilities for the objective win probabilities.

We assume $h(b)$ is logistically distributed with mean zero. The parametric assumption is used primarily for simplicity, and our interpretation of the results does not vary with this assumption. Given the parametric assumptions imposed, the scale parameter, which controls the dispersion of the belief distribution, is the only parameter to be estimated.

B.3 The Two Market Share Measures

We estimate the model two measures of market shares: the proportion of bets placed on a team, and the proportion of the dollar value of bets placed on a team. In the case that each bettor places wagers of the same size, as is commonly assumed in the literature, both of these measures of market shares are exactly the same. However, given our evidence, it is unlikely that bettors place wagers of the same size on each contract.

The estimated proportion of dollars bet on contracts is strictly increasing in the probability of the contract paying off. Bets on underdogs offer a high payoff with a low probability, while bets on favorites offer a low payoff with a high probability. For the market to clear without the bookmaker taking any risk, more dollars must be bet on the favorite than the underdog in equilibrium, as can be observed via Equations (B.1) and (B.2). However, the relationship between the proportion of bets placed on a contract and the probability of the contract paying off is an empirical question, since in practice, bettors may wager different amounts when betting on the favorite and the underdog.

As it applies to the interpretation of our estimated belief distribution, measuring a contract's market share as the proportion of bets placed on the contract yields an estimate of the distribution of beliefs in the population, giving each individual bettor equal weight. Measuring a contract's market share as the proportion of dollars wagered on the contract yields an estimate of the dollar-weighted distribution of beliefs in the population, where each bettor's belief is weighted by the number of dollars wagered by the bettor.

The empirical relationship between the proportion of bets placed and the dollar proportion of bets placed on a contract, as presented in Figure 3 and discussed in the main text, foreshadow the results of the estimation results. First, the flat relationship between the two suggests that the estimated distribution of beliefs is more dispersed than the dollar-weighted distribution of beliefs. Second, the non-monotonicity of the relationship suggests that belief heterogeneity alone will not be sufficient to explain the distribution of beliefs. Belief heterogeneity can explain the under-performance of extreme underdogs when the marginal bettor betting on extreme underdogs has more extremely distorted beliefs. However, belief heterogeneity cannot explain the *increase* in the proportion of bettors for extreme underdogs that we observe in the data.

B.4 Expected Returns from Estimated Belief Distribution

We construct the implied expected returns for each contract using the estimated belief distribution and observed market shares. Using Equation (B.8), we estimate the objective probability that team 1 wins (and hence expected returns) as $p = F\left(F^{-1}(\bar{p}) - \bar{b}\right)$, where F and F^{-1} are the standard normal CDF and inverse CDF, $\bar{b}$ is the belief of the marginal bettor and $\bar{p}$ is team 1's win probability as perceived by $\bar{b}$. We directly calculate $\bar{p}$ from contract payoffs in each game via Equation (B.7). We calculate $\bar{b}$ using observed market shares for each game and the calibrated belief distribution via the equation $\bar{b} = H^{-1}(s_2)$, where H^{-1} is the inverse CDF function of the calibrated belief distribution.

B.5 Empirical Results

We calibrate the belief distribution using both market share measures via Maximum Likelihood, which provide estimates of the scale parameter of the belief distribution. When using dollar market shares, we estimate the scale parameter as 0.0718 (standard error of 0.0012) - the corresponding estimated standard deviation of b is $\sigma_b = 0.13$. When using market shares, we estimate the scale parameter as 0.302 (standard error of 0.014) - the corresponding estimated standard deviation of b is $\sigma_b = 0.547$.

To assess the ability of the model to explain the Favorite-Longshot bias with both assumptions, we compare the model implied expected returns with the realized returns across contracts. We sort each contract into deciles based on the Moneyline. We plot the model implied expected returns and the average realized returns for each decile using both market share measures in Figure B.1. The first plot in the figure corresponds with the model implied returns assuming bettors each wager the same amount. The plot reveals a Favorite-Longshot Bias in expected returns that is quantitatively similar to that observed in realized returns, though the magnitude of the estimated Favorite-Longshot Bias is slightly larger than observed in the data.

The second plot in the figure corresponds with expected returns implied using the actual proportion of bets. The model-implied returns exhibit a non-monotonic Favorite-Longshot Bias and the model does a considerably worse job of capturing Favorite-Longshot Bias than using the dollar weighted belief distribution. The model underestimates the returns for extreme underdogs and overestimates the returns for more moderate underdogs. Extreme underdogs (in the first decile) actually earn higher equilibrium returns than more moderate underdogs in the model calibration, opposite what we observe in the data. The failure of heterogeneous beliefs to explain the Favorite-Longshot bias using the actual proportion of bets stems from the fact that the proportion of bets placed is not monotonic in odds, with a substantial uptick of bettors choosing to wager on the underdog in games where the outcome is expected to be more extreme.

B.6 Heterogeneous Beliefs and Choice Sets

One potential obstacle for the conclusions we draw is that they assume the distribution of beliefs across games is the same, with bettors choosing either to bet on the underdog or favorite in a particular game. In reality, bettors have a wider choice set available, which can challenge the assumption that the distribution of beliefs of bettors choosing to wager on a game is the same across games. For example, bettors may have subjective beliefs about the outcome of each game that is played on a particular date, and, based on these beliefs, may choose the betting contract *across* games that they find most attractive. In this example, heterogeneous beliefs may be consistent with our finding that, relative to games with more moderate underdogs, a higher proportion of bettors choose to wager on the underdog in games with more extreme underdogs. It is possible that with other alternative games to bet on, bettors that would otherwise bet on a game with an extreme expected outcome choose to bet on games with more moderate expected outcomes, which they subjectively believe offer more attractive returns. With an appropriately shaped distribution that beliefs are drawn from, where many bettors have moderately biased beliefs that lead them to find extreme favorites to be less attractive than other contracts, and where a sufficient number of bettors have beliefs that make them find extreme underdogs to be the most attractive bet, we may be able to find the pattern of increasing proportion of bettors choosing the underdog in more extreme games.[41]

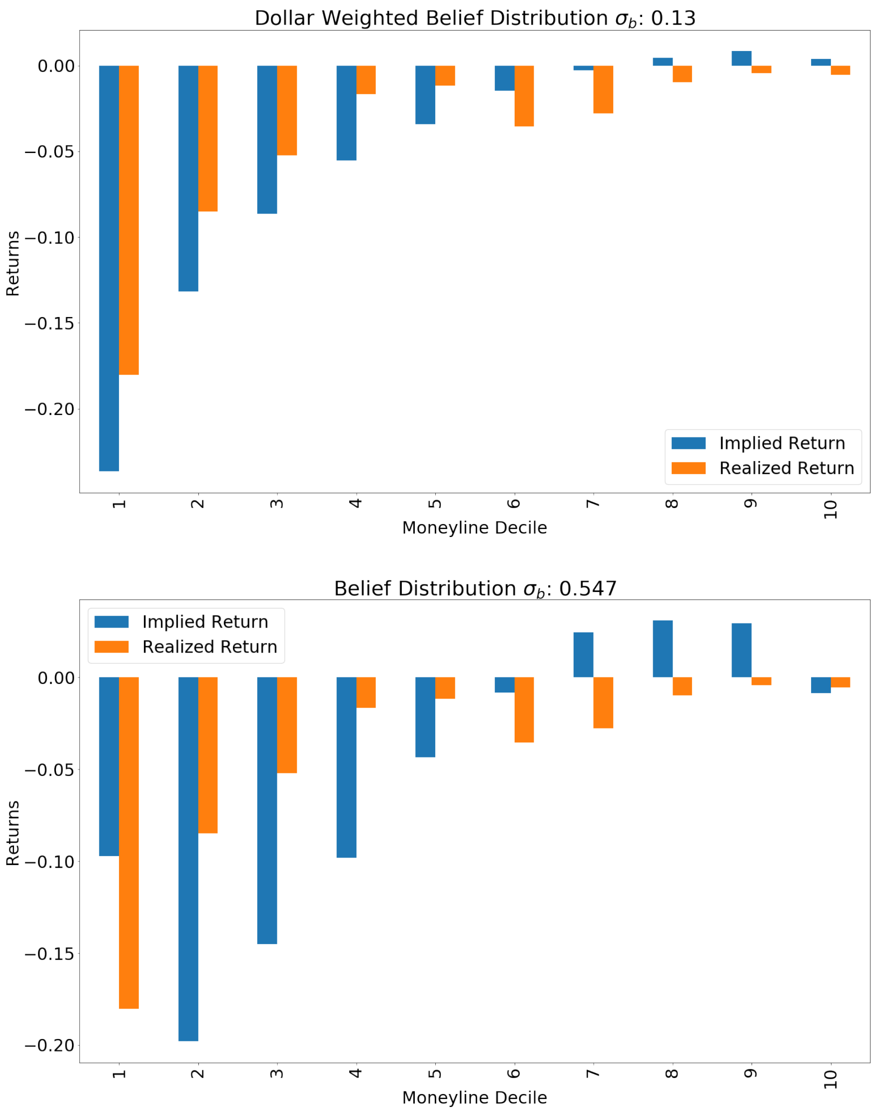
We do not have the betting volume data required to estimate a model in which the choice set includes all potential games on a date. However, an implication of this type of heterogeneous beliefs model is that the allocation of bettors into different contracts will vary depending upon the choice set available on each date. For example, when there is only one game available to bet on, all bettors must choose to bet either on the favorite or the underdog. When there are two games available to bet on, bettors with more extreme beliefs may choose to bet on the more extreme underdog (which earns the lowest returns), while bettors with more moderate beliefs may choose to bet on the other three available contracts. Under this alternative explanation for our empirical evidence, we expect the proportion of bets placed on extreme underdogs and the returns of these underdogs to vary based on the other betting opportunities available on a given date.

To test this alternative explanation, we consider the decile of the most extreme underdogs, and analyze (1) how the proportion of bettors choosing to bet on the underdog and (2) how betting returns, vary with the available choice set on a given date. We construct two measures to capture the relevant choice set for a particular contract. The first is simply the number of other betting contracts available on the same date as the contract considered. The second is the proportion of

[41] This alternative may be closer to the studies of belief heterogeneity at the horse racetrack, where bettors choose from a number of horses in a race.

Figure B.1: Heterogeneous Beliefs and Expected Returns

The figure plots the average returns and expected returns of Moneyline contracts for games that we have betting volume data for, sorted into deciles based on the Moneyline. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. Expected returns are computed using the observed market shares on each contract and the calibrated belief distribution. In the first plot, expected returns are estimated using the dollar-weighted belief distribution, estimated by assuming that bettors each wager the same amount. In the second plot, the expected returns are estimated using the belief distribution, weighing each bettor equally in the calculation for the belief distribution. Each plot also includes σ_b, the estimated standard deviation of the belief distribution implied by the model.



bets on underdogs available on the same date with a higher probability of paying off (with zeros if no other games are available to bet on). We regress contract returns and the proportion of bets placed on the underdog on these variables, including a control for the contract odds. In effect, these regressions capture whether, for a given odds contract, the betting returns and proportion of bets placed on the contract vary with the set of alternative contracts that are available. We standardize all independent variables in the regressions to have zero mean and unit standard deviation. If sorting across games explains our results, we expect lower returns and a higher proportion of bettors betting on the underdog corresponding with increases in the independent variables. This explanation for our results would suggest that with a greater amount of alternative betting opportunities in games with more moderate expected outcomes, a smaller (larger) proportion of bettors choose to bet on the favorite (underdog) in games with extreme expected outcomes, and the bettors choosing to bet on extreme underdogs will have more extreme beliefs.

Table B.1 reports the results from the regressions. The first three columns correspond with results where betting returns are the dependent variable, and the last three columns correspond with results where the proportion of bets placed are the dependent variable. The regression results suggest that returns are decreasing in odds, and the proportion of bets placed are increasing in odds. Since higher odds correspond with more extreme underdogs, these results are consistent with the Favorite-Longshot bias, and the increasing proportion of bettors betting on more extreme underdogs. Next, we turn to the variables of interest in the regression. First, the coefficients on the number of other betting contracts available is not statistically significant in any of the regressions, and in fact has the opposite sign as expected when the betting proportion on the underdog is the dependent variable. Turning to the proportion of other underdog contracts with a greater probability of paying off, the regressions suggest that a one standard deviation change in this variable corresponds with a 1% increase in the proportion of bettors wagering on the underdog. This is small compared with the average proportion of bettors (41%) choosing to bet on the underdog in this decile of contracts. Additionally, the coefficient on this variable is negligible in regressions where the independent variable is contract returns, suggesting that choice sets do not appear to explain any variation in the returns of extreme underdogs. Overall, the regression results do not support an alternative explanation for our results where bettors sort into different games based on their beliefs.

An additional point of note in this discussion is that bettors are also free to bet on multiple events, and are not restricted to betting on a single event. In a sample of 336 individual bettors, Andrikogiannopoulou and Papakonstantinou (2016) report that when betting, the average bettor in their sample places a little more than five bets per day. This provides some evidence against the idea that bettors select a single game where they find the expected returns to be most attractive, as sorting into games by beliefs might suggest.

Table B.1: Betting on Underdogs and Choice Sets

The first three columns in the table present regression results from a series of regressions of contract betting returns on the number of other contracts offered on the same date ("Num Contracts") and the proportion of contracts on underdogs with greater probability of paying off offered on the same date ("Proportion"), where the sample is the decile with the most extreme underdogs. The last three columns in the table report regression results with the same sample and independent variables, where the dependent variable is the proportion of bets placed on the underdog. *t*-statistics are reported in parentheses.

	Betting Returns			Proportion of Bets		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-0.23	-0.23	-0.23	0.41	0.41	0.41
	(-7.33)	(-7.33)	(-7.33)	(126.70)	(125.89)	(126.10)
Proportion	0.00		0.00	0.01		0.01
	(-0.01)		(0.03)	(3.91)		(3.93)
Num Contracts		-0.05	-0.05		-0.01	-0.01
		(-1.69)	(-1.69)		(-1.92)	(-1.95)
Odds	-0.11	-0.11	-0.11	0.03	0.03	0.03
	(-3.30)	(-3.52)	(-3.35)	(6.20)	(7.81)	(6.13)
N	7349	7,349	7,349	5,451	5,451	5,451

B.7 Discussion

While we believe preferences play an important role in the results, the results do suggest a potential role for belief heterogeneity in the Favorite-Longshot Bias as well. In our stylized model of heterogeneous beliefs, we estimate that the standard deviation of the dollar-weighted belief distribution and the unweighted belief-distribution, expressed in units of the standard deviation of point-spread distribution, are 13% and 54.7%. Compared with numbers found in financial markets, these numbers imply a reasonable degree of belief heterogeneity. For example, Giglio et al. (2021) report in a survey of investors that the standard deviation of expectations of the one-year return of the S&P 500 is 5.2%. The annualized volatility of monthly S&P 500 returns is about 15%, so belief heterogeneity about stock market returns, as a proportion of stock market volatility, is about 35%. Similarly, using the I/B/E/S analyst database, which provides analyst estimates of quarterly stock level earnings per share for a number of firms, (analyzing all listed US stocks in the database from 1990 to 2017), we find that the average standard deviation of quarterly earnings forecasts, as a percentage of the trailing 2-year standard deviation of a firm's earnings, is 35%. With these numbers from financial markets providing context, the degree of belief heterogeneity revealed by our model is quite reasonable. While belief heterogeneity is unlikely to be the whole story, it may play a role in explaining the patterns in the data in combination with preference-based explanations.

C Preference Calibration

C.1 What Parameters Can Fit the Data?

To get a sense of what parameter values may be able to fit the data, we form a grid of values for $(\alpha, \gamma) \in [0.5, 1] \times [0.5, 1]$ (which broadly correspond with the range of values for α and γ found in experimental settings), with values on the grid spaced in intervals of 0.01. For each (α, γ) pair on the grid and $\lambda \in \{1, 1.25\}$, we compute the proportion of contracts for which the model is able to provide implied payoffs that satisfies the equilibrium and optimality conditions for the given choice of parameters. We also compute the mean squared error of expected returns implied by the model for each set of parameters.[42] We plot the surface of valid contract proportions and the negated MSE values in Figure C.1.

[42]The optimality conditions imply that contract expected returns must be negative, and that prospect theory utility must be positive. This provides us with necessary and sufficient conditions to evaluate if a choice of parameters Θ generates a valid payoff that satisfies the equilibrium and optimality conditions given the loss probability for a game. Particularly, for game i and a given Θ, a valid payoff exists if and only if the largest potential payoff that has negative returns (calculated using the loss probability) has a positive prospect theory value and provides greater utility than the spread contract.

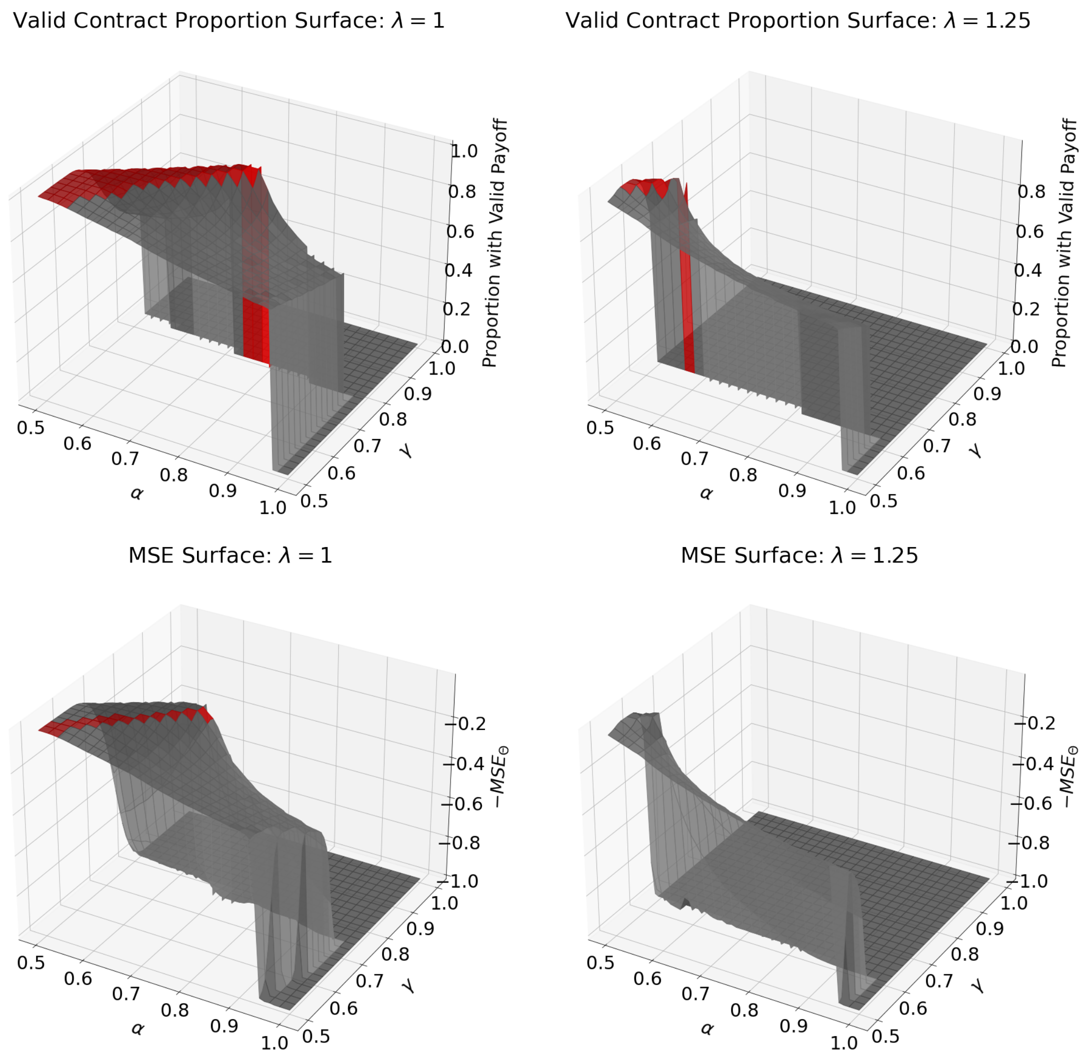
The top left panel in the figure plots the proportion of contracts for which the model is able to calculate valid payoffs for a given (α, γ) pair, for $\lambda = 1$ (which corresponds to no loss-aversion). Points in red on the surface indicate (α, γ) pairs for which the model is able to calculate valid payoffs for at least 99% of contracts. The plot indicates that the model is able to generate valid payoffs for at least 99% of contracts for a band of (α, γ) pairs that lie along the line $\alpha = \gamma$ and are less than 0.8. The top right panel in the figure plots the same quantities, for $\lambda = 1.25$ (which corresponds with moderate loss-aversion). For each (α, γ) pair, the corresponding proportion of contracts for which the model computes valid implied payoffs is lower than the $\lambda = 1$ case, and the model is only able to generate valid payoffs for more than 99% of contracts for a considerably smaller set of values.

The bottom left panel in the figure plots the negated MSE values against values of α and γ, for $\lambda = 1$. Points in red on the surface have an MSE < 0.014, while all other points are colored in gray. The observations with MSE lower than 0.014 all lie close to the diagonal line $\alpha = \gamma$, with values of α and γ less than 0.7. The bottom right panel plots the MSE surface for $\lambda = 1.25$, and once again reveals that the model fit is considerably worse when introducing loss-aversion. For each point on the grid, the MSE is higher than it is for the corresponding point with $\lambda = 1$.

Figure C.1 indicates that *both* diminishing sensitivity, captured by α, and probability weighting, captured by γ, are important for explaining the data. Providing a strong fit for the data requires parameter values for both that are less than 0.8. Additionally, the parameter value for each feature is closely related to the parameter value of the other feature, as indicated by the fact that the model performs best for values of α and γ that are close to each other. Lastly, the figure also demonstrates that loss-aversion ($\lambda > 1$) reduces the ability of the model to explain the data. We provide more economic intuition for these results after first addressing whether the model can quantitatively fit the data.

Figure C.1: Surface of Mean Squared Errors and Valid Contract Proportions

The top two panels in the figure plot the proportion of contracts against values of α, the diminishing sensitivity parameter, and γ, the probability weighting parameter, for which the model computes a valid payoff that satisfies the optimality conditions. The top left panel corresponds with values where the loss aversion parameter, $\lambda = 1$. The top right panel corresponds with values for $\lambda = 1.25$. The bottom two panels in the figure plot the mean squared error of expected returns against α and γ. The bottom left panel corresponds with values for $\lambda = 1$ and the bottom right panel corresponds with values for $\lambda = 1.25$.



C.2 Probability Weighting and Diminishing Sensitivity

Having calibrated the model, we seek to better understand what role the parameters in the model play in explaining the results. The two main parameters we focus on are the diminishing sensitivity parameter, α, and the probability weighting parameter, γ.

The probability weighting function increases the weight that the bettor gives to *unlikely* events in her decision, and decreases the weight that the bettor gives to *likely* events. The parameter γ governs the amount of probability weighting, where lower values generally correspond with increasing the weight that tail events play in a person's decision. Relative to the benchmark of objective probabilities, this generates a preference for lotteries with skewed payoffs, which in turn can help explain the lower equilibrium return the bettor demands for underdogs that offer high payoffs with low probability. As Barberis and Huang (2008) note, probability weighting can generate a positive demand for assets with negative returns that have positively skewed payoffs (like the underdogs in our sample), so probability weighting also helps explain why bettors may enter into the betting market on the extensive margin, and wager on underdogs in the first place.

Diminishing sensitivity, governed by the parameter α, leads the bettor to worry less about marginal losses the further she is from her reference point (here, her initial wealth), with smaller values of α corresponding with increased diminishing sensitivity. In the context of bets on the favorite, increased diminishing sensitivity leads the small, but high probability gains from betting to look more attractive relative to the possibility of losing the wager, to the point where the bettor gets positive utility from the bet. In the absence of diminishing sensitivity, bettors would not derive positive utility from betting on favorites, which offers negative returns that are only made less attractive when including probability weighting.

In our specification, there is a tension between probability weighting, which makes underdogs more attractive and favorites less attractive, and diminishing sensitivity, which makes favorites more attractive and underdogs less attractive. Overall, the two features are linked tightly, and increasing the amount of probability weighting also increases the amount of diminishing sensitivity required to explain the data. The relationship between probability weighting and diminishing sensitivity also helps to better understand some of the results in Table 5. For example, increasing probability weighting can potentially provide a better fit for the returns of extreme underdogs in decile 1, but doing so comes at the cost of being able to compute payoffs that satisfy the optimality conditions for the extreme favorites in decile 10. This relationship between probability weighting and diminishing sensitivity also explains why the fit of the model does not change substantially for difference choices of γ, when α is allowed to vary.

C.3 Probability Weighting, Loss-Aversion, and the Choice to Bet

In the non-Expected Utility component of preferences, we assume rank-dependent probability weighting (Quiggin (1982)). Losses relative to the reference point are weighted by $w(p)$, and gains relative to the reference point are weighted by $1 - w(p)$, where $w(\cdot)$ is the probability weighting function proposed by Tversky and Kahneman (1992). There are two notable points to make about this choice. First, this form of probability weighting has some differences from implementations of Cumulative Prospect Theory often used to explain financial markets facts, particularly in terms of the way that losses are treated. Second, recent experimental evidence casts doubt upon rank-dependence (Bernheim and Sprenger (2020)). Accordingly, we discuss how changing the assumed form of probability weighting may influence our conclusions.

First, note that the Tversky and Kahneman (1992) probability weighting function is not symmetric around 0.5, and crosses the line $w(p) = p$ at a point $\bar{p} < 0.5$. For $p < \bar{p}$, $w(p) > p$, and for $p > \bar{p}$, $w(p) > p$. This property also holds for other commonly used probability weighting functions (e.g., the Prelec (1998) function). For our implementation, the implication of this property is that bettors underweight loss probabilities greater than $\bar{p}$ and overweight loss probabilities less than $\bar{p}$. Moreover, increasing the degree of probability weighting (lower γ) also corresponds with a lower $\bar{p}$, as illustrated in Figure 8. In effect, increasing probability weighting also carries an additional implication similar to decreasing loss-aversion, λ.[43] This, in turn, contributes to the model's ability to explain the decision to bet, including on Spread contracts, as we decrease γ.

These properties of our implementation can be contrasted with an alternative implementation of probability weighting commonly assumed in financial market applications of Cumulative Prospect Theory, where losses are weighted by $w(p)$ and gains are weighted by $w(1 - p)$. Consider a bettor who evaluates a bet by computing

$$V = (1 - \eta) [p(-b) + (1 - p)yb] + \eta [w(p)v(-b) + w(1 - p)v(yb)] . \tag{C.1}$$

The bettor is identical to the previous bettor in how he evaluates potential bets, except that he probability weights gains using $w(1 - p)$, rather than $1 - w(p)$. For $\lambda = 1$ (no loss-aversion), the previous bettor is willing to bet on a negative expected return Spread contract that has a 50% probability of paying off, while the present bettor is not. Additionally, for $\alpha = \gamma$ (close to what we find in our calibration), this bettor only receives positive non-Expected Utility value from betting

[43] The relationship between probability weighting and loss aversion in our implementation is conceptually related to the non-identification result in Barseghyan et al. (2013). They use the form of reference-dependent preferences proposed in Kőszegi and Rabin (2007), and show in that specification, loss-aversion and rank-dependent probability weighting are equivalent to a probability distortion of the form $\Omega(p) = w(p) [1 + \lambda(1 - w(p))]$.

if bets have positive expected returns, or if $\lambda < 1$.[44]

Would this alternative bettor choose to bet, if $\lambda \geq 1$? The answer is a qualified yes, on some contracts. This bettor will only be willing to wager on favorites, if diminishing sensitivity is stronger than probability weighting ($\alpha < \gamma$), or underdogs, if probability weighting is stronger than diminishing sensitivity ($\gamma < \alpha$), but not both. He is unwilling to bet on Spread contracts. Table C.1 reports the proportion of favorite and underdogs contracts the bettor is willing to wager on for different parameterizations, given the estimated win probabilities and contract payoffs.[45] For $\alpha << \gamma$, the bettor is generally willing to bet on the favorite; for example, for $(\alpha, \gamma, \lambda) = (0.5, 0.65, 1)$ the bettor is willing to bet on 88% of favorites (and no underdogs). For $\gamma << \alpha$, the bettor is increasingly willing to bet on contracts on the underdog; for example, for $(\alpha, \gamma, \lambda) = (0.9, 0.65, 1)$, the bettor is willing to wager on 87% of underdogs (and no favorites). In a market populated with bettors with preferences of this form, we may be able to partially explain the data if bettors have heterogeneous preferences, as studied by Chiappori et al. (2019), with bettors exhibiting more diminishing sensitivity wagering on favorites, and bettors exhibiting more probability weighting wagering on underdogs.[46] However, it is worth noting that introducing even mild loss aversion substantially reduces the willingness to bet – for $\lambda = 1.25$, the bettor is never willing to wager on more than 50% of contracts at the offered prices.

An alternative way to proceed is to allow λ to be less than one, i.e., the bettor may be *loss-tolerant* rather than *loss-averse*. This may be motivated by evidence that part of the overall US population appears to be loss-tolerant (Chapman et al. (2018), who suggest this may be true of more than half the population), or, alternatively, as a way to capture gambling preferences (for example, as in Conlisk (1993)), which may be domain specific. Loss-tolerance can help rationalize the decision of bettors to wager on negative expected return Spread contracts. To have similar explanatory power as the main specification we present, a substantial degree of loss-tolerance is required; the parameter values $(\alpha, \gamma, \lambda) = (0.65, 0.63, 0.86)$ deliver an MSE of returns of 0.012, similar to the model performance in the main specification.

[44] For $\alpha = \gamma$, the non-EU component of value is positive if and only if $(1-p)^\gamma (yb)^\gamma - \lambda p^\gamma b^\gamma > 0$. In turn, for $\lambda \geq 1$, restricting betting amounts to be positive ($b > 0$), this condition can only be satisfied if $(1-p)y - p > 0$, i.e., the contract has positive expected returns.

[45] In these calculations, we use the given contract payoffs for most contracts. For some contracts, which are estimated to have positive expected returns, we substitute the largest payoff for which the contract earns negative returns. We use the same estimated win probabilities as before in these calculations.

[46] It may not be straightforward to apply the approach of Chiappori et al. (2019) to endogenize the amount bettors choose to wager, as we seek to do here. They note "Whether nonparametric tests and identification like those we develop below could be constructed with endogenous bet amounts or would require information at the individual level is an open question."

Table C.1: Alternate Preference Specification: Contracts with Positive Valuations

The table presents the proportion of contracts that have positive valuations under the alternative preference specification given in Equation (C.1), which differs from the main preference specification in the form that probability weighting takes. Panel A fixes the value of α as 0.65, and presents the proportion of contracts with positive valuations for a set of (α, γ, λ) triples. Panel B fixes the value of γ as 0.65 and presents the proportion of contracts with positive valuations for a set of (α, γ, λ) triples.

Panel A: Proportion with Positive Value, $\alpha = 0.65$

γ	$\lambda = 1$		$\lambda = 1.25$	
	Favorites	Underdogs	Favorites	Underdogs
0.50	0%	83%	0%	6%
0.55	0%	66%	0%	1%
0.60	0%	30%	0%	0%
0.65	0%	0%	0%	0%
0.70	66%	0%	1%	0%
0.75	79%	0%	8%	0%
0.80	85%	0%	19%	0%
0.85	89%	0%	33%	0%
0.90	90%	0%	43%	0%
0.95	92%	0%	50%	0%

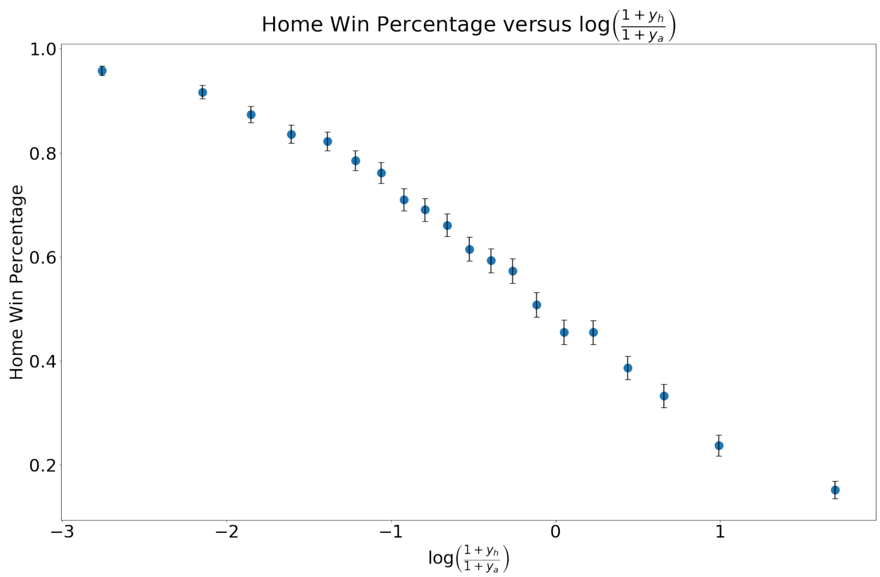
Panel B: Proportion with Positive Value, $\gamma = 0.65$

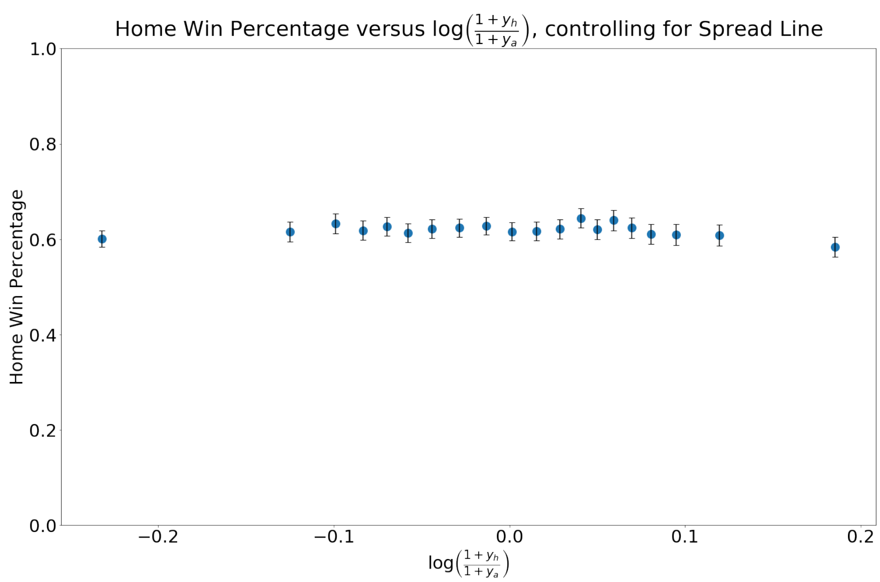
α	$\lambda = 1$		$\lambda = 1.25$	
	Favorites	Underdogs	Favorites	Underdogs
0.50	88%	0%	21%	0%
0.55	82%	0%	8%	0%
0.60	68%	0%	1%	0%
0.65	0%	0%	0%	0%
0.70	0%	28%	0%	0%
0.75	0%	58%	0%	1%
0.80	0%	76%	0%	4%
0.85	0%	84%	0%	14%
0.90	0%	87%	0%	26%
0.95	0%	89%	0%	39%

D Additional Tables, Figures, and Analyses

Figure D.1: Wins, Moneylines, and Spreads

The figure plots the relationship between the probability the home team wins and the quantity $\log\left(\frac{1+y_h}{1+y_a}\right)$ (the "Moneyline ratio"), where y_h and y_a are the payoffs associated with winning bets on the favorite and the underdog in the Moneyline. The first plot in the figure plots a binned scatterplot, where each game is sorted into twenty bins by sport based on the Moneyline ratio, and the figure plots the average Moneyline ratio versus the home win percentage in each bin. The second plot in the figure plots the same quantities, including a control for the Spread Line of the game. The error bars correspond with plus/minus two standard errors relative to the home win percentage in each bin.





D.1 Distribution of Point Differential Minus Spread Line

In this section, we discuss the assumption that the the point-differential minus spread line is distributed *iid* normal across different games within a sport.

First, as we document in our main results, spread contracts are priced such that there is approximately an equal probability of winning and losing the bet, with approximately equal returns regardless of the spread line (as captured by Figure **??**). This suggests that the quoted spread line reasonably accurately captures the median of the distribution. We similarly analyze how the second moment of the realized-point differential minus spread line varies across games. We divide games into deciles based on the quoted Spread Line in each sport, and plot the standard deviation of each decile in Figure D.2. The standard deviation does not exhibit any systematic pattern across games and appears to be relatively constant within each sport. Lastly, one implication of the assumption of the the distribution of the realized point differential minus the quoted spread line being identically distributed is that the quoted spread line is a sufficient statistic to summarize the probabilities of either team winning the game. This, in turn, means that, the spread line should be able to explain the prices of Moneyline contracts. We find that spread lines are able to explain more than 89% of the variation in Moneyline contract prices in each sport, consistent with this implication. Regression results are presented in Table D.1.

We plot a histogram of the realized point differential minus the quoted spread line for each sport in Figure D.3. The point-spread distributions are reasonably well behaved and appear to be approximately symmetric and bell-shaped. For each sport, the figure also overlays the Maximum Likelihood Estimate of the probability distribution function assuming that the data follow a normal distribution. The normal distribution provides a reasonably good, though certainly far from perfect fit of the data.

We can also estimate the implied win percentages for games using the Spread Line and the assumption that the realized point differential minus Spread line distribution is mean zero and identically normally distributed across games in each sport. To evaluate the identical normality assumption, in Figure D.4, we estimate the implied win percentage based on the normality assumption, then group games into 50 equally spaced bins based on their estimated implied win percentage, and plot the average win percentage against the average implied win percentage of each bin. In general, the estimated win percentages line up with the implied win percentages on a 45 degree line, especially for college sports. The estimates are more noisy for the NFL, and for the NBA, the implied win probabilities do appear to be higher than the true win probabilities for strong underdogs and lower than the true win probabilities for strong favorites.

Figure D.2: Point-Spread Standard Deviation By Spread Decile

The figure plots the standard deviation of the difference between the point-spread and spread-line for games grouped into deciles based on the quoted Spread Line on the game. Decile 1 correspond with the home team being strongly favored and Decile 10 corresponds with the home team being strong underdogs. Each panel in the figure corresponds with the values for a particular sport. The error bars corresponds with 95% confidence intervals for the the decile standard deviation.

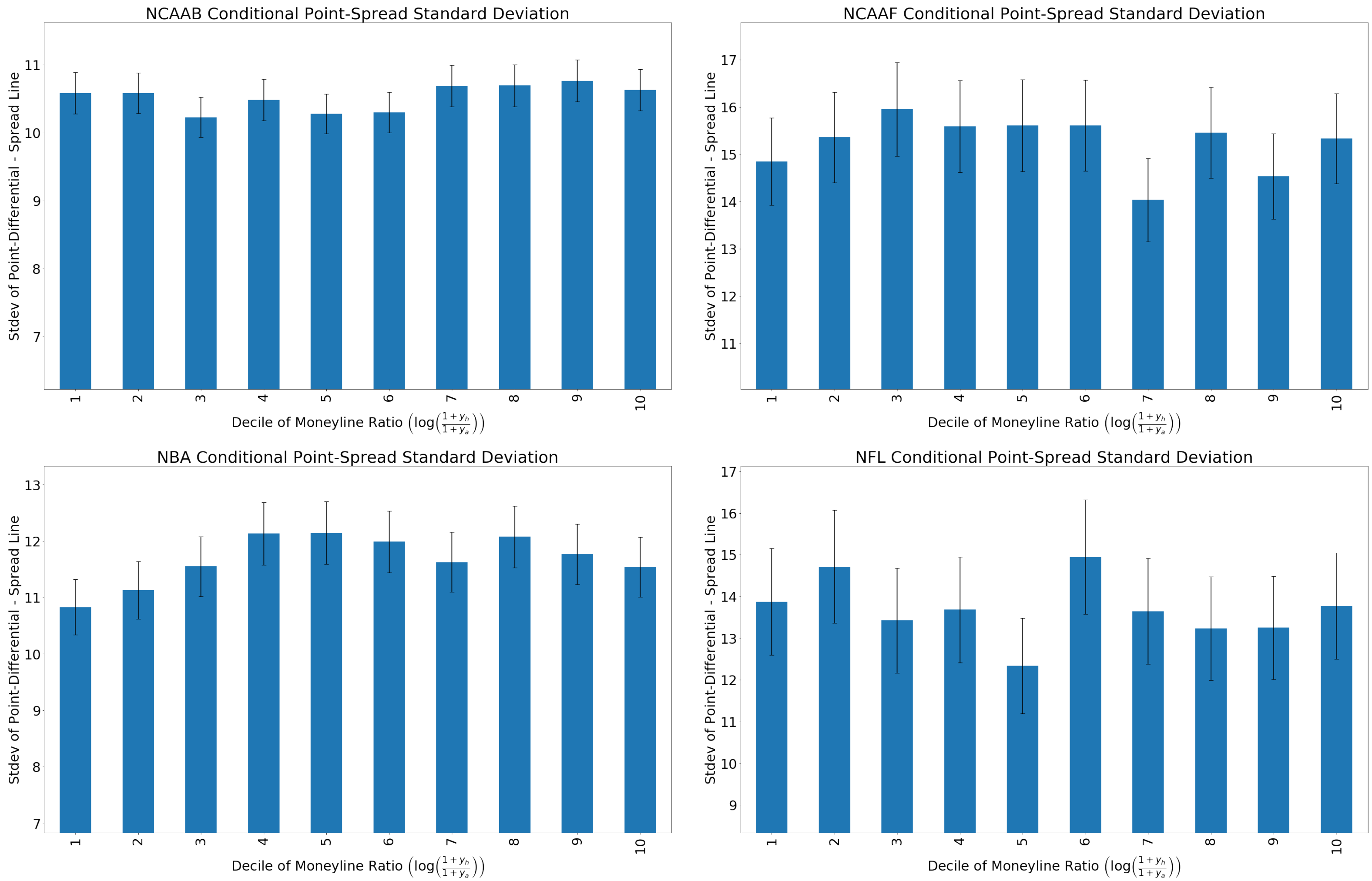


Table D.1: Relationship between Moneyline and Spread Lines

The table returns results from the regression of transformed Moneylines on Spread Lines within each sport. The transformed Moneyline variable is the log of the payoff for a winning $1 bet implied by the Moneyline. The sample includes both the Moneyline contract on the home team and the Moneyline contract on the away team for each game. For observations that correspond with the Moneyline contract on the away team, we multiply the spread line by negative one. Standard Errors are reported in parentheses.

	NCAAB	NCAAF	NBA	NFL	Combined
Intercept	0.845	0.885	0.801	0.767	0.836
	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)
Spread Line	0.076	0.056	0.078	0.071	0.071
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	89.2%	88.0%	89.3%	93.0%	87.5%
N	43,964	8,784	16,784	3,686	73,218

Figure D.3: Conditional Point-Spread Distribution by Sport

The figure plots a histogram of the difference between the point-spread and the posted spread line across games, with one panel for each sport. Each panel also overlays the best fit normal distribution for the data, calculated via Maximum Likelihood.

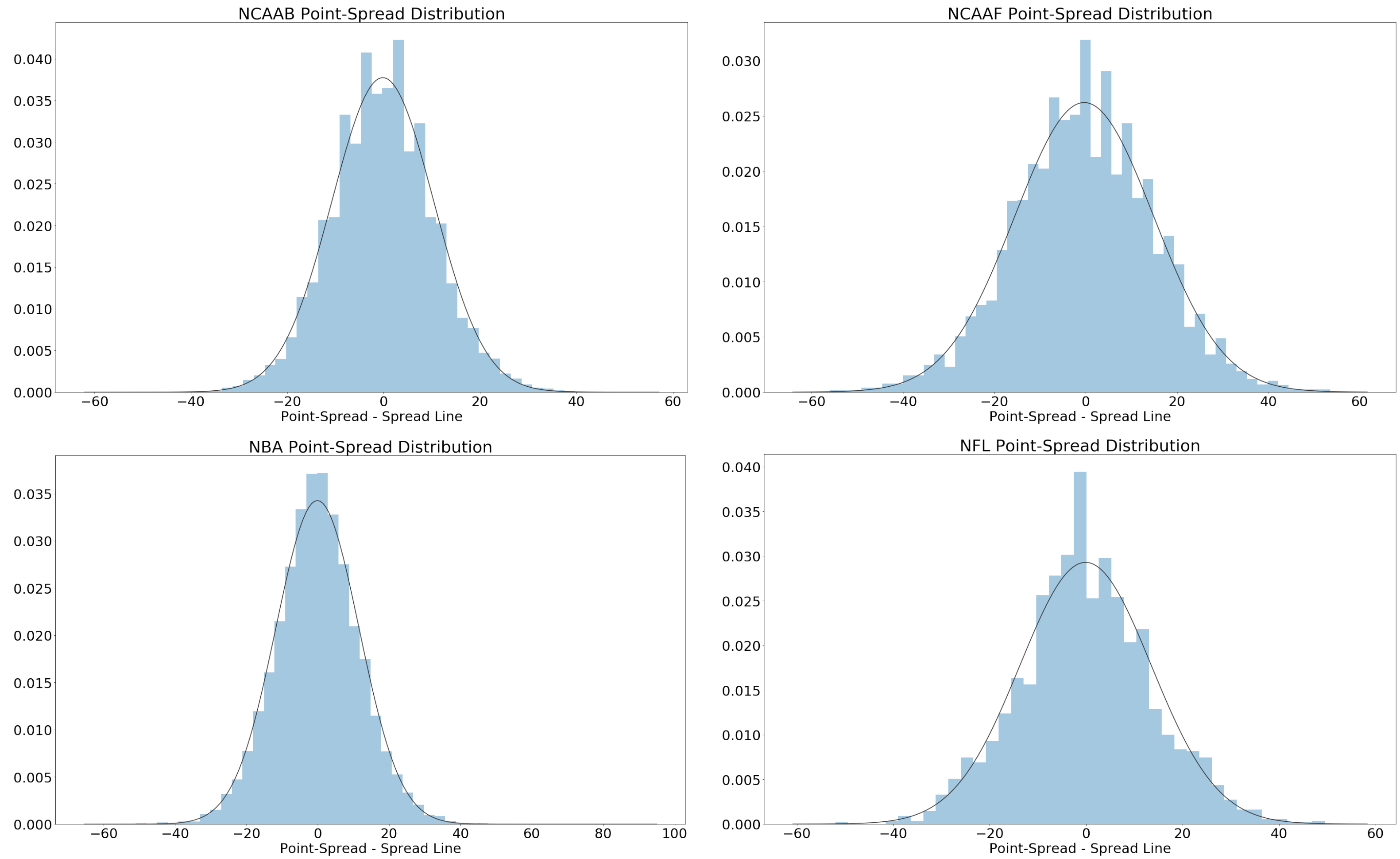


Figure D.4: Win Probabilities Implied by Normality

The figure plots the implied win probability of games assuming that the Spread line minus point-differential distribution is identically normally distributed across games versus actual win percentages. The plot is formed by estimating the standard deviation of the Spread Line minus point-differential distribution for each sport, σ_s. Then, for each game an implied win percentage is calculated, based on the Spread Line and the assumption that the Spread Line minus point-differential is distributed $N(0, \sigma_s)$. All contracts are split into 50 equally spaced bins. Each plot plots the average estimated win probability for each bin on the x-axis and the percentage of games one on the y-axis. Each plot also plots the 45 degree line as a dotted red line.

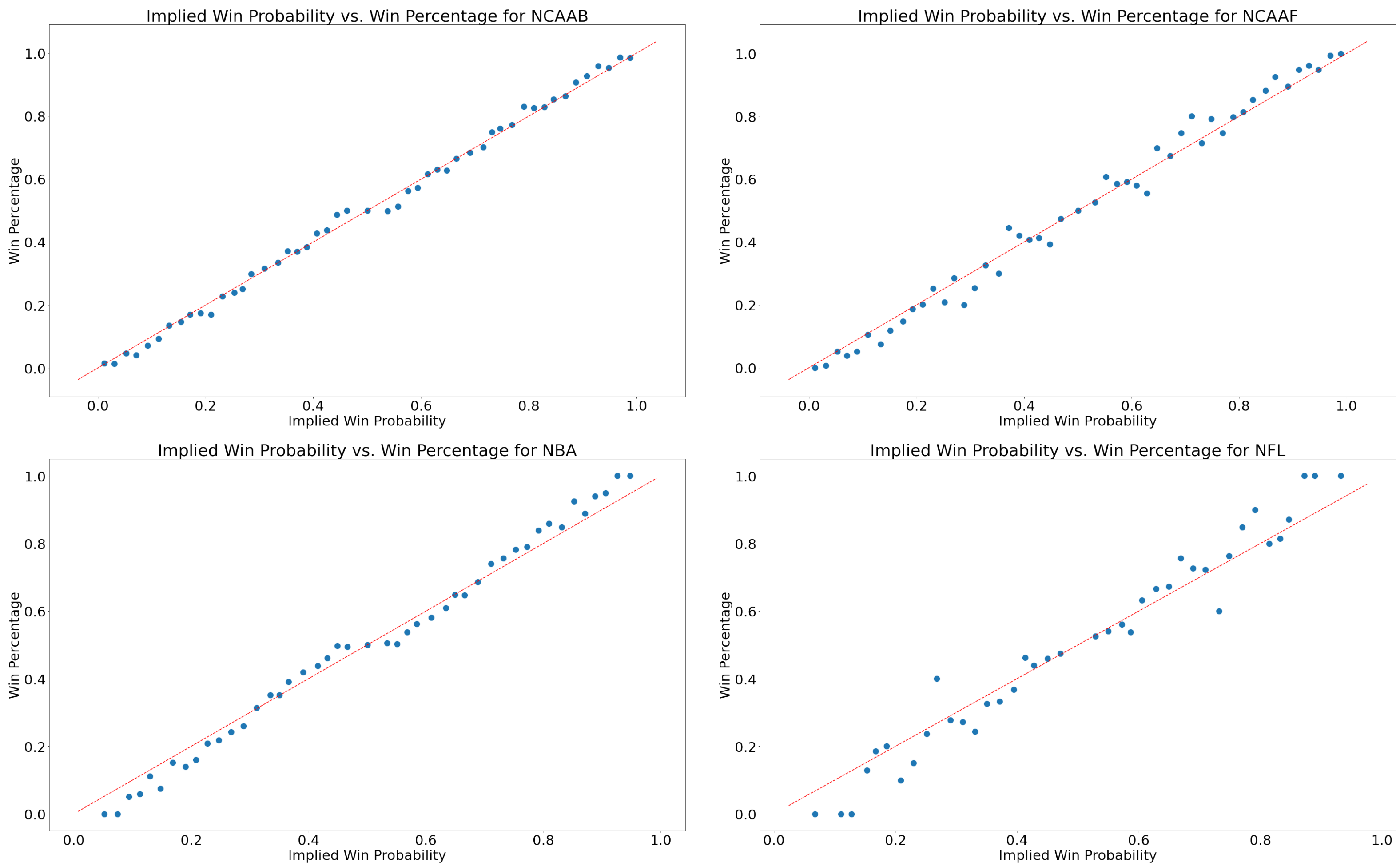


Figure D.5: Accuracy of Over/Under Contracts by Sport

The figure presents binned scatter plot of the Over/Under line versus the total number of point scored in a game for each sport. Each game in the sample is sorted into 20 equal sized bins based on the Over/Under line of the game. Each point on the plot corresponds with the average Over/Under Line and the average point total of each game in one of the bins. The 45 degree line is also plotted on the graph in red. The error bars correspond with plus/minus two standard errors relative to the mean number of points scored in a bin.

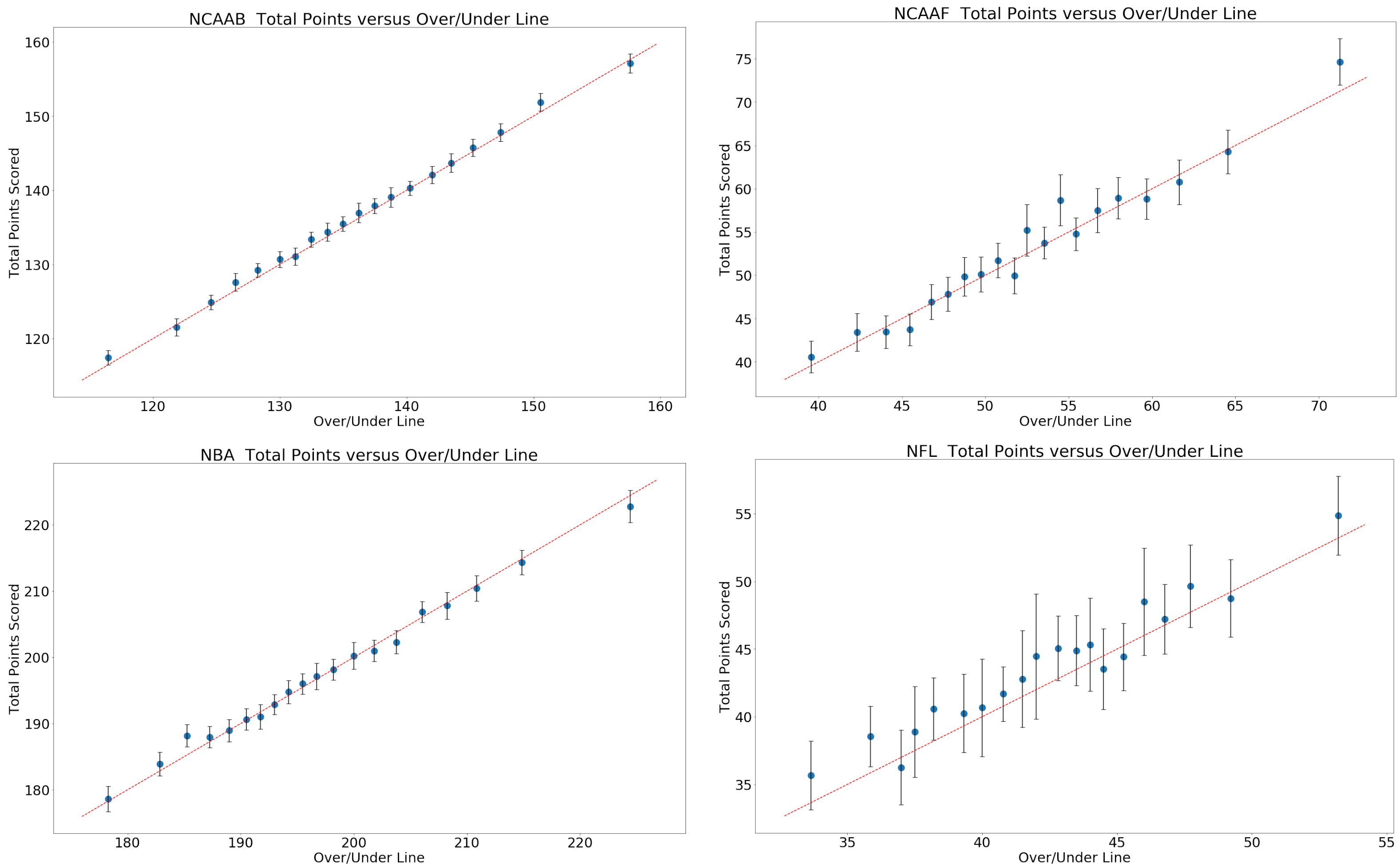


Table D.2: Index Option Sample

The table reports the list of instruments included in our sample of index options along with the corresponding date ranges and ticker symbols.

Name	Ticker	Start Year	End Year
CBOE INT RATE 30 YR T B	TYX	1996	2010
CBOE TREASURY YIELD OPT	TNX	1996	2011
DOW JONES INDEX	DJX	1997	2019
NASDAQ 100 INDEX	NDX	1996	2019
NYSE ARCA MAJOR MARKET	XMI	1996	2008
PSE WILSHIRE SMALLCAP I	WSX	1996	2000
RUSSELL 2000	RUT	1996	2019
S&P100 INDEX	OEX	1996	2019
S&P500 INDEX	SPX	1996	2019
S&P MIDCAP 400 INDEX	MID	1996	2019
S&P SMALLCAP 600 INDEX	SML	1996	2019

D.2 Details on Options Data

For our analysis on the options implied volatility surface, we use all available equity options in the Option Metrics database, and we use data on index options for eleven indices, whose details are listed in Table D.2.

Listed index options are European options, while listed equity options are American. Implied volatilities for European options are calculated using the Black-Scholes formula. Implied volatilities for American options are calculated by OptionMetrics using the Cox, Ross and Rubinstein (1979) binomial tree model.

The volatility surface file is provided by OptionMetrics and constructed for each security on each day by using a kernel smoothing algorithm that interpolates implied volatilities from listed options to provide implied volatilities for call options and put options with fixed expirations and option deltas.

Figure D.6: Cross-Section of Betting Returns with Opening Lines

The figure reproduces first Panel of Figure 1, the first Panel of Figure ??, and Figure ??, using opening lines to sort contracts and calculate returns rather than the closing lines.

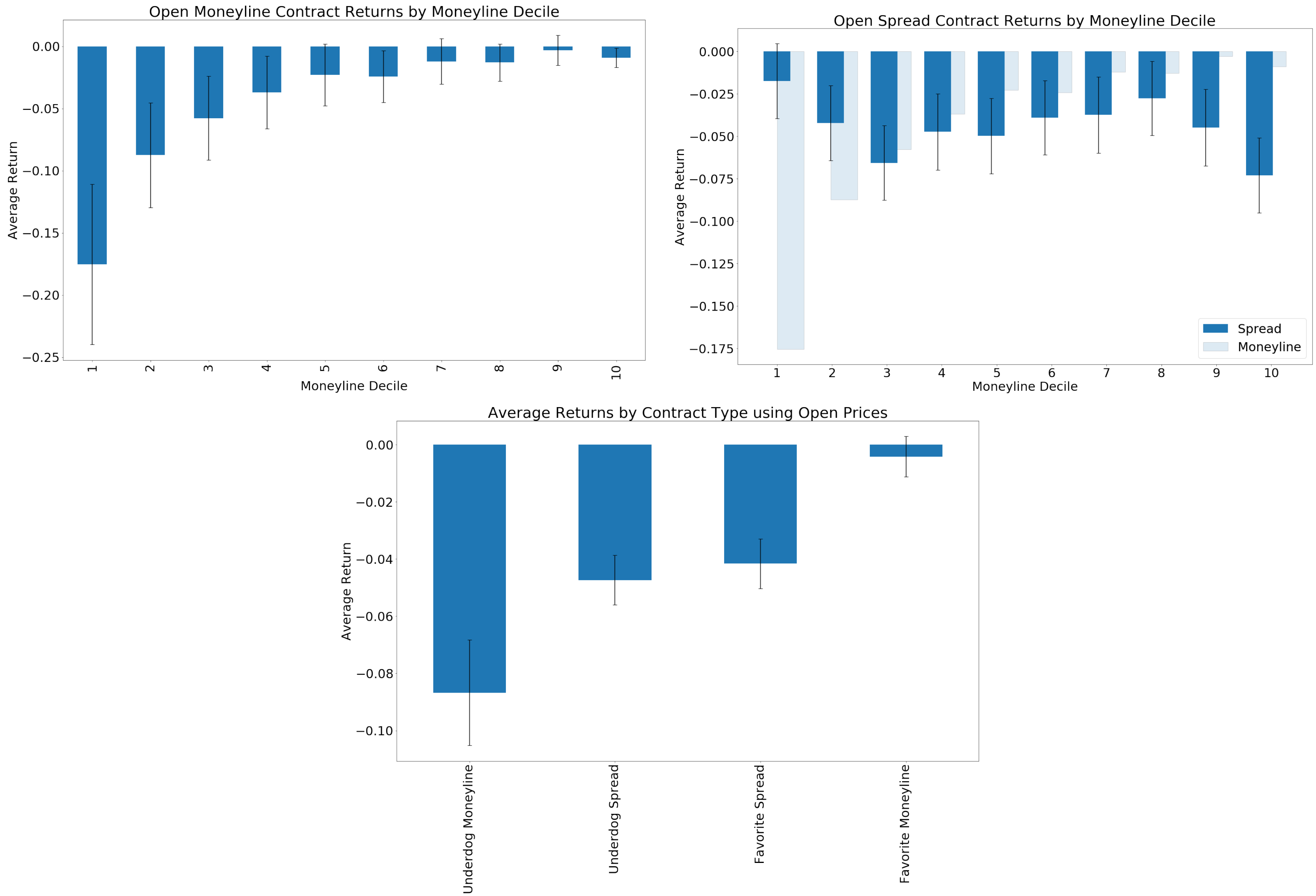


Figure D.7: Average Moneyline Contract Return Across Games

The figure plots the average returns of Moneyline contracts, sorted into deciles based on the Moneyline and split into bets on home and away teams. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. The error bars correspond with two standard errors above and below the average return for a particular decile. The first plot is the subset of bets made on the home team and the second plot is the subset of bets made on the away team.

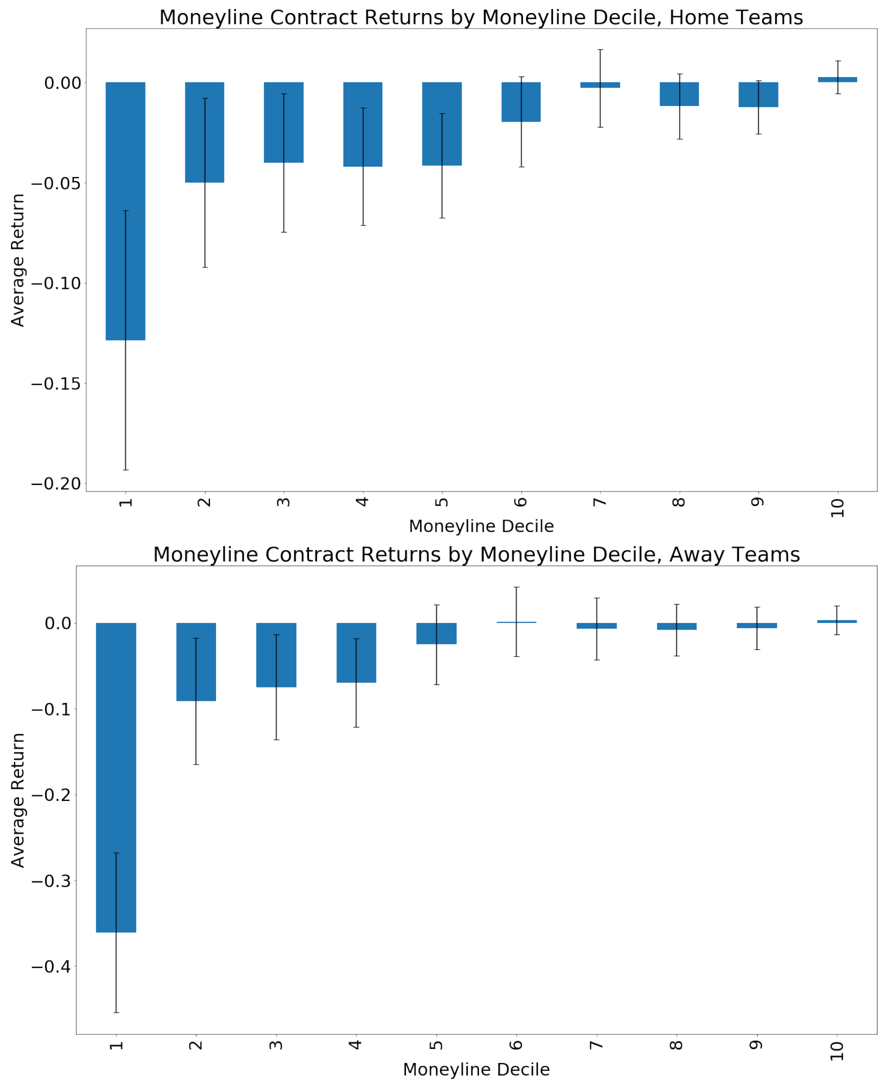
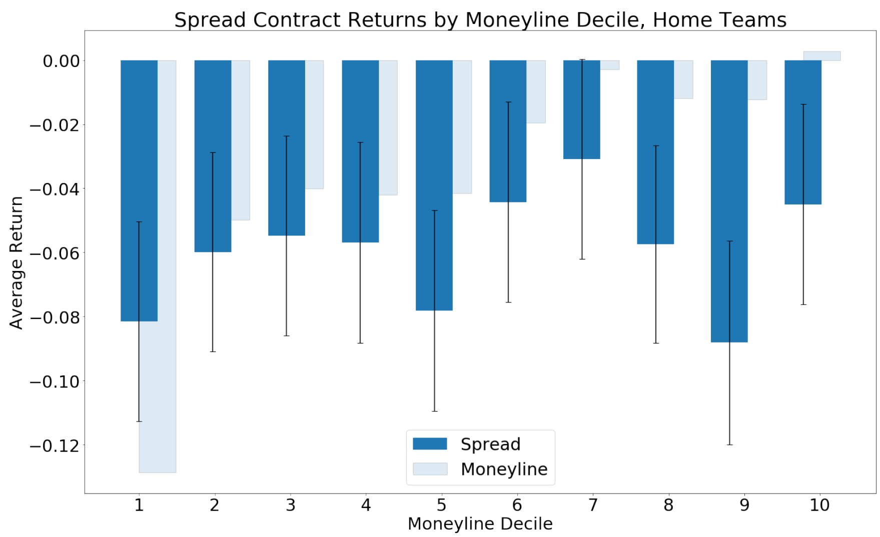
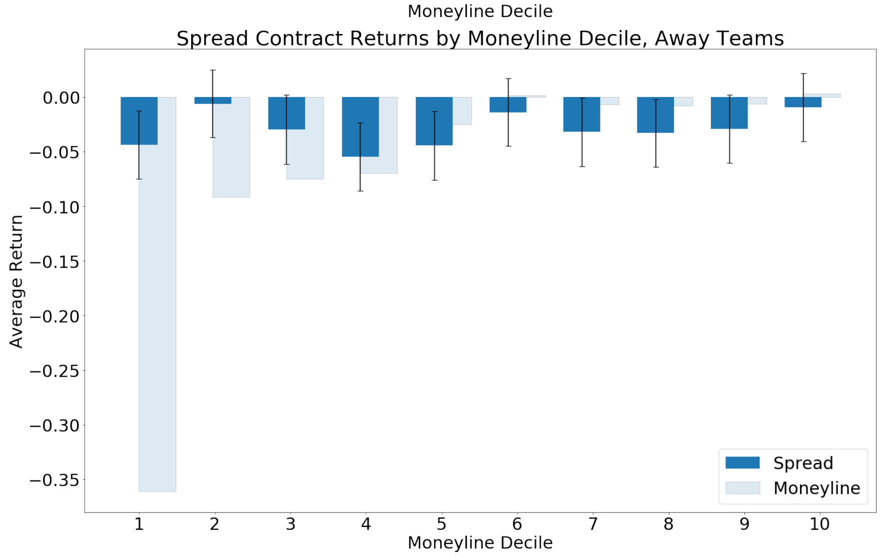


Figure D.8: Average Spread Contract Return Across Games

The figure plots the average returns of contracts sorted into deciles based on the Moneyline and split into bets on home and away teams. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. The darker bars in the foreground correspond with Spread contract returns. The error bars correspond with two standard errors above and below the average return for Spread contracts in a given decile. The lighter bars in the background correspond with Moneyline contract returns, as presented in Figure D.7. The first plot is the subset of bets made on the home team and the second plot is the subset of bets made on the away team.





D.3 Bet Sizes and Odds

In order to estimate bet sizes in our sample, we use the market clearing condition assuming that bookmakers take no risk. Naturally, this assumption may affect the patterns of bet sizes that we estimate in the data. To alleviate potential concerns associated with this assumption, we conduct an out of sample test of the relationship between odds and bet sizes in a sample of soccer matches from Betfair, an exchange with no bookmakers; the results are plotted in the main text in Figure 9. Here, we present additional details about the sample and the data.

For the sample of soccer matches, our data contain the total dollar volume of bets placed on the home and away teams for each game in our sample, as well as the total number of bets on the home and away teams for each game. This allows us to directly measure the average bet size of wagers placed on the home and away teams without making assumptions about bookmakers.

The sample ranges from 2006 to 2011 and contains data on games in eighteen major soccer leagues: the Belgian Jupiler league, the Dutch Eredivisie, the English Championship League, the English Premier League, the French Ligue 1, the French Ligue 2, the German Bundesliga 1, the German Bundesliga 2, the Greek super league, the Italian Serie A, the Italian Serie B, the Portuguese Super Liga, the Scottish Premier League, the Spanish Primera Division, the Spanish Segunda Division, and the Turkish Super League.