What Can Betting Markets Tell Us About Investor Preferences and Beliefs? Implications for Low Risk Anomalies

Tobias J. Moskowitz and Kaushik Vasudevan*

August 29, 2020

Abstract

A robust finding in financial markets is that riskier assets earn lower risk-adjusted returns than less risky assets in equilibrium. A number of theories have emerged to explain this phenomenon that focus on market frictions or preferences for lottery-like payoffs. A related phenomenon from betting markets is the Favorite-Longshot Bias, where returns for betting on riskier "longshots" are lower than betting on "favorites". We synthesize the evidence from the two settings and study their joint implications. Rational theories of risk-averse investors with homogeneous beliefs predict no cross-sectional relationship between diversifiable risk and return, and therefore cannot simultaneously explain these facts. We evaluate preferences versus belief-based explanations for these facts, and conclude that preferences likely play a dominant role. Specifically, our evidence points to Cumulative Prospect Theory preferences as a unifying explanation for the facts in both markets, with quantitative calibrations in the betting data matching those used to explain financial market phenomena.

*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, Stefano Giglio, Jon Ingersoll, Ed Kaplan, Bryan Kelly, Ben Matthies, Nils Rudi, Kevin Zhao and seminar participants at Yale for helpful comments and suggestions.

1 Introduction

A robust empirical finding in traditional financial markets is that, on a risk-adjusted basis, riskier assets earn lower returns than less risky assets in equilibrium. This finding stands in contrast with the predictions of classical models of asset pricing without frictions, such as the Capital Asset Pricing Model, in which all individuals are rational and risk-averse. Starting with Black et al. (1972), a long literature in finance has documented the risk-adjusted underperformance of risky stocks in the cross-section, measuring risk in a variety of ways, such as idiosyncratic stock price volatility (Ang et al. (2006, 2009)) and market beta (Frazzini and Pedersen (2014), and Asness et al. (2020)). The risk-adjusted underperformance of risky assets is also found in other asset classes. For example, riskier options contracts (out-of-the-money put *and* call options) written on the same underlying earn lower returns than less risky contracts (Bondarenko (2014), Ni (2008), Boyer and Vorkink (2014), Frazzini and Pedersen (2020) and Baele et al. (2016)). Frazzini and Pedersen (2014) also document evidence of a similar relationship between risk and return in US Treasuries, Corporate Bonds, Equity Indices, Currencies, and Commodities. A substantial theoretical literature has emerged to explain these results, often relying on models of heterogeneous beliefs and non-standard preferences (most notably, a preference for "lottery-like" payoffs), or alternatively, on capital market frictions, such as leverage constraints.[1]

Another prominent phenomenon of a similar flavor is the Favorite-Longshot Bias in betting markets. First documented by Griffith (1949) and subsequently confirmed as an empirical regularity, the Favorite-Longshot Bias describes a phenomenon observed in betting at the horse racetrack; the returns for betting on a *longshot* to win a race are substantially lower than the returns for betting on a *favorite* to win a race. At the horse racetrack, bets on longshot horses are risky bets that offer a high payoff associated with a low probability event, while bets on favorite horses are less risky bets, offering a low payoff associated with a high probability event. A substantial literature in economics and finance, starting with Weitzman (1965), has used the Favorite-Longshot Bias, and betting markets more broadly, to study risk preferences and decision-making under uncertainty. As Thaler and Ziemba (1988) articulate, betting markets are particularly well-suited to study decision making under uncertainty because bets are gambles on observable, idiosyncratic outcomes that have well-defined termination points and that are not affected by the beliefs and preferences of bettors. These features facilitate the study of how beliefs and preferences influence prices in a cleaner way than is possible in financial markets, where rich dynamics, the exposure of assets to systematic, macroeconomic risk, and the often unobservable nature of the terminal outcomes that assets' payoffs are tied to, confound analysis of the role that beliefs and preferences play in asset price behavior. A central question for the Favorite-Longshot Bias is whether the pattern in returns represents the market having systematically incorrect beliefs of expected outcomes, or whether the pattern reflects a bettor preference for lottery-like payoffs.

[1] Brunnermeier et al. (2007) and Barberis and Huang (2008) are prominent examples of theoretical models that suggest investors have a preference for idiosyncratically skewed assets. A prominent theory of market frictions is the theory of leverage constraints (Black (1972) and Frazzini and Pedersen (2014)). The theory suggests investors seeking higher returns on their investment portfolio increase the weight of risky assets in their portfolio, rather than holding an optimally leveraged position in the market portfolio, due to an inability or unwillingness to borrow. In the model, the result of this behavior is that the equilibrium return on systematically risky assets is lower.

In this paper, we bring additional evidence to bear on the low risk anomaly and the Favorite-Longshot Bias, seeking to bridge the evidence from betting markets to help explain the facts in financial markets. We study a sample of betting contracts written on 36,609 college and professional basketball and football games, spanning two decades, and analyze how the relationship between risk and return in betting markets can inform theories of risk and return in financial markets. Betting contracts in our setting are bets on idiosyncratic sporting events that have observable and exogenous terminal outcomes. In addition, basketball and football game betting have unique features that facilitate direct comparisons to financial markets. First, bets are on the difference in the number of points scored by the two teams playing, which is an observable quantity. We are, therefore, able to observe the underlying *distribution of fundamentals* upon which the contracts are written, unlike at the horse racetrack. Second, there are multiple types of betting contracts. We focus on two in particular, Moneyline contracts and Spread contracts. Both contracts are contingent claims written on the same quantity, the difference in the number of point scored by both teams. However, a Moneyline contract is a fixed-odds bet on which team will win the game (by scoring more points), where the underdog (less favored team) offers a low probability of a high payoff, while the favorite offers a high probability of a low payoff. Spread contracts, on the other hand, bet on the same outcome, but are structured as a bet on the median outcome of the game. A bookmaker sets a spread "line" for each game at approximately the expected difference in the number of points between the favorite and the underdog, where a bet on the favorite pays off when the point difference exceeds the Spread line, while a bet on the underdog pays off when the point difference is lower than the Spread line. Importantly, the potential payoffs and riskiness of Spread bets on the favorite and underdog are exactly the same across contracts.[2] Comparing the Moneyline versus Spread contracts on the same game's outcome provides a unique contrast between two claims on the same payoff state space: one that has different probabilities and payoffs associated with each side of the contract, and one where each side has the same probability and payoffs. This unique comparison allows us to examine any pricing differences on two contingent claims on the same outcome, where the only difference is the idiosyncratic risk and payoffs of the two contracts.

We document three empirical facts in our setting and connect them to evidence in financial markets. First, we show a Favorite-Longshot Bias in our setting. Studying the returns of the cross-section of Moneyline contracts, we find that (less risky) bets on more favored teams earn higher returns and (more risky) bets on less favored teams earn lower returns. This fact parallels the empirical fact that in the cross-section of stocks, less risky stocks earn higher returns and riskier stocks earn lower returns. Second, we document that in the cross-section of Spread and Moneyline contracts written on the same game, riskier contracts earn lower returns and less risky contracts earn higher returns. This fact parallels an empirical feature of options markets, that for options contracts written on the same underlying asset, (more risky) out-of-the-money

[2] Many studies have focused on Spread contracts in teams sports to study the informational efficiency of betting markets (Zuber et al. (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)). Despite the availability of Moneyline contracts in team sports, they have been studied less extensively, likely in part because Spread contracts are the more popular contract to bet on. Works that study the pricing of Moneyline contracts in team sports include Pope and Peel (1989), Woodland and Woodland (1994), Kuypers (2000), and Moskowitz (2015), the latter studying Spread contracts, Moneyline contracts, and a third-type of contract, the Over-Under contract. A unique aspect of our work here is in studying the *joint* implications of pricing patterns in Spread and Moneyline contracts.

options earn lower returns and (less risky) in-the-money options earn higher returns. Third, we show that in the cross-section of Spread contracts, there is no strong relationship between how favored a team is and the returns to betting on that team, in contrast with the pattern observed in the returns of Moneyline contracts *on the same games*. That is, the Favorite-Longshot Bias only exists for the Moneyline contracts, which face different risks, and *not* the Spread contracts, where risk is uniformly set to approximately a 50-50 median outcome. This result does not have a direct parallel in traditional financial markets. However, it is an important result, because it demonstrates that, on average, market estimates of the expected point difference do not reflect a favorite-longshot bias. Given that point differences are the basis on which teams win and lose, the expected point difference is the key piece of information required to determine win probabilities and price all contracts, including Moneylines, within a sport. Hence, our third fact provides evidence against the interpretation that the Favorite-Longshot Bias is explained by incorrect beliefs about game outcomes on the part of the market. Rather, the evidence strongly points toward a preference-based explanation, standing in contrast to recent work studying betting at the horse racetrack.[3]

To illustrate this result clearly and compare it with evidence from financial markets, we construct an implied volatility surface for sports betting contracts akin to the implied volatility surface used to analyze 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. While there are many explanations for the smile in the options implied volatility surface, these are primarily related to misspecification of the probability distribution of the underlying assets' return. Our quantitatively similar results in a simpler state space, with no dynamics, and where the probability distribution is quite easy to measure, highlight that other features of preferences and beliefs likely play an important role.

Our reduced form empirical evidence demonstrates substantial similarity between the cross-section of risk and return in traditional financial markets and betting markets, despite their institutional differences. Putting the facts from both settings side-by-side informs which theories can simultaneously explain both sets of facts. Rational theories that rely on capital market frictions to explain the facts in financial markets, such as leverage constraints (Black (1972), Frazzini and Pedersen (2014)), assume that investors have homogeneous beliefs, are risk-averse, and evaluate risks in the context of the broader set of risks they face. These theories predict no relationship in equilibrium between *idiosyncratic* risks and expected returns. Only systematic, non-diversifiable risks matter. Hence, these theories cannot explain the relationship between risk and return in the cross-section of betting contracts. A unifying explanation for the results in both financial markets

[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 Favorite-Longshot Bias is driven by informational inefficiency. Our interpretation of the results we present is not that betting markets prices are necessarily perfectly efficient: just that market prices reflect the key piece of information required to determine win probabilities and price Moneyline contracts, the expected point difference, in a manner inconsistent with informational inefficiencies driving the Favorite Longshot Bias. This interpretation does not speak to other potential inefficiencies in betting markets, such as the pricing of compound lotteries, a bias for betting on the home team, etc.

and betting markets requires relaxing the assumptions about agents' beliefs and preferences. Moreover, we reject that informational inefficiencies explain the patterns in betting markets.

In the second part of the paper, we quantitatively calibrate preference- and belief-based models to accommodate the Favorite-Longshot Bias and risk-return evidence in financial markets. We use Cumulative Prospect Theory (CPT) preferences (Tversky and Kahneman (1979, 1992)) to model our betting results, which have been used to study a variety of settings, including betting behavior at the horse racetrack (Jullien and Salanié (2000); Snowberg and Wolfers (2010)), individual betting behavior (Andrikogiannopoulou and Papakonstantinou (2019)), options prices (Kliger and Levy (2009) and Baele et al. (2016)), and stock prices (Barberis et al. (2001), Barberis and Huang (2008), Barberis et al. (2016), and Barberis et al. (2019)).[4] We build upon the approach of Jullien and Salanié (2000) and uniquely exploit the average size of bets on each contract, which is not available in other studies of aggregate betting behavior. We find that a CPT specification that includes diminishing sensitivity and probability weighting, with parameter values similar to those that capture the financial markets facts, is able to explain the Favorite-Longshot Bias in betting. In contrast with prior evidence, which suggests that the average bettor may overweight losses in his decision (Jullien and Salanié (2000)), we find mixed evidence for loss-aversion. This difference emerges because we incorporate data on the average amount wagered per bet for each game.

We then explore how belief heterogeneity can explain the findings, using a model of risk-neutral bettors with heterogeneous beliefs about the expected point-differential of games.[5] The logic of belief heterogeneity as an explanation for the Favorite-Longshot Bias is that the marginal bettor that chooses to wager on more extreme underdogs has more extreme beliefs, which are in turn reflected in more negative returns to betting on underdogs. Assuming each bettor bets the same dollar amount on each contract, we first estimate the dollar-weighted distribution of beliefs. We find that belief heterogeneity can capture the empirical Favorite-Longshot Bias, with dispersion in beliefs that is lower than disagreement estimated in financial markets from other studies.[6] However, using data on the actual dollar proportion of bets placed on each team, we find that the heterogeneous beliefs model has more difficulty matching the data. Specifically, a much greater proportion of bets are on the most extreme underdogs (which earn highly negative returns) than on moderate underdogs, which is inconsistent with heterogeneous beliefs models, where the marginal bettor in games with more extreme outcomes should have more extreme beliefs. This evidence suggests that while belief heterogeneity may help explain some of the observed patterns, it cannot explain the popularity of extreme underdogs. Consequently, we conclude that preferences for lottery-like payoffs likely play a dominant role in

[4]Many empirical studies in financial markets primarily focus on one or two of the components of CPT, or are implemented in static contexts (including our own). Barberis et al. (2019) construct a novel equilibrium model that incorporates *all* elements of prospect theory in a multi-period setting, and demonstrate that CPT has a wide-ranging ability to explain a number of stock market anomalies.

[5]In betting markets, earlier studies focusing on belief heterogeneity include Shin (1991, 1992) and Ottaviani and Sørensen (2009, 2010), who focus on the role of private information in betting. More recently, Gandhi and Serrano-Padial (2015) apply the insight that the choice of which horse to bet on in a race can be thought of as analogous to the choice among horizontally-differentiated products, a problem traditionally studied in the literature on industrial organization, to estimate the degree of belief heterogeneity that can rationalize the Favorite-Longshot Bias at the horse racetrack.

[6]In the asset pricing literature, a non-exhaustive sampling of papers studying heterogeneous beliefs include Harrison and Kreps (1978); Scheinkman and Xiong (2003); Diether et al. (2002); Geanakoplos (2010); Simsek (2013), Martin and Papadimitriou (2019), and Giglio et al. (2019). Egan et al. (2020) use a methodology similar to ours to extract beliefs about expected stock market returns using data on investor flows into leveraged ETFs tracking the S&P500.

explaining the facts.

As Barseghyan et al. (2018) note, relatively little work has been done in assessing the stability of risk preferences across field contexts. Our work takes a step in this direction by illustrating that anomalies in traditional financial markets and betting markets parallel one another. Moreover, a similar set of preferences and beliefs calibrated to explain the facts in financial markets can simultaneously rationalize the betting markets facts. While there are institutional differences between the settings, betting markets share a number of similar features with financial markets: large transaction volume, widely available information, market making activity, arbitrage activity (from professional bettors and even hedge funds), and professional analysts. Additionally, bettors prefer to make money rather than lose money, and other preferences related to entertainment and loyalty are secondary motives (as they are in the stock market as well, see e.g., Dorn and Sengmueller (2009) and Grinblatt and Keloharju (2009)).[7] Given the similarities of the features between the settings, there is reason to believe that the facts in the two settings may share similar economic drivers. 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 et al. (2005), and Moskowitz (2015)). Focusing on beliefs and preferences, our findings here suggest that a unifying theory for the pricing of risk in equilibrium across markets features Cumulative Prospect Theory preferences, with heterogeneous beliefs potentially also playing a role.

The rest of the paper proceeds as follows. Section 2 describes the betting markets setting and our data in more detail. Section 3 presents empirical evidence of the Favorite-Longshot Bias in our setting, and relates it to evidence from financial markets. Section 4 discusses the implications of our results for theories across both settings. Section 5 studies Cumulative Prospect Theory as an explanation for the Favorite-Longshot Bias, and Section 6 analyzes the role of belief heterogeneity. Section 7 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 e.g., 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.[9]

[7] The majority of sports betting volume is comprised of investors who use this market professionally and not simply for entertainment, such as professional gamblers and even hedge funds. See e.g., 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.

[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.

We study bets placed on four types of games: National Collegiate Athletic Association (NCAA) Football games, NCAA Basketball games, National Basketball Association (NBA) games and National Football League (NFL) games. For all of the betting contracts we study, a bookmaker controls the "lines" (prices). At the start of betting, bookmakers set an opening line, 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. 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.[10]

Our primary focus is on two types of betting contracts: Moneyline contracts and Spread contracts. The payoffs of Spread and Moneyline contracts 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 $P_A - P_B \lesseqgtr 0$). Spread contracts are bets on whether the point-differential between the two teams exceeds the "Spread Line" set by the bookmaker for the game (i.e., a bet on $P_A - P_B \lesseqgtr \bar{x}$, where $\bar{x}$ is set by the bookmaker). 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 some line set by the bookmaker (i.e., a bet on $P_A + P_B \lesseqgtr T$). We describe the mechanics and pricing of the three types of contracts in more detail below.

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 is favored or team B is favored to win.

For readers familiar with fixed-odds contracts in other settings, Moneylines can be directly converted to

[10]This style of betting is different than 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 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 being 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 Favorite-Longshot Bias for both.

both fractional odds, as quoted in the UK, and decimal odds, as quoted in continental 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, 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 a game for a given wager. 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. The \$10 difference between the amount bet and the amount that can be won is known as the "juice" or "vigorish" or "vig," and is the commission that sportsbooks collect for taking the bet, which is a transactions cost.

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 an approximately 50% probability of paying off. 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 a 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. Bets on the underdog in the Moneyline, which offer a low probability of a high payoff, are the riskiest contract type (measuring risk in terms of idiosyncratic variance or skewness), with risk decreasing in the probability the underdog wins. Moneyline contracts on the favorite, which offer a high probability of a low payoff, are the least risky contract, with contract risk decreasing in the probability the favorite wins the game. The two spread contracts on the favorite and underdog are intermediately risky compared with Moneyline contracts on the favorite and the underdog. Moreover, the Spread bets on the favorite and underdog face the same risk, since they are 50-50 bets or bets on the median outcome. Additionally, since Spread contract payoffs are set to be the same across games, there are no cross-sectional differences in Spread contract risk across games (all are approximately 50-50 bets). However, Moneyline contracts exhibit significant cross-sectional heterogeneity in risk across games as there are significant differences in the quality of the two teams playing across games. We exploit both of these risk differences: 1) for a given game, we compare Moneyline favorite versus underdog against the Spread bet on the favorite versus the underdog; 2) we compare Moneyline favorites versus underdogs across games, where Spread bets across games hold risk constant.

2.5 Assumptions about Transactions and Bookmaker Behavior

For all of our analysis, we use the prices set 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 when using opening prices rather than closing prices.

In our empirical calibrations, we make the assumption that bookmakers do not take any risk, and simply balance the book for each game by adjusting prices such that they make a profit regardless of the outcome of the event by collecting a fee baked into the prices. This assumption is common to the literature studying betting in markets with bookmakers, and appears to be true on average.[11] We further discuss the assumption in our empirical calibrations, when comparing the average size of wagers and the total dollar amount wagered on contracts.

2.6 Data

The data come from SportsInsights.com. The data begin in 2005 and end in May 2013 for all four settings. There are 36,609 total games in the sample: 21,982 NCAA Basketball games, 8,392 NBA Basketball games,

[11] Levitt (2004) and Pope and Peel (1989) suggest that bookmakers may set prices between informationally efficient prices and those that balance the book to earn profits. Snowberg and Wolfers (2010) summarize the empirical results for the Favorite-Longshot Bias in horse betting across a number of studies that use data from both parimutuel and bookmaker-based markets, and find that the pattern is nearly identical across the settings.

4,392 NCAA Football games and 1,843 NFL Football games.

The betting lines are drawn from the Las Vegas legalized sportsbooks and online betting sportsbooks, where all bookmakers offer nearly identical lines on a given game. The data include all games from the regular season, pre-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. 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. The three sportsbooks are collectively considered the "market setting" sportsbooks that dictate pricing in the U.S. market.[12]

3 Empirical Analysis of Betting Contract Returns

3.1 Three Facts about Betting Contract Returns

We begin by documenting three empirical facts in betting markets that serve as the cornerstone for the analysis.[13] We relate each of these facts to analogous facts in financial markets.

First, we document the Favorite-Longshot Bias in our setting. We find that in the cross-section of Moneyline contracts across games, riskier bets (bigger underdogs) earn significantly lower average returns relative to less risky bets (on bigger favorites). We sort Moneyline contracts into deciles based on the Moneyline, with decile 1 corresponding with extreme underdogs and decile 10 corresponding with extreme favorites.[14] Figure 1 plots the average return in each decile. The average returns are monotonically decreasing in risk. Less risky contracts earn significantly higher average returns than more risky contracts. The biggest underdogs (riskiest contracts) in decile 1 have average returns of -22.60%. Contracts in decile 10, which correspond to betting on the biggest favorites (the safest contracts), earn an average return of -0.51%. The fact that all returns are negative is due to the vig or transaction cost in betting markets, which at 10% is substantially punitive that average returns are negative. Moreover, this result has a direct parallel in financial markets. In the cross-section of stocks, riskier stocks tend to earn lower risk-adjusted returns than less risky stocks. In the second panel of Figure 1, we plot the CAPM alphas (the risk-adjusted returns) of US stocks sorted into decile portfolios based on their risk at each point in time (CAPM beta), drawing our numbers from Frazzini and Pedersen (2014). The pattern in the plot is similar to the pattern in sports betting. Stocks with the lowest betas substantially outperform, earning a monthly CAPM alpha of 0.52 basis points per month, while stocks with the highest betas underperform substantially, earning a CAPM alpha of -10 basis points per month.

Second, we document that in the cross-section of contracts written *on the same game*, riskier contracts earn lower average returns and less risky contracts earn higher average returns. Figure 2 plots the average return of spread and Moneyline contracts on the favorite and the underdog in each game, including error

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

[13] We use closing prices in the calculations presented here. We replicate the results with opening prices in Appendix Figure A.7. The patterns remain the same.

[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 further when interpreting the results.

bars corresponding with 95% confidence intervals for the average return for each game. The riskiest type of contract (betting on the underdog in the Moneyline) earns an average return of -8.5 percent. The least risky contract (betting on the favorite in the Moneyline) earns an average return of -1.5 percent. The two intermediately risky contracts (which have the same risk), the spread contracts on the underdog and favorite, each earn average returns of -4.1 and -4.8 percent, which are not statistically different from each other. This result has a direct parallel in options contracts; riskier (out-of-the-money) options earn lower risk-adjusted returns than less risky (at-the-money) options. In the second panel of Figure 2, we plot the delta hedged excess returns of one-month maturity equity options and index options, sorted into five buckets based on the absolute value of their delta, drawing our numbers from Table III of Frazzini and Pedersen (2020). The portfolios are ordered in terms of "moneyness", from "deep out of the money" (the riskiest option contracts) to "deep in the money" (the least risky option contracts). The plot shows that delta-hedged returns are decreasing from the deep out of the money portfolio (monthly delta hedged excess returns of -17.26% and -29.23% for equity options and index options) to the deep in the money portfolio (monthly delta hedged excess returns of -3.36% and -1.26%). Frazzini and Pedersen (2020) attribute this return pattern to the embedded leverage of the options contracts; investors seeking leveraged exposure buy the (riskier) out of the money options, in turn lowering the equilibrium return for them. This explanation, however, cannot explain the sports betting results because the sports contracts are idiosyncratic. Leverage exposure is about systematic risk that is compensated in equilibrium. There is no such exposure in betting markets, hence an alternative explanation must be present.

Third, we document that in Spread contracts, where there is no cross-sectional heterogeneity in the risk of contracts, the average returns of different contracts do *not* appear to vary with the likelihood that the favorite wins the game. As we do with Moneyline contracts, we sort spread contracts into deciles based on the Moneyline, with decile 1 corresponding with extreme underdogs and decile 10 corresponding with extreme favorites. This is the exact same sort as before, but now instead of betting on the Moneyline, we bet on the Spread contract on the *same games*. Figure 3 plots the average return of contracts in each decile. The returns in the deciles range from -2.8% to -6.3%, but the returns are not statistically different across any of the deciles and there is no systematic pattern. To the extent there is a pattern, it is in the opposite direction of the Favorite-Longshot Bias. This lack of or opposite pattern is much weaker than the Favorite-Longshot Bias in Moneyline contracts, is not statistically or economically significant, and can be resolved completely by comparing bets made on home teams across games and bets made on away teams across games (see Appendix Figure A.9).[15] This result indicates that it is the *risk* of the contract that matters to investors and not some other characteristic of the game. For example, this result rules out that favorites versus longshots are mispriced, since such mispricing should show up in the Spread contract as well as the Moneyline contract. Put differently, the only difference between Figure 3 (Spread contracts) and Figure 1 (Moneyline contracts) is the risk of the contract. Everything else remains exactly the same because these are two different contracts on the same game. Hence, we have a controlled experiment betting on the same outcome for the same contest, where the only variation is the risk of the contract. The stark difference

[15] The evidence suggests that the pattern is explained by the fact that there are systematically lower returns for betting on the home team.

in the results points to risk, in this case idiosyncratic risk, being the only difference between the two bets. This result does not have a clear parallel in financial markets. However, it is a useful placebo test for our setting that rules out other confounding factors.

3.2 Quantifying the Results: The Betting Implied Volatility Surface

Each of the betting contracts that 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 threshhold, 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.

For this exercise, we assume that the realized point differential minus the Spread line is independently, identically normally distributed across games. We can statistically reject this assumption in the data (e.g., point differentials are discrete), but it holds reasonably well as an approximation. We empirically analyze this distributional assumption (which is also used in our calibration of belief heterogeneity) in Appendix A.1. The exact distributional assumption is not especially important for this exercise, but is used as a tool to express the magnitude of the pricing patterns we observe in terms of standard deviation.

To construct the implied volatility for each 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 embedded in the contract price, $\hat{p}$. Using p and $\hat{p}$, we calculate the implied volatility of a contract, $\hat{\sigma}$, as the value of $\hat{\sigma}$ that satisfies Equation (4):

$$F\left(F^{-1}(p;\sigma);\hat{\sigma}\right) = \hat{p} \tag{4}$$

where F is the cumulative distribution function of the point differential minus the Spread line distribution, and σ is its standard deviation. For each sport, we fit σ using Maximum likelihood.[16] We calculate the implied probability, $\hat{p}$, directly from contract prices as the probability of a contract payoff that sets the contracts expected return equal to the expected return of a Spread contract with 50% chance of paying off. To estimate the objective 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

[16] An analogous 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 alternative method yields similar results, but is more reliant upon the parametric assumption that the point difference minus Spread line is *iid* normal across games. The method we pursue primarily relies upon the accuracy of the estimated probabilities, which we are able to estimate without resorting to parametric assumptions. In our method, the assumption of normality is only used to quantify the magnitude of probability distortions in terms of the standard deviation of a normal distribution.

present evidence of the accuracy of the estimated win probabilities in Appendix Section A.2.[17]

For our analysis here, we exclude contracts where p is between 0.45 and 0.55. This is because our methodology is not particularly well-suited to handle win probabilities very close to 50%. This can be seen, for example, by the fact that Equation (4) has an undefined solution at $p = 0.5$. The restriction we impose excludes 31% of contracts. However, our interest is particularly in studying contracts with low and high values of p, which we are still able to do despite this restriction.

We compute implied volatilities for each Moneyline contract that satisfies our restriction on win probabilities. For each sport, we sort contracts into one of 90 equally spaced bins based on the estimated win probability of the contract (i.e., one bin for contracts with win probability between 0% and 1%, one bin for contracts with win probability between 1% and 2%, etc.). Figure 4 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 ("in-the-money") and for contracts that have a very low probability of paying off ("out-of-the-money"). Additionally, the implied volatility surface appears to be asymmetric. That is, the implied volatility for a contract with a win probability of, say, 70%, is higher than the implied volatility for a contract with a loss probability of 70%. This asymmetry suggests that bettors treat gains and losses differently. We discuss this result further when we quantitatively calibrate the preferences of a bettor with Cumulative Prospect Theory.

We provide context for the magnitude of the implied volatility surface smile and compare it with the magnitude of the implied volatility smile found in options markets. Panel A of Table 1 displays the average implied volatilities for contracts expected to pay off approximately 5-15% of the time and 85-95% of the time, using our non-parametric estimates of p. 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 vols that are 0.0% (for NBA contracts) to 13.8% (for NCAA Football contracts) higher than the true sample volatility, with an average premium of 6.8% across sports. Contracts expected to pay off 85% to 90% of the time have implied volatilities that are 12.4% (for NBA contracts) to 30.7% (for NFL contracts) higher than the true sample volatility, with an average premium of 22.8% across sports.

For comparison, Panel B of Table 1 presents the simple average implied volatility for "standardized" 10 (and -10) delta and 90 (and -90) delta call (and put), one month maturity options for a set of 13 indices. These correspond roughly to payout probabilities of 10 and 90%, respectively. Panel C presents the same quantities averaged across all equity options from the OptionMetrics IvyDB Implied Volatility Surface file.[18]

[17]We obtain similar results by directly estimating win probabilities via kernel regression, but find that directly modeling returns better captures the behavior of contract expected returns and win probabilities, especially for contracts with very low and very high probabilities of paying off.

[18]We follow Frazzini and Pedersen (2020) 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

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 options, which have payoff probability of approximately 50%. The results presented in Panel B show that 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 vol 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 at 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.

There are many reasons for volatility smiles and smirks to exist in options markets, which primarily come from the difficulty in assessing the true probability distribution of underlying asset returns. Implied volatilities are calculated via the Black-Scholes formula, which assumes that stock returns are log-normally distributed. The true distribution of asset returns is often thought to be fatter tailed than the the log-normal distribution, coming from different features of the asset return process, such as stochastic volatility and jumps. Exposure to aggregate risk also may play a role in the implied volatility smile. For example, deep-out-of-the money put options provide a hedge against market crashes; the implied volatility smirk in index options is also, in part, driven by risk premia associated with market crashes.

However, the volatility smile can also come from other features, such as non-traditional preferences or belief heterogeneity. These features are qualitatively consistent with the delta-hedged underperformance of risky, out of the money options, documented for example by Boyer and Vorkink (2014). The implied volatilities we document are of a similar magnitude to those found in options on single name stocks. This evidence further supports non-traditional beliefs or preferences may play a role in options markets. 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 is unlikely to drive our results. Rather, our results point to other explanations that might be at play, which we spend the rest of the paper exploring.

4 Interpreting the Evidence

The empirical facts that we present bring new evidence to the Favorite-Longshot Bias. In this section, we begin by discussing the implications that our results have for explanations posed to explain the Favorite-Longshot Bias and then proceed to discuss the implications for a unifying theory to explain the facts in

correspond with any single traded option. See the appendix for more details.

financial markets and betting markets.

4.1 Implications for Theories of the Favorite-Longshot Bias

A central debate in the literature on the Favorite-Longshot Bias is whether the it represents a market-level informational inefficiency (see Thaler and Ziemba (1988) for a discussion of various explanations for the Favorite-Longshot Bias). This debate generally centers around two questions: does the Favorite-Longshot Bias occur as a result of the market incorrectly assessing the outcomes of sporting events, and as such, represent a mispricing? Or, is the pattern a reflection of something else, such as a preference for lottery-like payoffs among bettors? And, in either case, is there a profitable opportunity for informed parties to exploit? Although earlier studies have found profitable systematic betting opportunities, more recent evidence indicates no – the Favorite-Longshot Bias does not represent a profit opportunity, as the transaction costs from trying to exploit the pattern exceed any potential profits. We similarly in our sample do not find reliable evidence that there is a profitable betting opportunity from the Favorite-Longshot Bias.[19] The debate is still very much open, however, as to whether mispricing or preferences are driving the Favorite-Longshot Bias. Snowberg and Wolfers (2010) use the prices of exotic bets on the order in which horses will finish as evidence that the market does not properly reduce compound lotteries. They use this as supporting evidence for the interpretation, more broadly, that the Favorite-Longshot Bias represents a misperception of horse win probabilities on the part of the market. Taking a rich non-parametric approach, Chiappori et al. (2019) find evidence that preferences and probability misperceptions both play a role.

The results we present are evidence against the Favorite-Longshot Bias representing a market-level informational efficiency. The key piece of evidence for this interpretation is the third empirical fact we documented: the absence of a systematic pattern of higher returns for betting on favorites and lower returns for betting on underdogs in Spread contracts *in the same games* that we observe a Favorite-Longshot Bias in Moneyline contracts. This result suggests that market prices are reasonably accurate in forecasting the expected difference in points scored in the game. To the extent that market prices might be inaccurate in forecasting the point difference, the expected point differences embedded in prices are not systematically optimistic about underdogs and pessimistic about favorites in a way that would explain the Favorite-Longshot Bias. This point is notable, because in a game, which team wins and which team loses is entirely determined by the point difference of the game. Accordingly, the expected point difference is the key piece of information required to determine win probabilities and appropriate prices for Moneyline contracts. The fact that the expected point differences from the Spread lines on each game do not display a Favorite-Longshot Bias suggests that incorrect assessments of the expected outcome of the game are unlikely to explain the Favorite-Longshot Bias in our setting.

There are two alternative explanations that stand out against our interpretation. One alternative explanation is that the two markets are informationally segmented. A second, more subtle alternative explanation is that market participants accurately perceive the expected point difference for games, but they mistakenly

[19]Early studies that find evidence of profitable betting strategies e.g., Hausch et al. (1981). More recent studies, such as Levitt (2004) and Snowberg and Wolfers (2010), do not find evidence of profitable betting opportunities. Admittedly, we did not attempt to devise betting strategies more complex than simple sorts of contracts based on the expected outcomes of games.

believe that the distribution of point differences has more dispersion than it does in reality. Bettors with such mistaken beliefs would properly price Spread contracts, but would underestimate the probability of the favorite winning and overestimate the probability of underdogs winning in a manner that would result in a pattern of equilibrium returns similar to the Favorite-Longshot Bias in Moneyline contracts.

We provide additional evidence to support our interpretation versus these potential alternative explanations. First, we show Spread contract lines span nearly all relevant information regarding win probabilities that is embedded in Moneyline contracts. This supports the claim that expected point differences are the key piece of information required to determine win probabilities. To do so, we run regressions of the form

$$\mathbb{I}_{\text{home win},i} = \alpha_s + \beta_{\text{Spread,s}}\text{SpreadLine}_i + \beta_{\text{ML,s}} \log\left(\frac{1 + y_{h,i}}{1 + y_{a,i}}\right) + \epsilon_i \tag{5}$$

where for game i, $\mathbb{I}_{\text{home win},i}$ is a 0/1 indicator if the home team won the game, 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 $\log\left(\frac{1+y_{h,i}}{1+y_{a,i}}\right)$ ("the Moneyline ratio"), captures how favored the home team is versus the away team as implied by the Moneyline, with smaller values corresponding with the home team being more favored and larger values with the home team less favored. $\beta_{Spread,s}$ and $\beta_{ML,s}$ are separate regression coefficients for each sport s. We standardize the independent variables within sports to have unit mean and standard deviation.[20]

We estimate three versions of the regression: one including only the Spread line as an independent variable, one including only the Moneyline ratio as an independent variable, and one including both. Table 2 reports the results from the regressions. The first two columns show that the Spread Line and Money lines have very similar return predictability for wins, with nearly identical, highly-significant regression coefficients (ranging from -0.18 for the NFL to -0.22 for NCAA Basketball, indicating that a one-standard deviation change in the independent variables corresponds with an 18-22% increased win probability for the home team) and R^2 values (20.85% for the Spread line regression and 20.66% for the Moneyline regression). The third column shows the results from the multivariate regression. The coefficient on the Spread line remains significant for three of the four sports (insignificant for the NFL), while the coefficient on the Moneyline ratio is only significant for NCAA football. The R^2 for the multivariate regression is 20.87%, indicating that the multivariate regression adds almost no value over the univariate regressions, and that the Money lines and Spread lines capture the same predictive information for wins. The F-statistic for comparing the multivariate regression with the univariate Spread regression is 0.0367, meaning that we cannot reject the univariate regression in favor of the multiple regression, consistent with the Spread line and Moneyline capturing common information for wins. We plot a binned scatterplot of the Moneyline variable versus home win percentages in Appendix Figure A.1, both including and not including a control for the Spread Line. The figure verifies that win probabilities are monotonically decreasing in the Moneylines, but that the Spread Line captures almost all of the Moneyline's explanatory power, and that alternative functional forms

[20]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 linear. Alternative formulations have *worse* predictive power for wins.

for the relationship between win probability and Moneyline implied probability are unlikely to explain the results.

Second, we show that the distribution of point differences appears to be relatively constant across games. In Figure A.2, we plot the standard deviation of the point-differential minus Spread line of all games, sorted into deciles by the Moneyline ratio in each sport. For the most part, the standard deviations are largely constant across deciles. There is no systematic pattern in standard deviations across the deciles (e.g., point-spreads being more or less dispersed in games where the home team is more favored). To the extent there are differences in the standard deviation across deciles, they are economically small compared to the implied volatilities shown in Figure 4 and Table 1.

Third, we show that market prices in the Over/Under market accurately reflect the total number of points scored by both teams in games. We sort games into twenty bins based on the Over/Under line of the game, and plot a binned scatterplot of the Over/Under line versus the total number of points scored in Figure 5. All points on the line lie on the 45 degree line, illustrating that the Over/Under contracts accurately capture point totals in games. We do the same exercise, and find similar results when separately analyzing the results in each sport in Figure A.4 (though the estimates are noisier given the smaller sample sizes). The second panel in Figure 5 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 of score of games regardless of how close or lopsided the game is expected to be.

Taken together, the additional evidence paints a picture of the information captured by market prices for game outcomes. Given the results that the information in Spread lines and Moneylines regarding win probabilities nearly completely overlaps, the evidence is inconsistent with the alternative interpretation of Moneyline and Spread line markets being informationally segmented from one another. Informational segmentation is also not consistent with the fact that the bookmakers that set prices in the Spread line market are the same bookmakers that set prices in the Moneyline market.

Moreover, while we cannot explicitly rule out that the second alternative explanation that the market accurately captures the expected point difference inaccurately captures the higher moments of the point spread distribution, the evidence suggests this alternative explanation is unlikely to be true. Our results show that (1) the standard deviation of point spreads appears relatively constant within a sport regardless of the expected outcome of the game and (2) the market accurately estimates the total number of points in an unbiased way, though total points can vary substantially across games. For the alternative explanation to be true, the market must display extreme sophistication in forming expectations about the number of points each team will score in each game (a conditional random variable that varies substantially across games), but does not display the same sophistication in assessing the relatively static features of the point spread distribution across games. This seems highly unlikely given that the market treats point spreads as close to sufficient for determining win probabilities as Spread lines and Moneylines capture almost entirely overlapping information for the game outcome.

The results indicate the efficient pricing of contracts by the market. However, it does not address the possibility that different bettors sort into different contracts based on heterogeneous (potentially incorrect) beliefs. This possibility we explore in more detail later in the paper. However, the lack of a reliable profit opportunity in our setting suggests that even in the case where bettors sort into contracts based on potentially incorrect beliefs, the Favorite-Longshot Bias still does not reflect informational inefficiency, at least in the sense that it does not appear to permit a reliable profit opportunity to informed arbitrageurs.

4.2 Implications for Theories in Financial Markets

In the previous section, we show financial markets facts that are analogous to those in sports betting. Putting the results together side-by-side highlights the similarities between the two settings. What types of theories are able to simultaneously explain the facts in both settings?

"Rational" theories that explain the underperformance of riskier assets in traditional financial markets via market frictions assume that people are risk-averse and have homogeneous, correct beliefs. In the absence of frictions, they evaluate the riskiness of any asset in the context of their wealth portfolio, and diversify away any idiosyncratic risk. Under these assumptions, the only priced risks in financial markets should be systematic, non-diversifiable risk, and equilibrium returns should be higher for riskier assets to compensate investors for bearing those risks. Such a paradigm, however, is not able to explain the existence of any patterns in the cross-section of returns of sports betting contracts, which are bets on purely idiosyncratic risk. Hence, theories from financial markets based on market frictions alone are unable to simultaneously accommodate the facts in betting markets.

Departing from the rational paradigm, our results also provide some direction for the types of behavioral models that can explain the data. For example, the fact that incorrect market forecasts of the expected point difference of the game do not explain the Favorite-Longshot Bias suggests that a unifying theory of the facts is unlikely to be driven by excessive optimism of fundamentals about risky assets. However, one way to potentially accommodate the results from sports betting markets and financial markets is non-traditional preferences. For example, under *narrow framing*, agents evaluate gambles in isolation from the other risks they face elsewhere (Kahneman (2003)); this assumption can explain the existence of patterns in the cross-section of asset returns where there are no difference in exposure to aggregate risk. There is also evidence of narrow framing in traditional financial markets (Barberis and Huang (2001) and Barberis et al. (2006)). Assuming narrow framing, a preference for lottery-like payoffs, as in Cumulative Prospect Theory, can generate predictions that less risky assets earn higher returns than riskier assets in both markets.

While we believe narrow framing is likely to be at play in explaining betting market patterns, it is worth pointing out that Cumulative Prospect Theory, and particularly its probability weighting component, may deliver the patterns in betting returns in a way that the rational paradigm cannot. As Barberis and Huang (2008) show, an individual with CPT preferences that evaluates risks in the contexts of all other risks he faces may still demand positively, idiosyncratically *skewed* assets with negative returns (such as bets on extreme underdogs). Such assets can increase the overall skewness of the individual's portfolio, which the individual finds desirable.

An alternative way to accommodate our results may come from heterogeneous beliefs. Heterogeneous beliefs can potentially rationalize the Favorite-Longshot Bias, if bettors sort into contracts based on their beliefs about the outcome of the game. Survey evidence in financial markets also suggests the presence of substantial belief heterogeneity, which has been shown to explain pricing patterns found in the cross-section of stock returns. Belief heterogeneity may therefore be a unifying explanation.[21]

For the remainder of the paper, we explore departures from standard rational models in the form of non-expected utility preferences and heterogeneous beliefs, to better understand their ability to explain the patterns in sports betting and traditional financial markets. Our calibrations are particularly focused on the pricing of Moneyline contracts in our setting, where we observe cross-sectional variation in risk and return. We begin by focusing on a preference based-channel for the Favorite-Longshot Bias using Cumulative Prospect Theory preferences, and use the calibrated preferences to model the evidence in financial markets. Then, we model belief heterogeneity and estimate the degree of disagreement implied by the data to match the results.

Finally, our data offer a unique advantage over previous studies in that we are able to analyze the proportion of bets placed on each contract and the size of bets placed. This feature of the data allows us to test these models in a novel way. For instance, we find evidence that bettors may be loss-tolerant rather than loss-averse when using actual bets as opposed to assumping that bettors bet equal amounts across contracts. Additionally, belief heterogeneity assumes that more extreme beliefs correspond with more extreme odds, yet our data on proportion of bets placed shows that almost half of bettors wager on the most extreme underdogs, which is inconsistent with this explanation and suggests that preferences likely play an important role.

5 Cumulative Prospect Theory and the Favorite-Longshot Bias

We begin by studying the preferences of bettors that are able to rationalize the Favorite-Longshot Bias in our data, and more specifically the preferences of a bettor that has correct beliefs and is indifferent between wagering the average amount on the two teams playing.

Our focus is on Cumulative Prospect Theory (CPT) preferences (Tversky and Kahneman (1979) and Tversky and Kahneman (1992)). First, our empirical results qualitatively align with the predictions of CPT, and particularly its predictions that people have a preference for positively skewed payoffs and that they may treat gains and losses differently. Second, Cumulative Prospect Theory has been shown to be consistent with facts about the cross-section of asset returns in traditional financial markets (Baele et al. (2016), Barberis and Huang (2008), Barberis et al. (2019)). Third, other work suggests that features of prospect theory are able to explain the Favorite-Longshot Bias at the horse racetrack (Jullien and Salanié (2000) and Snowberg and Wolfers (2010)). However, these studies assume forms of CPT that are not directly comparable with

[21] In addition to heterogeneous beliefs, bettors may also have heterogeneous preferences. For example, Chiappori et al. (2019) examine preference heterogeneity as an explanation for the Favorite-Longshot Bias at the horserace track and Andrikogiannopoulou and Papakonstantinou (2016) analyze heterogeneity in preferences by examining the behavior of individual bettors. Chiappori et al. (2019) find that allowing for preference heterogeneity has limited additional explanatory power for the Favorite-Longshot Bias in quantitative calibrations. Andrikogiannopoulou and Papakonstantinou (2016) find evidence of preference heterogeneity across individuals. We do not explore this alternative channel of preference heterogeneity, but recognize that preference heterogeneity may also play a role in pricing patterns.

structural estimates from financial markets and other settings.[22] Our goal is not to evaluate whether non-expected utility models provide a better fit for the data than expected utility models (as in Jullien and Salanié (2000) and Chiappori et al. (2019)), but rather to understand whether CPT is able to explain the Favorite-Longshot Bias with parameters that are consistent with those able to capture the related financial markets facts.

5.1 CPT Preferences and Betting

Consider a wager of b on a simple binary outcome. The bet pays off with probability p, in which case the bettor receives yb dollars. With probability $(1-p)$, the bet does not pay off, and the bettor loses the b she wagered. Under expected utility preferences, a bettor would evaluate this bet by computing

$$V = pU(W + yb) + (1-p)U(W-b) \tag{6}$$

where W is the bettor's total wealth and U is her utility function. In contrast, a bettor with Cumulative Prospect Theory preferences evalutes this bet by computing

$$V = w(p)v(yb) + w(1-p)v(b) \tag{7}$$

where $w(\cdot)$ and $v(\cdot)$ are known as the probability weighting function and the value function. Tversky and Kahneman (1992) propose the functional forms

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \tag{8}$$

and

$$v(x) = \begin{cases} x^\alpha & x \geq 0 \\ -\lambda(-x)^\alpha & x < 0 \end{cases} \tag{9}$$

where $\alpha, \delta \in (0,1)$ and $\lambda > 1$.[23]

Relative to the expected utility preference specification in Equation (6), the CPT preference specification in Equation (7) has four important differences. First, under prospect theory, the bettor evaluates her gains or losses relative to a reference point, rather than evaluating her terminal wealth, as under expected utility (i.e., v is computed over gains and losses, whereas $U(\cdot)$ is computed over terminal wealth). Here, we assume

[22]Using data on individual bettors, Andrikogiannopoulou and Papakonstantinou (2016, 2018, 2019) are able to observe betting amounts, and use them to calibrate the preferences of individual bettors. Their work highlights heterogeneity in preferences across individuals, and documents that individuals respond to their past betting performance in a manner consistent with cumulative prospect theory. The primary goal of their work, however, is not to explain patterns in equilibrium prices and returns via bettor preferences, which is the goal of our analysis.

[23]Under the usual formulation of Cumulative Prospect Theory preferences, outcomes are ranked by how preferred they are, separated into gains and losses, and the probability weighting function is applied over cumulative probabilities. Computing the probability weighting function over cumulative probabilities, proposed by Quiggin (1982), allows for computing utility for non-binary gambles and respects first-order stochastic dominance, unlike the original formulation of prospect theory. However, the Moneyline contracts we study are binary gambles, so applying the probability weighting function over cumulative probabilities reduces to applying the function over probabilities. Our application does not rely upon rank-dependence.

that the reference point is zero, the bettor's payoff if they do not take the bet.[24]

Second, under prospect theory, the bettor exhibits *loss-aversion*. That is, $v(\cdot)$ is kinked at zero, so the bettor is more sensitive to small losses than gains of the same magnitude. The parameter λ governs how loss-averse the bettor is. In contrast to the expected utility specification, $U(\cdot)$ is differential everywhere and exhibits no such kink.

Third, $v(\cdot)$ is concave over gains and convex over losses. The bettor is risk-averse over gains, but risk-loving over losses. α governs the concavity and convexity of the value function over gains and losses. Concavity over gains and convexity over losses leads to a feature called *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. In contrast, under expected utility, $U(\cdot)$ is typically assumed to be concave everywhere, especially in finance applications.[25]

Finally, under CPT, the bettor engages in *probability weighting*. She uses distorted probabilities to compute her utility for a bet, rather than objective probabilities, as under expected utility. Probability weighting is motivated by psychological evidence from the laboratory and field evidence that people tend to systematically overweight low-probability events and underweight high-probability events (Fehr-Duda and Epper (2012), and Barberis (2013) discuss some of the evidence). The parameter γ governs how much the person distorts probabilities, with smaller values of γ corresponding with more overweighting of tails of the distribution. Under CPT, the standard interpretation of probability weighting is not necessarily that the bettor does not know objective probabilities, but rather that the bettor gives additional weight to tail events in her decision.

5.2 Taking CPT Preferences to the Data

5.3 Methodology

Consider two teams that are playing against each other in a game, indexed as team 1 and team 2. Before the game is played, bettors are indifferent between betting wagering b_1 dollars on team 1, which offers a potential payoff y_1 per dollar wagered, and wagering b_2 dollars on team 2, which offers a potential payoff of y_2 per dollar wagered. With CPT preferences, the bettors' indifference condition is expressed as

$$w(p)(y_1 b_1)^{\alpha} - w(1-p)\lambda b_1^{\alpha} = w(1-p)(y_2 b_2)^{\alpha} - w(p)\lambda b_2^{\alpha} \tag{10}$$

[24]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 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 e.g., Grinblatt and Han (2005), Barberis and Xiong (2009), Imas (2016), and Barberis et al. (2019) for analysis of this idea as it relates to stock price behavior, and Andrikogiannopoulou and Papakonstantinou (2019) for evidence from individual bettors. See e.g., O'Donoghue and Sprenger (2018) for a broader discussion of reference dependence.

[25]Under expected utility, $U(\cdot)$ may also be convex, which corresponds to risk-loving preferences. In the betting context, when assuming expected utility preferences, the literature has found evidence of a convex utility function e.g., Weitzman (1965). This evidence has been interpreted as local risk-love, as suggested by Friedman and Savage (1948). However, the literature in finance almost always assumes that $U(\cdot)$ is concave.

We re-write this condition to derive an expression for $p(\cdot)$, the probability that team 1 wins the game, as a function of preference parameters ($\Theta \equiv (\alpha, \gamma, \lambda)$) and characteristics of the game ($\mathbf{y} \equiv (y_1, y_2, b_2/b_1)$).

$$p(\mathbf{y}; \Theta) = \frac{\left((y_2 b_2/b_1)^{\alpha} + \lambda\right)^{\frac{1}{\gamma}}}{\left((y_2 b_2/b_1)^{\alpha} + \lambda\right)^{\frac{1}{\gamma}} + \left(y_1^{\alpha} + \lambda\,(b_2/b_1)^{\alpha}\right)^{\frac{1}{\gamma}}} \tag{11}$$

Indexing each game in the sample $i = 1, \ldots, n$ and the team that ex-post wins each game as team 1, we can follow the approach of Jullien and Salanié (2000), and estimate Θ as the parameters that maximize the likelihood of the sample given in Equation (12).

$$\mathcal{L}(\Theta) = \prod_{i}^{n} p(\mathbf{y_i}; \Theta) \tag{12}$$

To take this specification to the data, we require estimates of $\mathbf{y}$ for each game. y_1 and y_2, the payoffs associated with winning bets, can be directly determined from contract prices. The other remaining piece of $\mathbf{y}$ is the ratio of the amount per bet wagered on the losing team to the amount per bet wagered on the winning team, b_2/b_1.

To estimate b_2/b_1, we make the assumption that the bookmaker takes no risk. With this assumption, market clearing conditions for Moneyline contracts on a game are

$$(1 + y_1)\, n_1 b_1 = n_1 b_1 + n_2 b_2 \tag{13}$$

$$(1 + y_2) n_2 b_2 = n_1 b_1 + n_2 b_2 \tag{14}$$

where n_j is the number of bets placed on team j and b_1 and b_2 are the *average* amount per bet wagered on teams 1 and 2. We can re-write these conditions to derive the expression

$$\frac{b_2}{b_1} = \frac{(1 + y_1)\, n_1}{(1 + y_2) n_2} \tag{15}$$

For nearly all games in our sample (32,655 out of 36,609 games), our data contain the proportion of bets placed on each team, and hence allow us to compute the ratio n_1/n_2, which allows us to compute b_2/b_1. We discuss the data on the proportion of bets placed on each team in Appendix A.4.

Previous work that estimates the preferences implied by the Favorite-Longshot Bias uses the assumption that bettors are indifferent between betting the same amount on each betting contract ($b_1 = b_2$).[26] This assumption is largely made because of lack of data on betting amounts. For our analysis, we depart from the assumption of equal bet sizes and study the preferences of a bettor that is indifferent between wagering the average amount wagered on the home and away teams.

To understand how average estimated bet sizes vary across games, for each game, we use Equation (15) to construct estimates of the "bet size ratio", which is the ratio of the average amount per bet on the home

[26]See Weitzman (1965); Ali (1977); Golec and Tamarkin (1998); Jullien and Salanié (2000); Snowberg and Wolfers (2010) and Chiappori et al. (2019). An exception is Bradley (2003), who notes the potential impact of the equal bet size assumption for estimating preferences using race data and performs back-of-the-envelope estimates of the preferences of a CPT bettor that optimizes his bet size.

team to the average amount per bet on the away team. Larger values of the bet size ratio correspond with larger bet sizes on the home team relative to the away team. For each game, we also compute the "odds ratio", which is the ratio of the odds of the home team winning to the odds of the away team winning (calculated by converting Moneylines to decimal odds). Smaller values of the odds ratio correspond with the home team being more favored in a game, while larger values corresponds with the home team being less favored. We sort games into twenty equally spaced bins based on their odds ratio, and plot a binned scatterplot of the average log odds ratio against the average log bet size ratio of each bin in Figure 6.

The figure reveals that the bet size ratio is nearly monotonically decreasing in the odds ratio. That is, when the home team is favored, the average amount wagered on the home team is larger than the amount wagered on the away team, and vice-versa when the home team is less favored. All else equal, the evidence suggests bettors wager more per bet when betting on a more favored team. This departs from the usual assumption of equal betting amounts across different betting contracts.

One limitation in our calculation of bet sizes is that it requires the assumption that bookmakers take no risk in betting markets. In Appendix A.4, we show that the same relationship between bet sizes and odds appears in each of the sports in our data set. Additionally, using betting data for a sample of soccer matches from 2006 to 2011 from Betfair, one of the world's largest betting exchanges, we verify that the same relationship between bet sizes and odds persists in a similar setting where we are able to measure the average wager size exactly without any assumptions about bookmaker behavior.[27] This evidence suggests that the assumption that bookmakers take no risk is not likely to have a major impact on our estimates of bet size ratios.

5.4 Estimation Results

We pool all games in our sample together and maximize the likelihood function in Equation (12). Naively maximizing the likelihood function, we obtain point estimates for $(\alpha, \gamma, \lambda)$ of (0.81, 0.79, 0.92). As Figure 7 shows, these parameters are able to do a reasonably good job of explaining the Favorite-Longshot Bias in our sample.[28]

However, the diminishing sensitivity parameter, α, and the probability weighting parameter, γ, are not identified from one another in our estimation.[29] To illustrate this point, we form a grid of values of α and γ ranging from 0.5 to 1, and maximize the likelihood function for each (α, γ) pair by allowing λ to vary. We plot the log-likelihood surface against α and γ in Figure 8. Points on the surface colored red cannot be rejected at the 5% confidence level by likelihood ratio tests versus the unrestricted estimates. The figure illustrates that for any given value of α on the surface, there are values of γ and λ that produce sample likelihoods that are statistically indistinguishable from the unrestricted estimates. In the red, "optimal region" of the

[27]We thank Angie Andrikogiannopoulou for graciously providing the Betfair data used for this comparison.

[28]The realized returns in each decile in Figure 7 differ slightly from those presented in Figure 1 because the former excludes games for which we do not have betting volume data. The games for which we do not have volume data tend to be college games where one team is very heavily favored relative to the other (e.g., a major conference team playing against a much lesser team).

[29]See Section 4 of Barseghyan et al. (2018) for more discussion. As they note, separately identifying the probability weighting and utility parameters is possible when including a third choice that individuals are indifferent between. However, this is just out of reach of our data given that we cannot compare betting amounts across games or contract types.

likelihood surface, we generally observe $\alpha - 0.03 < \gamma < \alpha$. In the optimal region, point estimates of λ (not shown in the figure) range from 0.90 to 0.97.

Because α and γ are not uniquely identified, we first proceed by fixing the value of α, and studying the corresponding maximum likelihood estimate of γ and λ. Tversky and Kahneman (1992) report an estimate of $\alpha = 0.88$ for the median participant in their experiment. In a meta-study, Booij et al. (2010) document that the majority of estimates of α from experimental studies range from 0.5 to 0.95. Barberis et al. (2019) choose a value of $\alpha = 0.7$ when studying the ability of CPT to explain a number of stock return anomalies. Hence, we consider $\alpha \in \{0.5, 0.7, 0.88, 1\}$, with the last value corresponding with a linear value function (studied by Snowberg and Wolfers (2010)).

The first five columns of Panel A in Table 3 present the maximum likelihood estimates of γ and λ and their standard errors corresponding with each value of α. Focusing first on estimated values of γ, the values are all statistically reliably less than one at the 5% confidence level, consistent with bettors overweighting low probability events and underweighting high probability events in their decision-making. Point estimates of γ are approximately 0.02 to 0.03 less than the assumed α (e.g., the estimated γ is 0.68 for $\alpha = 0.7$). Focusing on the estimated values of λ, the values range from 0.92 to 0.95, and are statistically significantly less than 1 in all cases, suggesting that bettors *underweight* potential losses in their betting decision; that is they are loss-tolerant, rather than loss-averse.

Next, we fix the value of γ and study the corresponding maximum likelihood estimates of α and λ. Tversky and Kahneman (1992) report a median value of 0.65 for γ among participants in their study. Barberis et al. (2019) use a value of $\gamma = 0.65$ to study CPT's explanatory power for stock pricing anomalies. Booij et al. (2010) report values from experimental studies that are around the same. Baele et al. (2016) estimate a value of $\gamma = 0.65$ when studying CPT's ability to explain options prices. In the betting literature, Snowberg and Wolfers (2010) report a probability weighting parameter of 0.93 using the Prelec (1998) probability weighting function with a risk-neutral value function.[30] This corresponds with moderate probability weighting, and translates to approximately $\gamma = 0.9$ in our setting. Accordingly, we focus on $\gamma \in \{0.65, 0.9, 1\}$, with $\gamma = 1$ corresponding with the case in which there is no probability weighting (bettors use objective probabilities).

The last four columns of Panel A in Table 3 present MLE estimates of α and λ corresponding with the chosen γ values. For $\gamma = 0.65$, the corresponding estimates of α and λ are 0.67 and 0.94. These values are 0.93 and 0.92 for $\gamma = 0.90$ and 1.03 and 0.92 for $\gamma = 1$. The numbers broadly mirror the story told in Panel A. One interesting point is that, under objective probabilities, α is estimated to be greater than one. This corresponds with bettors being risk-loving over gains and risk-averse over losses, which is inconsistent with diminishing sensitivity.

In Appendix Table A.3, we report CPT parameter estimates using average bet sizes separately for each sport in our sample. Consistent with the full-sample results, we find point estimates of $\lambda < 1$ for each sport in our sample. Across sports, we obtain similar estimates of α and γ as those in the full sample, though in the NBA and NFL sub-samples, we find $\alpha < \gamma$.

[30] This specification is $w(p) = \exp\left\{-\{-\ln p\}^{\beta}\right\}$, where β is the probability weighting parameter. Prelec (1998) derives this functional form from a set of axioms for probability weighting functions.

5.5 Comparison of CPT Estimates with Evidence From the Racetrack

Previous work characterizing the preferences of bettors using aggregate data makes the assumption that bettors bet the same dollar amount on each available betting contract, and only focuses on one or two features of CPT. Snowberg and Wolfers (2010) focus on probability weighting and use a risk-neutral value function. Jullien and Salanié (2000) adopt the following specification of Cumulative Prospect Theory:

$$V = G^{-}(1-p)U(W-1;\theta) + G^{+}(p)U(W+x;\theta) \tag{16}$$

where p is the probability the bet pays off, $G^{+}(\cdot)$ and $G^{-}(\cdot)$ are separate probability weighting functions for gains and losses, W is the bettor's wealth, and $U(y;\theta) = (1-e^{-\theta y})/\theta$ i.e., $U(\cdot)$ is a CARA value function and θ is the bettor's risk aversion.

Among the specifications they test, Jullien and Salanié (2000) find the best fit for the data is produced where $G^{+}(p) = p^{\alpha}$ is a slightly convex power function ($\alpha \approx 1.2$) and $G^{-}(p) = p^{\beta}$ is a very concave power function ($\beta \approx 0.3$).[31] They interpret this as evidence for Cumulative Prospect Theory, since it suggests that bettors treat gains and losses differently. The strong concavity of the probability weighting function over losses suggests that bettors overweight probabilities of losses in their decisions, and especially for small loss probabilities.

Our results are similar to the previous calibrations, in the sense that they suggest a role for probability weighting and suggest that bettors may treat gains and losses differently. Focusing on the case of a risk-neutral value function ($\alpha = 1$), our results suggest economically small, but statistically significant probability weighting, as in Snowberg and Wolfers (2010). Taking this one step further, our estimates show that introducing curvature into the value function, the degree of probability weighting implied by the data is even greater. Similarly, our results also suggest that bettors may treat gains and losses differently ($\gamma \neq 1$), a feature shared by the preferred specification in Jullien and Salanié (2000).

However, a notable difference between our work and prior work is that we assume that bettors are indifferent between wagering the *average* amount bet on the two teams playing in a game, rather being indifferent between wagering the same amount. Relaxing the assumption of equal betting amounts across contracts yields substantial differences in the estimated preference parameters and has particularly different implications for how bettors treat potential losses in their decision making. To illustrate, in Panel B of Table 3, we re-estimate the CPT preference parameters as before, but enforce the assumption that the bettor is indifferent between wagering the same dollar amount on the two teams playing in each game. The panel reveals two differences from our main specification. First, λ, the loss-aversion parameter is estimated to be greater than one in all cases, with point-estimates ranging from 1.5 to 3.2. This result indicates the presence of loss-aversion. Second, while diminishing sensitivity (α) and probability weighting (γ) still are not separately identified, the wedge between the estimated value of γ and α is much higher ($\alpha - \gamma \approx 0.3$).

Assuming equal betting amounts, the results suggest that bettors overweight losses in their decision-making, since $\lambda > 1$. While the assumed functional form of preferences is not the same as in Jullien and

[31] Gandhi and Serrano-Padial (2015) replicate these results in a larger sample

Salanié (2000), the implication is similar to the concave probability weighting function they estimate over losses. With data on betting amounts, our main specification yields the opposite result ($\lambda < 1$), suggesting mild loss-tolerance. The reason for the difference comes from the fact that, on average, bettors tend to bet a substantially larger dollar amount on favorites than on underdogs. Quantitatively, loss-aversion and diminishing sensitivity predict that bettors should bet a smaller dollar amount on favorites than we find in the data. If bettors are homogeneous, then the assumption that bettors wager the average amount bet on contracts more closely reflects their actual choices and therefore preferences.

5.6 Comparison with Financial Markets Facts

Building on previous work, and incorporating bet sizes into our calibration, our results on CPT preferences are comparable with results obtained in other settings, particularly in financial markets.

Focusing on probability weighting, we find evidence of diminishing sensitivity when using average betting amounts. Studies have found 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 et al. (2019)). With a value of $\gamma = 0.65$, we estimate a diminishing sensitivity parameter of $\alpha = 0.67$ and a loss-aversion parameter of $\lambda = 0.94$. The estimate of α lies approximately in the middle of experimental estimates of diminishing sensitivity (see Booij et al. (2010)), and is very close to the parameter value of 0.7 used by Barberis et al. (2019) to explain stock market anomalies, including relative underperformance of idiosyncratically risky stocks. More broadly, assuming reasonable values of either probability weighting or diminishing sensitivity, we are able to obtain reasonable estimates for the other parameters.

We find evidence of slight loss-tolerance, however, consistently obtaining point estimates of $\lambda < 1$ when using average betting amounts.[32] A body of work suggests that loss-aversion may be able to explain a number of financial markets facts.[33] The importance of loss-aversion for the particular facts we are concerned with here, the relative performance of riskier assets in financial markets, is less clear. For example, in explaining the cross-section of options returns, Baele et al. (2016) find that loss-aversion has little explanatory power in their context, and that probability weighting is the quantitatively more important component of prospect theory for explaining the facts. On the experimental side, many experimental studies find evidence of loss-aversion (as summarized by Booij et al. (2010)). However, in recent work, Chapman et al. (2018) conduct a study of participants that are more representative of the US sample than normally studied in experiments, and obtain an estimate of $\lambda = 0.98$ for their median participant. They find that around 50% of the US population may be loss tolerant, but do find greater evidence of loss-aversion for those with higher cognitive abilities. It may be that loss-aversion plays a role in traditional financial markets, but the average bettor

[32] Our finding of mild loss-tolerance should be qualified, in the sense that we make parametric assumptions about the form of preferences. Relaxing the parametric form assumed of probability weighting, for example, can lead to different inferences about loss-aversion as well, as Barseghyan et al. (2018) discuss. Similarly, while the status quo is a common choice of reference point, a different choice of reference point may also result in different conclusions about loss aversion.

[33] See e.g., Benartzi and Thaler (1995) and Barberis et al. (2001) in the context of the equity risk premium, Barberis et al. (2006) in the context of stock market participation, and Barberis and Huang (2001) and Barberis et al. (2019) in the context of asset pricing anomalies.

that selects into sports betting is more loss-tolerant than the average financial market participant.

On the optimistic side for Cumulative Prospect Theory as a unifying explanation of the facts, our estimates suggest that prospect theory, and especially probability weighting and diminishing sensitivity, can reconcile the Favorite-Longshot Bias with a similar parmetrization shown to explain facts in financial markets. The evidence on loss-aversion (or loss-tolerance), however, is more mixed. On the more cautious side, our estimation does not allow us to separately identify the role that diminishing sensitivity and probability weighting play. While our evidence is consistent with the parameters found in financial markets, different parameterizations may also be consistent with the Favorite-Longshot Bias.

6 Heterogeneous Beliefs

In the previous section, we show that the empirical patterns we document are consistent with homogeneous bettors with Cumulative Prospect Theory preferences. Here we focus on the ability of a model with risk-neutral bettors with heterogeneous beliefs to explain the Favorite-Longshot Bias, similar in spirit to Gandhi and Serrano-Padial (2015). Previous studies link belief heterogeneity with the Favorite-Longshot Bias from a theoretical perspective, e.g., Ali (1977), Shin (1991, 1992), and Ottaviani and Sørensen (2009, 2010). This research places particular focus on the role that private information plays in the FLB. Gandhi and Serrano-Padial (2015) take an empirical approach to estimate the degree of belief heterogeneity by noting that in a discrete choice environment, the choice of which horse to bet can be framed as isomorphic to a model of horizontally differentiated demand. The logic behind the heterogeneous beliefs explanation for the Favorite-Longshot bias 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 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.

Whereas Gandhi and Serrano-Padial (2015) directly model belief heterogeneity about expected returns across contracts, we model bettors as having heterogeneous beliefs about a more fundamental quantity - the expected point difference of a game. As a preview of the results from the empirical estimation, we find that assuming that each bettor wagers the same amount, belief heterogeneity is able to empirically rationalize the Favorite-Longshot Bias, consistent with the findings in Gandhi and Serrano-Padial (2015). Because we have betting volume data, however, we are able to relax the assumption of equal betting amounts and use data on the actual proportion of bets placed on each contract. Doing so, we find that the proportion of bettors choosing to bet on extreme underdogs is actually higher than the proportion of bettors choosing to bet on more moderate underdogs, with 44% of bettors choosing to bet on the underdog in the most extreme games. This evidence is difficult to reconcile with heterogeneous beliefs alone, and the magnitude of betting volume is difficult to explain in this framework, where 44% of betting volume is on the 10% most extreme underdogs. Hence, while heterogeneous beliefs may contribute to the Favorite-Longshot Bias, other factors, likely preferences for risk or skewness, seem to play an important role.

6.1 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 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.

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).[34] As before, betting on team 1 offers a potential payoff of y_1 per dollar wagered and betting on team 2 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.[35]

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 $h(b)$.

[34] 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.

[35] 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.

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

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

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 (13) and (14) in terms of market shares (below). We then directly solve for market shares for each game by equating the market clearing conditions and since $s_1 + s_2 = 1$. Both measures of market shares have different interpretations for our results, which we discuss in more depth after presenting the model setup.

$$s_1(1 + y_1) = s_1 + s_2 \tag{17}$$

$$s_2(1 + y_2) = s_1 + s_2 \tag{18}$$

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

$$\text{likelihood}(j; b) = \begin{cases} \int_{-\infty}^{\bar{b}} h(b) & \text{if } j = 1 \\ \int_{\bar{b}}^{\infty} h(b) & \text{if } j = 2 \end{cases} \tag{19}$$

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

$$\mathcal{L} = \prod_{k=1}^{K} \prod_{j_k \in J_k} \text{likelihood}(j_k; b_k)^{s_k^j} \tag{20}$$

and the corresponding log-likelihood of the sample as

$$\log \mathcal{L} = \sum_{k=1}^{K} \sum_{j_k \in J_k} s_k^j \log(\text{likelihood}(j_k; b_k)) \tag{21}$$

6.2 Empirical Implementation

To estimate belief heterogeneity in the data, we need to express the likelihood function in Equation (19) in terms of parameters that we would like to estimate, and observables. Because bettors are risk-neutral, $\bar{b}$'s indifference condition is expressed as

$$\bar{p}y_1 - (1 - \bar{p}) = (1 - \bar{p})y_2 - \bar{p} \tag{22}$$

Re-arranging terms yields

$$\bar{p} = \frac{y_2 + 1}{y_1 + y_2 + 2} \tag{23}$$

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

$$\bar{b} = F^{-1}(\bar{p}) - F^{-1}(p) \tag{24}$$

where F^{-1} is the inverse CDF of the standard normal distribution, p is the objective probability that team 1 wins the game. As in Section 3.2 (and discussed in Appendix A.2), we use the nonparametrically estimated win probabilities for the objective win probabilities.

We assume $h(b)$ is logistically distributed with mean zero, following Gandhi and Serrano-Padial (2015).[36] The parametric assumption is used primarily for simplicity, and our interpretation of the results does not vary with the parametric assumption. Given the parametric assumptions imposed, the scale parameter, which controls the dispersion of the belief distribution, is the only parameter to be estimated.

6.3 Identifying Variation

For each game, denote the observed payoffs and market shares for game i as $\left(y_1^i, y_2^i, s_1^i, s_2^i\right)$. Given the assumptions of the model, $H\left(\bar{b}^i\right) = s_2^i$, where $H(x) = \int_{-\infty}^{x} h(b)$ is the CDF of the belief distribution and $\bar{b}^i$ is the marginal bettor that is indifferent between betting on the two contracts offered on the game i. From each game, we observe variation in market shares, and implicitly as a function of observed contract returns, variation in the marginal bettor in each game, $\bar{b}^i$. Our estimation uses the variation in market shares and $\bar{b}^i$ to trace out the shape of $h(b)$.

Variation in market shares and the marginal bettor itself emerges from variation in the expected outcome across games, and the differences in expected returns across contracts based on the expected outcome. The observed difference in returns across contracts based on the expected outcomes suggests that the marginal bettor must have more extreme beliefs when the expected outcome is more extreme. Hence, the observed Favorite-Longshot Bias is crucial in order to estimate the implied belief heterogeneity.

6.4 Understanding Market Share Measures

We use 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 likely not the case 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 (17) and (18). However, the relationship between the proportion of bets placed on a contract and the probability of

[36] While they use a logistic mixture distribution, here we simply use a logistic distribution for simplicity.

the contract paying off is an empirical question, since in practice, bettors may wager different amounts when betting on the favorite and the underdog.

We compare the two measures of market shares in order to understand how they may differ. We sort contracts into twenty bins based on their estimated dollar market shares. Figure 9 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 reverse is true in the tails (for extreme favorites and extreme underdogs.) The relationship between the proportion of bets and the dollar proportion of bets on contracts 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) are approximately 90% and 10%, the average proportion of bets placed on contracts in these bins are 56% and 44%. This is especially notable given that the average returns for contracts in these bins are -0.40% and -24%, meaning that almost half of all bets are placed on the contract with considerably more extreme negative returns, though only about 10% of the dollar volume is placed on these contracts. While theoretically, the Favorite-Longshot Bias may emerge with only a small proportion of bettors willing to bet on extreme underdogs (as noted in Gandhi and Serrano-Padial (2015)), our empirical evidence suggests that many bettors bet on extreme underdogs.

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 relationship between the proportion of bets placed and the dollar proportion of bets placed on a contract foreshadow the results of our calibration. 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 underperformance of extreme underdogs when the marginal bettor betting on extreme underdogs has more extremely distorted beliefs. Belief heterogeneity cannot explain the *increase* in the proportion of bettors for extreme underdogs that we observe in the data.

6.5 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 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 dollar market shares, we estimate the scale parameter as 0.302 (standard error of 0.014) - the corresponding estimated standard deviation of b is $\sigma_b = 0.547$.

With these estimates, we construct the implied expected returns for each contract using the calibrated belief distribution and observed market shares. Using Equation (24), 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 (23). 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.

To assess the ability of each model to explain the Favorite-Longshot bias, 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 10. The first plot in the figure corresponds with the model implied returns using the estimated dollar weighted belief distribution. The figure 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 the Favorite-Longshot Bias observed in the data.

The second plot in the figure corresponds with expected returns implied by the unweighted belief distribution. 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. Particularly, 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.

6.6 Interpreting the Results

The results suggest that using the belief distribution implied by dollar market shares is able to do a better job at capturing the Favorite-Longshot Bias than the belief distribution using market shares based on the proportion of bets placed. This result stems from the relationship between the dollar proportion of bets and the proportion of bets observed in Figure 9. Conceptually, the heterogeneous beliefs based explanation of the Favorite-Longshot Bias operates by suggesting that bettors with more extreme beliefs bet on underdogs. This explanation suggests that the more extreme the outcome, the fewer bettors with more extreme beliefs. While this holds true when weighting bettors by the number of dollars they wager, Figure 9 suggests that nearly half of bettors bet on the underdog in more extreme games. Regardless of parametric assumptions on the belief distribution, belief heterogeneity alone cannot reconcile the increasing number of bettors choosing to bet on extreme underdogs. Additionally, belief heterogeneity does not provide a reason why bettors that bet on underdogs may bet a smaller dollar amount than bettors wagering on favored teams, as we document in the previous section. Both of these sets of results suggest that preferences likely play a role in the decision of bettors to bet on extreme underdogs.

However, the results also suggest a potential role for belief heterogeneity in the Favorite-Longshot Bias. 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. (2019) 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%. Compared with these numbers, the belief heterogeneity we estimate in the sports betting data is of a reasonable magnitude. 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.

A caveat to our interpretation of the calibration is that we do not model the choice of which game to bet on, but rather model the distribution of beliefs as the same across games conditional on bettors' choices to bet on those games. We do not have the data to make comparisons in proportion of Moneyline bets placed across games that would be required to model such a choice. It is possible that bettors with the most extreme beliefs select into the games with the largest underdogs. However, this selection would not, by itself, explain why such bettors choose to wager small dollar amounts. We would also expect that bettors with more "correct" beliefs would choose to wager on the favorite in such games, because extreme favorites offer the most attractive returns in the sample. Even when considering the possibility that bettors select into different games, preferences likely play a role.

While heterogeneous beliefs may play a role in financial markets,[37] our study of belief heterogeneity in the context of the Favorite-Longshot Bias is most closely related to the fact that on the same underlying, out-of-the-money options earn lower returns than in-the-money options. Similar to the Favorite-Longshot Bias, such a pattern can theoretically emerge under certain conditions if agents with heterogeneous beliefs about an asset's returns sort into options based on their optimism (Shefrin (2008)).[38] Our results add additional evidence toward the idea that non expected-utility preferences likely play an important role in option prices (Kliger and Levy (2009), Boyer and Vorkink (2014), and Baele et al. (2016)).

We do not focus here on disentangling the role of belief heterogeneity and preferences. However, our results do suggest that there may be a role for both. Understanding the interplay between preference and belief-based explanations in betting markets and financial markets is an interesting area for further research. However, our results highlight that a preference for lottery-like payoffs, as suggested by Cumulative Prospect Theory, likely play an important role in explaining the facts.

[37] For example, there is a substantial literature documenting that belief heterogeneity in the presence of short-sale constraints can contribute to trading volume and can lead to assets being overvalued relative to fundamentals (Miller (1977), Harrison and Kreps (1978), Scheinkman and Xiong (2003), and Diether et al. (2002)).

[38] The pricing results in such a model are not consistent with the dominant, no-arbitrage paradigm in option pricing. However, there is evidence against no-arbitrage in options markets. For example, Garleanu et al. (2008) show theoretically and empirically that customer demand pressures, combined with risk-averse market makers that cannot perfectly hedge their inventory, may lead to no-arbitrage violations, such as the low returns of out-of-the-money options.

7 Conclusion

In this paper, we seek to bridge two well-known, empirical phenomena in financial markets and betting markets. In financial markets, risky assets earn lower risk-adjusted returns, and in betting markets, the Favorite-Longshot bias is a similar phenomenon, where riskier bets on underdogs earn lower returns. Moreover, this pattern in betting returns disappears when comparing equal-risk contracts across the same games, where cross-sectional differences in risk are removed, thus highlighting that risk, and not any other characteristic, is the chief attribute driving these patterns. Quantitatively comparing the returns in betting markets with the financial markets facts using an implied volatility surface, the magnitude of the risk-return relationship in betting contracts and the magnitude of the risk-return relationship in index and single name equity options markets are similar.

Accumulating the results informs theories that are able to simultaneously explain both sets of facts. For example, rational theories which predict no relationship between idiosyncratic risk and returns, cannot explain these facts. Our evidence points towards Cumulative Prospect Theory as a a unifying explanation of the facts, and perhaps also some role for belief heterogeneity, though heterogeneous beliefs alone cannot explain all of the facts.

8 Tables and Figures

Figure 1: Risk and Return in the Cross-Section of Assets

The figure presents returns versus measures of risk in the cross-section of assets in both sports betting markets and financial markets. The first panel plots the average 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. The bars correspond with 95% confidence intervals for the average reutrn within a particular decile. The second panel plots the CAPM Alpha (the monthly risk-adjusted return, in basis points) of US equities, sorted into deciles based on their CAPM Beta. Decile 1 corresponds with contracts that have the highest CAPM Beta (the riskiest stocks) and Decile 10 corresponds with the the lowest CAPM Beta (the least risky stocks). The numbers used in the plot are taken from Table III in Frazzini and Pedersen (2014) and the sample runs from January 1926 to March 2012.

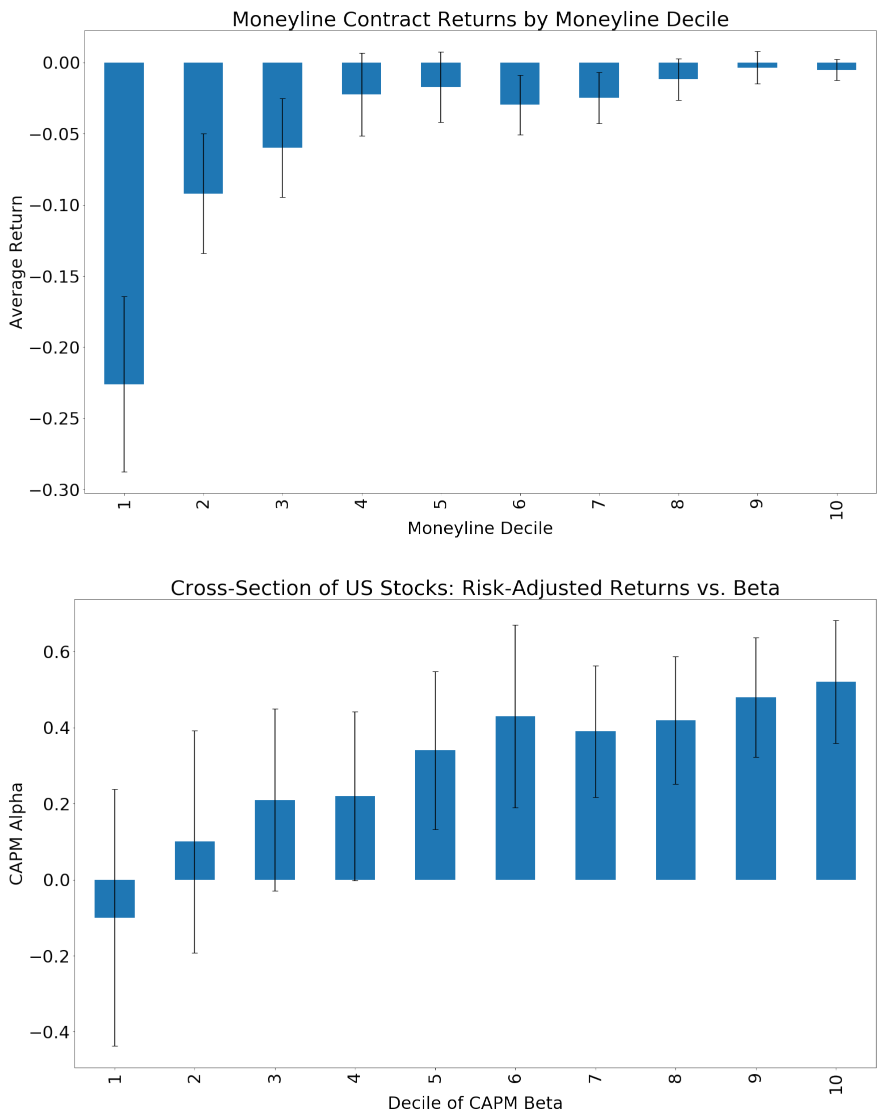
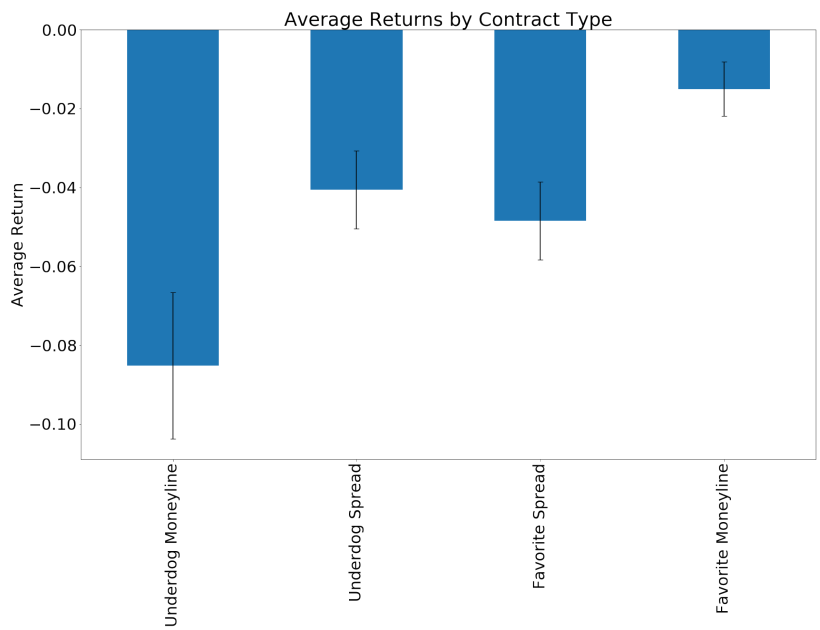


Figure 2: Risk and Return in the Cross-Section of Contracts

The figure presents the average return of contracts on the same asset, sorted based on characteristics related to their risk. The first panel plots the average returns of Spread and Moneyline contracts on the underdog and favorite of each game, with one column corresponding with each contract type. The bars correspond with 95% confidence intervals. The second panel plots delta hedged excess returns, in percent terms, of one month maturity options sorted into one of five portfolios based on the absolute value of their Black-Scholes Delta values: Deep out of the Money (Absolute Delta of 0-20), Out of the money (Absolute Delta of 20-40), At the Money (Absolute Delta of 40-60), In the Money (Absolute Delta of 60-80), and Deep in the Money (80-100). Equity options correspond with all options on equities in the OptionMetrics Ivy DB, subject to a set of liquidity filters. Index options correspond with options on a set of 13 indices, also subject to liquidity filters. Each portfolio is rebalanced on the first trading day following expiration Saturday, and all options are value weighted within a given portfolio based on the value of their open interest. The data in the plot are taken from Table III of Frazzini and Pedersen (2020).



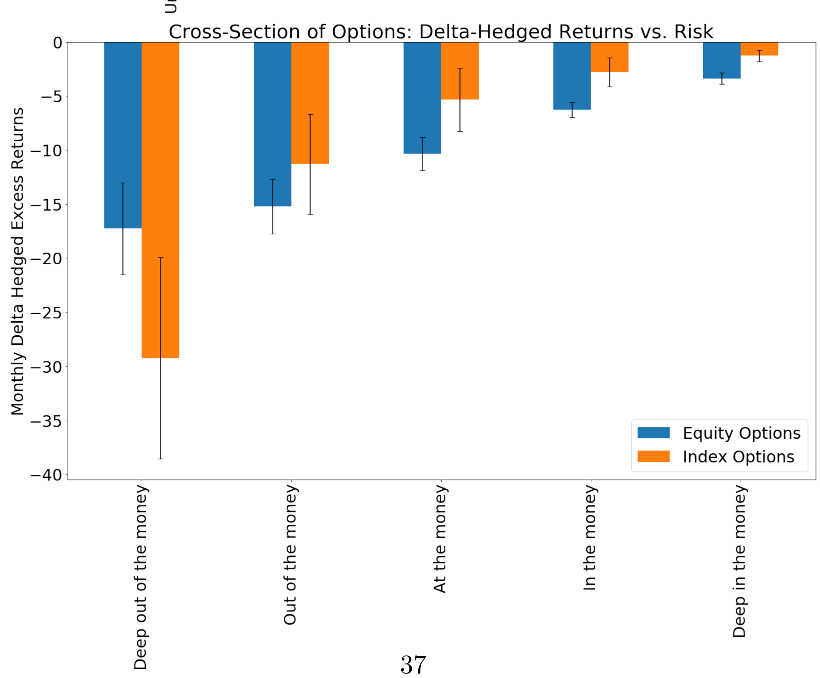


Figure 3: Average Spread Contract Return Across Games

The figure plots the average returns of Spread 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 bars correspond with two standard errors above and below the average return for a particular contract type.

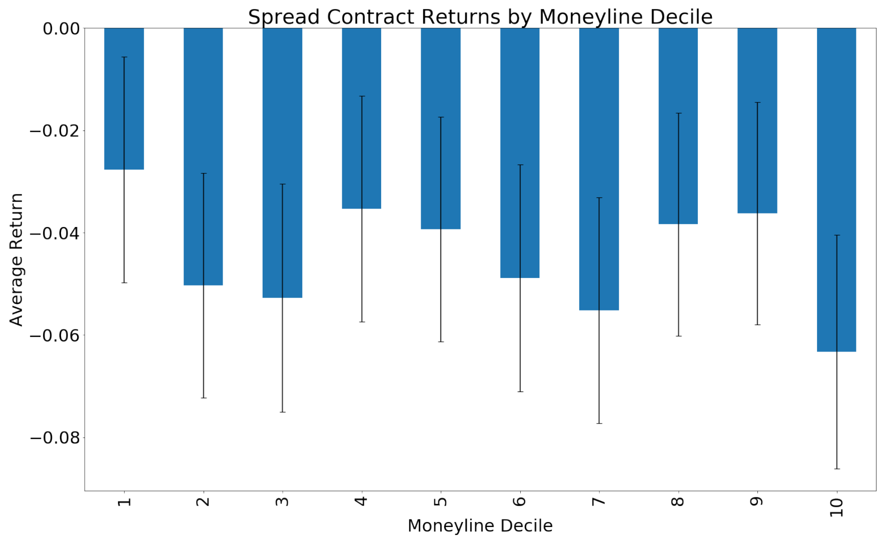


Figure 4: Implied Volatility Surface by Sport

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 (4), 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.

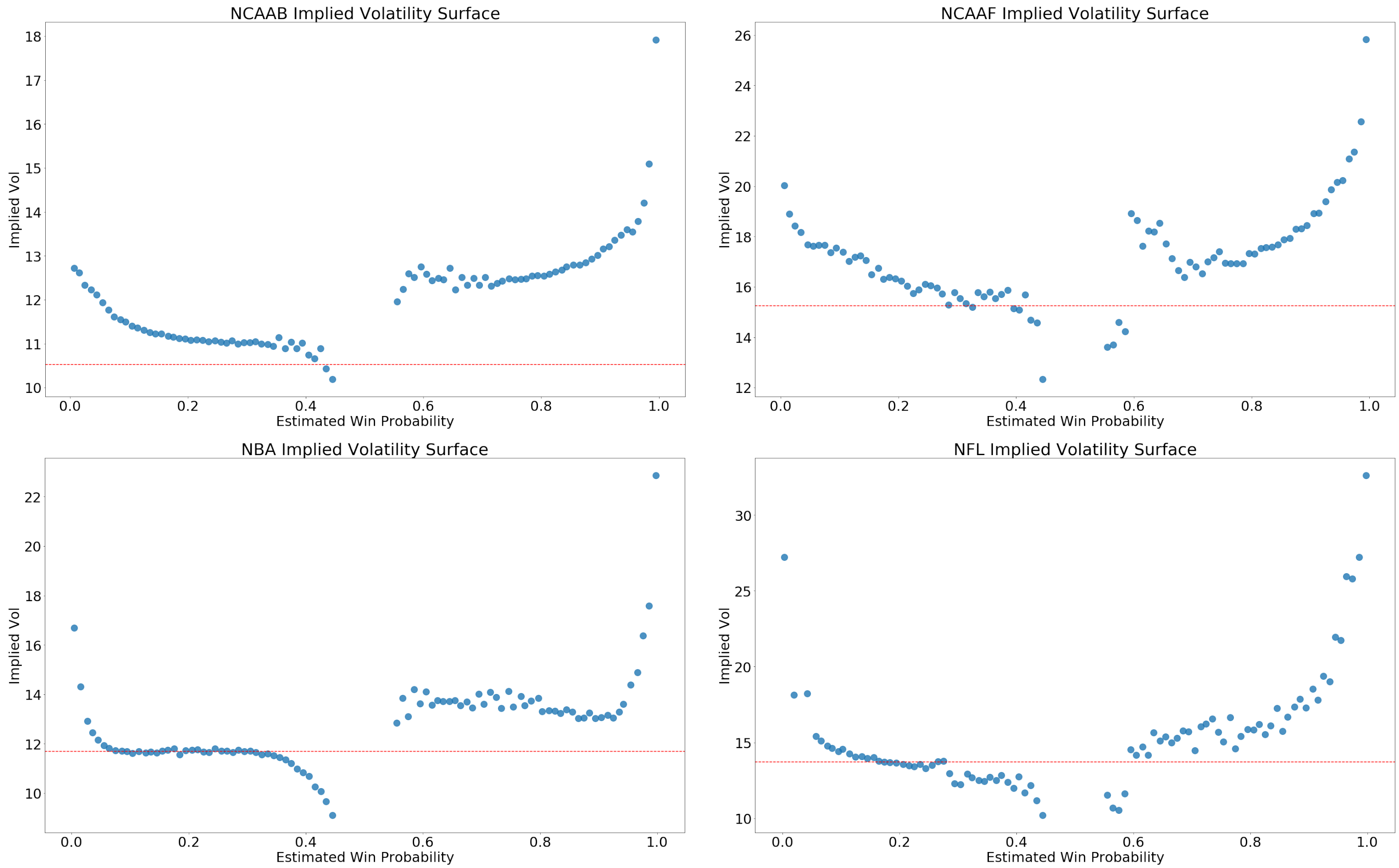


Figure 5: Accuracy of Over/Under Contracts

The first panel in the figure is a binned scatter plot of the Over/Under line versus the total number of point 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 bars correspond with plus/minus two standard deviations relative to the mean.

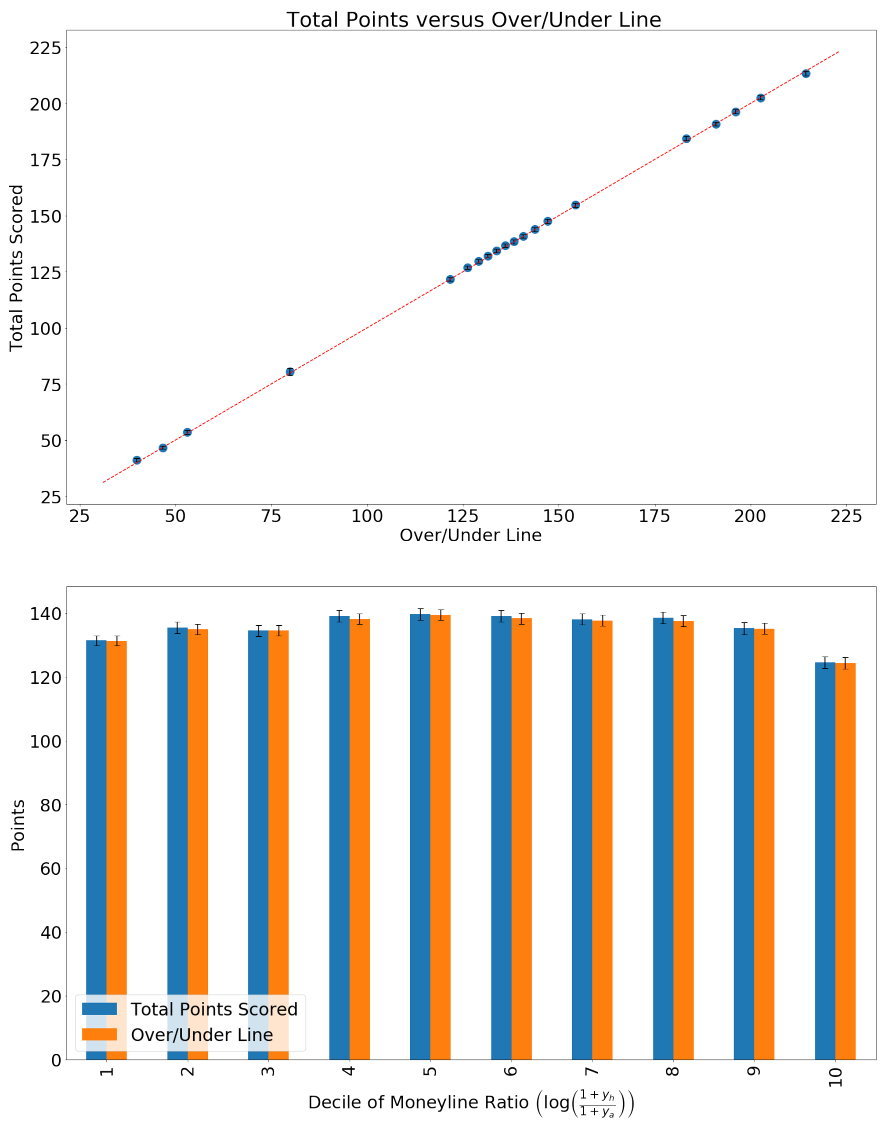


Figure 6: Odds and Average Bet Sizes

The figure plots a scatter plot of the log ratio of 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") versus the log ratio of the odds on the home team to the odds on the away team (the "odds ratio"). Each game is sorted into one of twenty equally sized bins based on the odds ratio of the game. Each point corresponds with the average log bet size ratio and average log odds ratio within a bin. The bars correspond with plus or minus two standard errors above and below the mean.

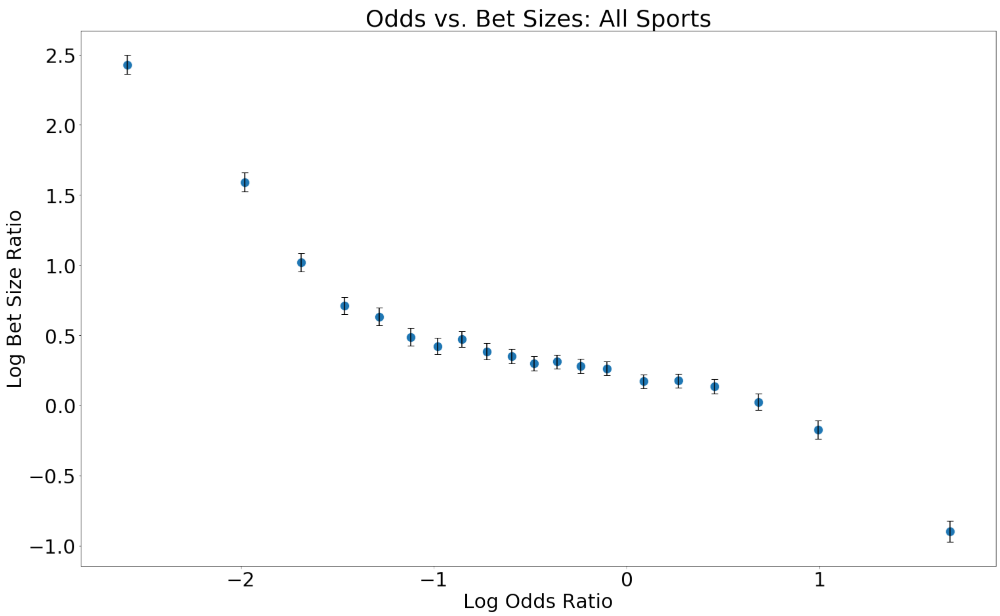


Figure 7: Preference Implied Returns and Realized Returns

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 estimated by using win probabilities calculated using Equation (7), which is derived by assuming a Cumulative Prospect Theory bettor that is indifferent between betting the average amount wagered on the two teams playing against each other in a game. The parameter values used are $(\alpha, \gamma, \lambda) = (0.81, 0.79, 0.93)$, which maximize the likelihood of our sample.

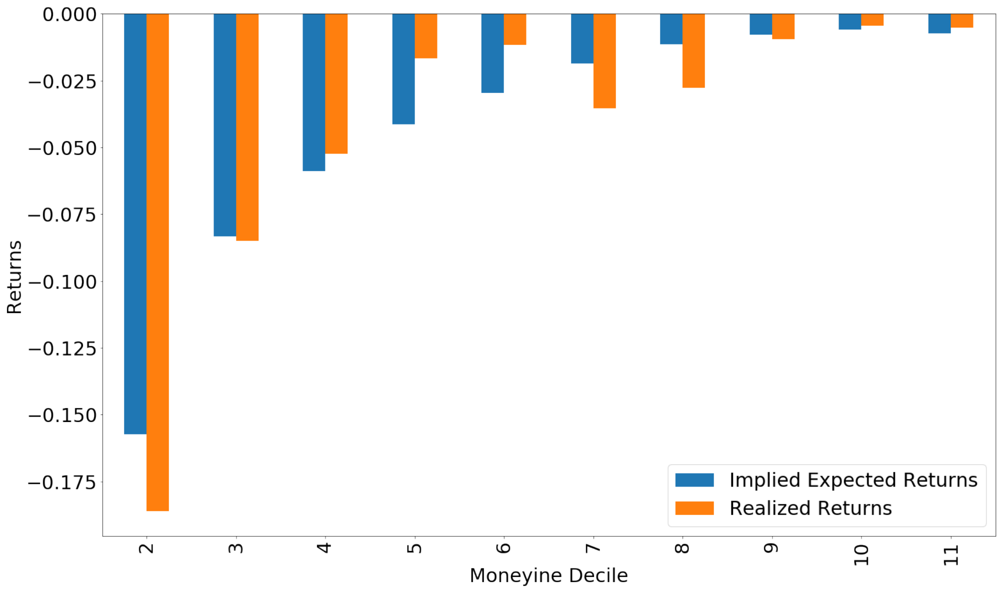


Figure 8: Log-Likelihood Surface for CPT Preference Parameters

For each α, γ pair on the grid, the figure plots the log likelihood found by maximizing the likelihood function in Equation (12) with respect to λ and restricting α and γ to be the given values. Points on the surface colored red cannot be rejected at the 5% confidence level by a likelihood ratio test versus the alternative parameters obtained by maximizing the likelihood function without restrictions.

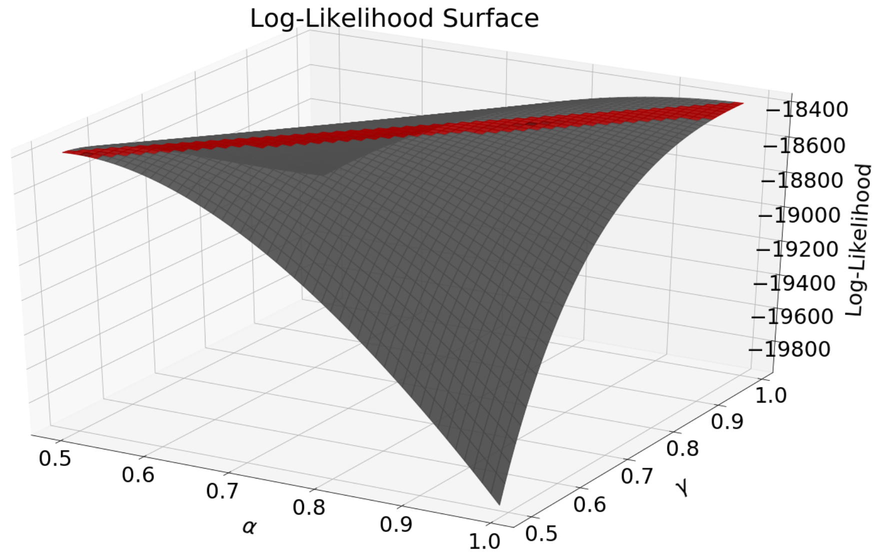


Figure 9: Comparison of Market Share Measures

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 (17) and (18) 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.

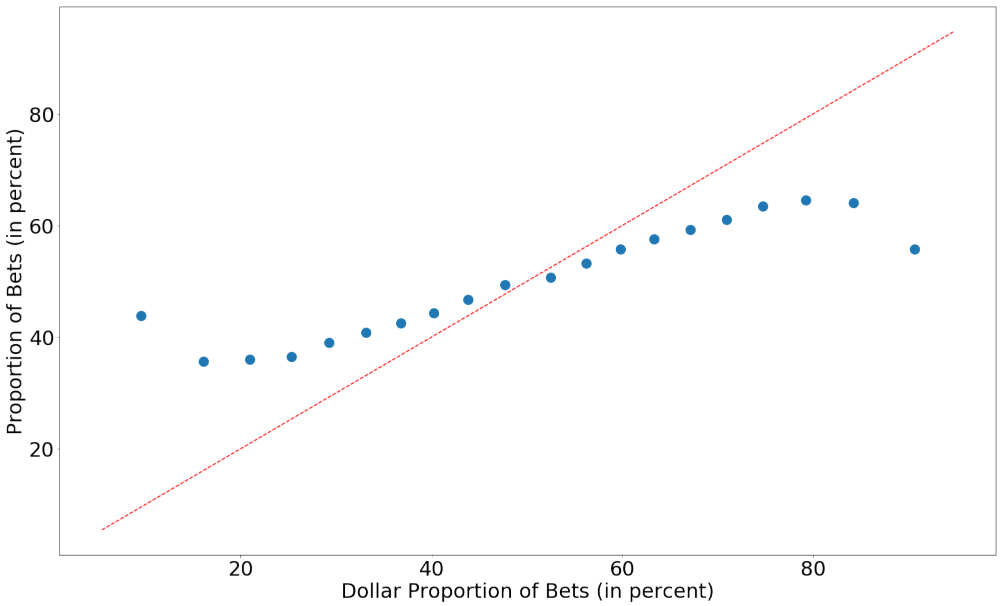
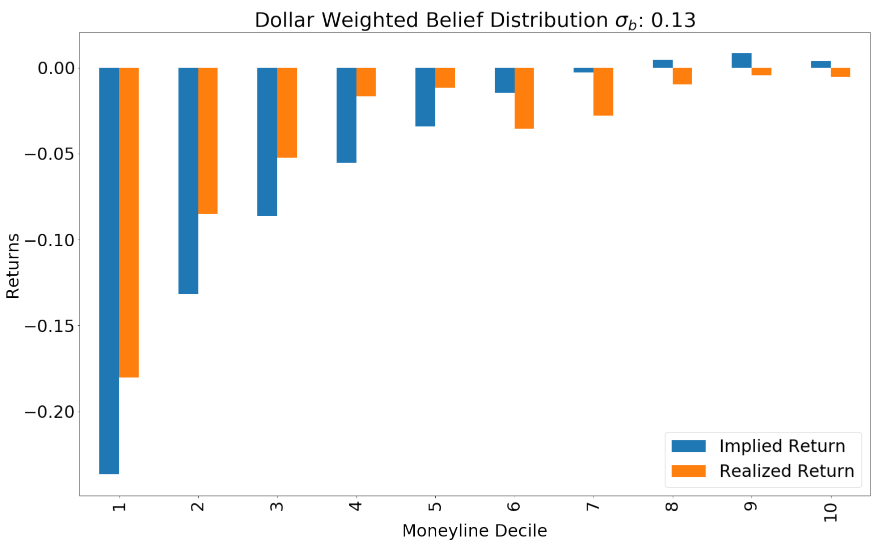


Figure 10: Heterogeneous Beliefs and Expected Returns

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.



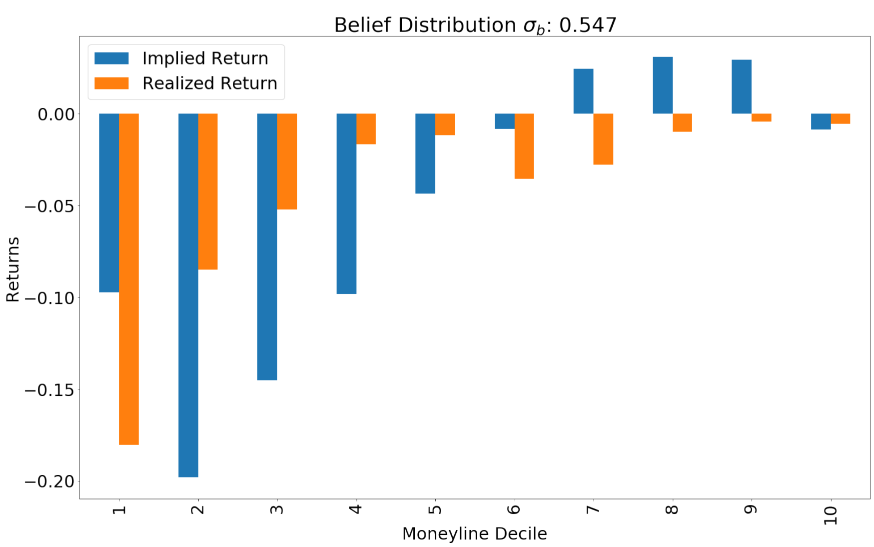


Table 1: Implied Volatility Premium: Sports Betting Contracts and Options

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.2. The last two rows of the panel correspond with the Implied Volatility premia, as the percent different between the average implied volatilties 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	11.5	17.4	11.7	14.3	-
IV for 85%-95% WP Contract	13.1	18.8	13.1	17.9	-
IV Premium for 5%-15% WP	9.2%	13.8%	0.0%	4.0%	6.8%
IV Premium for 85%-95% WP	24.9%	23.2%	12.4%	30.7%	22.8%

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 2: Spread lines, Moneylines, and Predicting Wins

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 Spread line of the game and $\log\left(\frac{1+y_h}{1+y_a}\right)$, where y_h and y_a are the payoffs associated with winning Moneyline bets on the home and away teams. Independent variables are standardized to have zero mean and unit standard deviation within each sport, and separate regression coefficients are estimated by sport. t-statistics are reported in parentheses.

	(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	36609	36609	36609
R^2	20.66%	20.85%	20.87%
F-statistic	1361.37	1377.75	877.72

Table 3: CPT Parameter Estimates

The table displays Maximum Likelihood Estimates for Cumulative Prospect Theory preference parameters found by maximizing Equation (12). Panel A presents MLE estimates of parameters assuming that bettors are indifferent between betting the average amount wagered on the home and away team in each game. Present B presents MLE estimates of parameters assuming bettors are indifferent between betting the same dollar amount on the home and away team in each game. In each panel, the first five columns present parameters estimates of γ, the probability weighting parameter, and λ, the loss-aversion parameter, for a given α, the diminishing sensitivity parameter. The last four columns of each panel present MLE estimates of α and λ for a given value of γ. Standard erorrs are reported in parentheses.

Panel A: CPT Parameter Estimates using Average Bet Sizes

	Fixed α Estimates					Fixed γ Estimates		
Fixed α	0.50	0.70	0.88	1.00	Fixed γ	0.65	0.90	1.00
γ	0.485	0.679	0.854	0.970	α	0.670	0.928	1.03
	(0.007)	(0.010)	(0.012)	(0.014)		(0.009)	(0.013)	(0.015)
λ	0.952	0.938	0.925	0.917	λ	0.940	0.922	0.915
	(0.023)	(0.024)	(0.025)	(0.026)		(0.023)	(0.025)	(0.026)

Panel B: CPT Parameter Estimates Assuming Equal Bet Sizes

	Fixed α Estimates					Fixed γ Estimates		
Fixed α	0.50	0.70	0.88	1.00	Fixed γ	0.65	0.90	1.00
γ	0.234	0.430	0.603	0.717	α	0.924	1.189	1.296
	(0.085)	(0.079)	(0.077)	(0.076)		(0.081)	(0.081)	(0.081)
λ	3.162	2.173	1.843	1.715	λ	1.768	1.571	1.521
	(1.573)	(0.637)	(0.411)	(0.366)		(0.292)	(0.225)	(0.209)

References

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

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

_ and _ , "Individual reaction to past performance sequences: evidence from a real marketplace," *Management Science*, 2018, *64* (4), 1957–1973.

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

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

_ , _ , _ , and _ , "High idiosyncratic volatility and low returns: International and further US evidence," *Journal of Financial Economics*, 2009, *91* (1), 1–23.

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

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

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

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

_ and Ming Huang, "Mental accounting, loss aversion, and individual stock returns," *Journal of Finance*, 2001, *56* (4), 1247–1292.

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

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

_ , Lawrence J Jin, and Baolian Wang, "Prospect Theory and Stock Market Anomalies," *Available at SSRN 3477463*, 2019.

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

_ , _ , and Tano Santos, "Prospect theory and asset prices," *Quarterly journal of Economics*, 2001, *116* (1), 1–53.

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

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

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

_ , Michael C Jensen, Myron Scholes et al., "The capital asset pricing model: Some empirical tests," *Studies in the theory of capital markets*, 1972, *81* (3), 79–121.

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

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

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

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

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

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

Camerer, Colin F, "Does the Basketball Market Believe in theHot Hand,'?," *American Economic Review*, 1989, *79* (5), 1257–1261.

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

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

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

Diether, Karl B, Christopher J Malloy, and Anna Scherbina, "Differences of opinion and the cross section of stock returns," *Journal of Finance*, 2002, *57* (5), 2113–2141.

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

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

Egan, Mark L, Alexander MacKay, and Hanbin Yang, "Recovering Investor Expectations from Demand for Index Funds," Technical Report, National Bureau of Economic Research 2020.

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

Frazzini, Andrea and Lasse H Pedersen, "Embedded leverage," Technical Report 2020.

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

Friedman, Milton and Leonard J Savage, "The utility analysis of choices involving risk," *Journal of Political Economy*, 1948, *56* (4), 279–304.

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

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

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

Garleanu, Nicolae, Lasse Heje Pedersen, and Allen M Poteshman, "Demand-based option pricing," *The Review of Financial Studies*, 2008, *22* (10), 4259–4299.

Geanakoplos, John, "The leverage cycle," *NBER Macroeconomics Annual*, 2010, *24* (1), 1–66.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, "Five facts about beliefs and portfolios," Technical Report, National Bureau of Economic Research 2019.

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

_ **and** _ , "Bettors love skewness, not risk, at the horse track," *Journal of Political economy*, 1998, *106* (1), 205–225.

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

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

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

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

Harrison, J Michael and David M Kreps, "Speculative investor behavior in a stock market with heterogeneous expectations," *Quarterly Journal of Economics*, 1978, *92* (2), 323–336.

Hausch, Donald B, William T Ziemba, and Mark Rubinstein, "Efficiency of the market for racetrack betting," *Management Science*, 1981, *27* (12), 1435–1452.

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

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

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

Kahneman, Daniel, "Maps of bounded rationality: Psychology for behavioral Economics," *American Economic Review*, 2003, *93* (5), 1449–1475.

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

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

_ and _ , "Reference-dependent risk attitudes," *American Economic Review*, 2007, *97* (4), 1047–1073.

Kuypers, Tim, "Information and efficiency: an empirical study of a fixed odds betting market," *Applied Economics*, 2000, *32* (11), 1353–1363.

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

Martin, Ian and Dimitris Papadimitriou, "Sentiment and speculation in a market with heterogeneous beliefs," 2019.

Miller, Edward M, "Risk, uncertainty, and divergence of opinion," *Journal of finance*, 1977, *32* (4), 1151–1168.

Moskowitz, Tobias J, "Asset pricing and sports betting," *Chicago Booth Research Paper*, 2015, (15-26).

Ni, Sophie Xiaoyan, "Stock option returns: A puzzle," *Available at SSRN 1259703*, 2008.

O'Donoghue, Ted and Charles Sprenger, "Reference-dependent preferences," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 1, Elsevier, 2018, pp. 1–77.

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

_ and _ , "Noise, information, and the favorite-longshot bias in parimutuel predictions," *American Economic Journal: Microeconomics*, 2010, *2* (1), 58–85.

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

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

Pope, Peter F and David A Peel, "Information, prices and efficiency in a fixed-odds betting market," *Economica*, 1989, pp. 323–341.

Prelec, Drazen, "The probability weighting function," *Econometrica*, 1998, pp. 497–527.

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

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

Scheinkman, Jose A and Wei Xiong, "Overconfidence and speculative bubbles," *Journal of Political Economy*, 2003, *111* (6), 1183–1220.

Shefrin, Hersh, *A behavioral approach to asset pricing*, Elsevier, 2008.

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

_, "Prices of state contingent claims with insider traders, and the favourite-longshot bias," *Economic Journal*, 1992, *102* (411), 426–435.

Simsek, Alp, "Belief disagreements and collateral constraints," *Econometrica*, 2013, *81* (1), 1–53.

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

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

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

_ **and** _, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, 1992, *5* (4), 297–323.

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

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

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

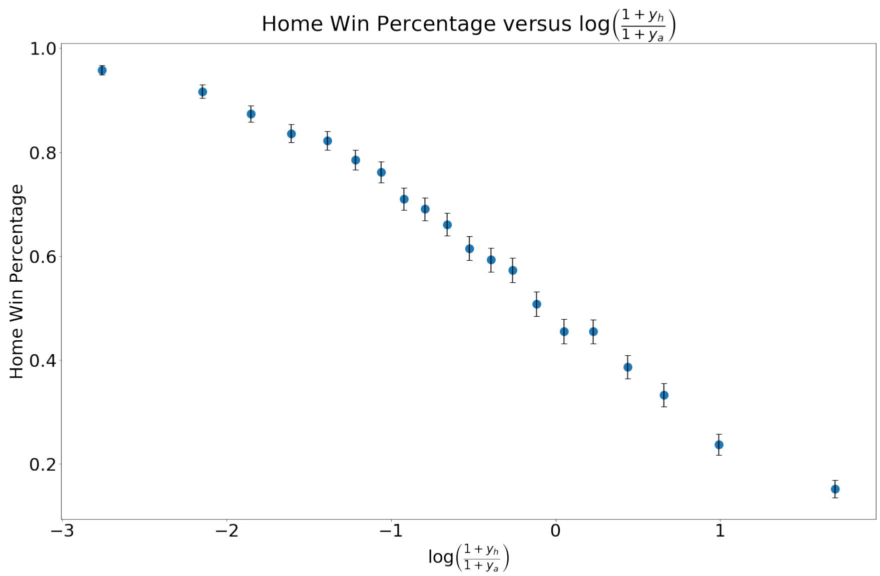
Woodland, Linda M and Bill M Woodland, "Market efficiency and the favorite-longshot bias: the baseball betting market," *Journal of Finance*, 1994, *49* (1), 269–279.

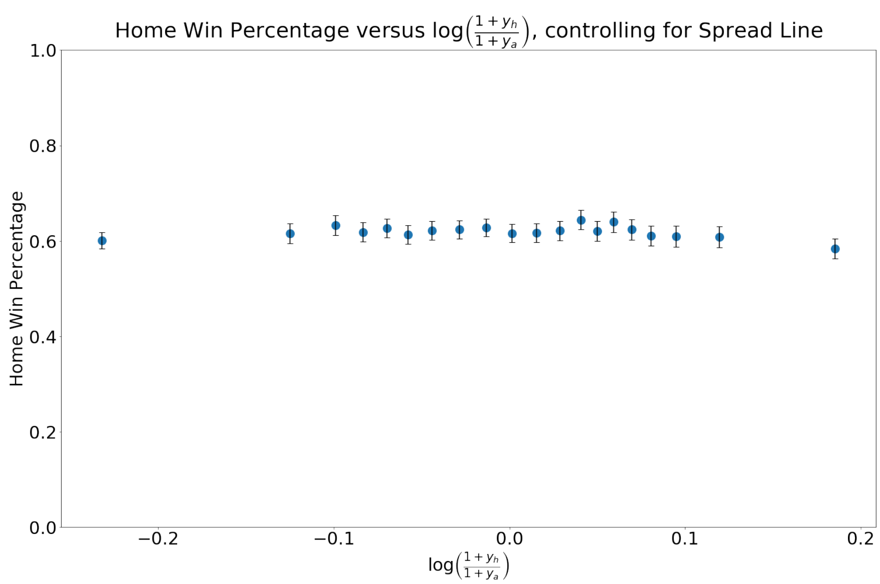
Zuber, Richard A, John M Gandar, and Benny D Bowers, "Beating the spread: Testing the efficiency of the gambling market for National Football League games," *Journal of Political Economy*, 1985, *93* (4), 800–806.

A Appendix

Figure A.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 bars correspond with plus/minus two standard errors relative to the home win percentage in each bin.





A.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 3). 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 A.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 A.1.

We plot a histogram of the realized point differential minus the quoted spread line for each sport in Figure A.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 not perfect fit of the data. In the exercise below, we use the calibrated normal distribution to calculate the implied volatility surface for each sport. However, the results do not crucially rely upon the normality assumption and do not change substantially when using other distributions.

Figure A.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 bars corresponds with 95% confidence intervals for the the decile standard deviation.

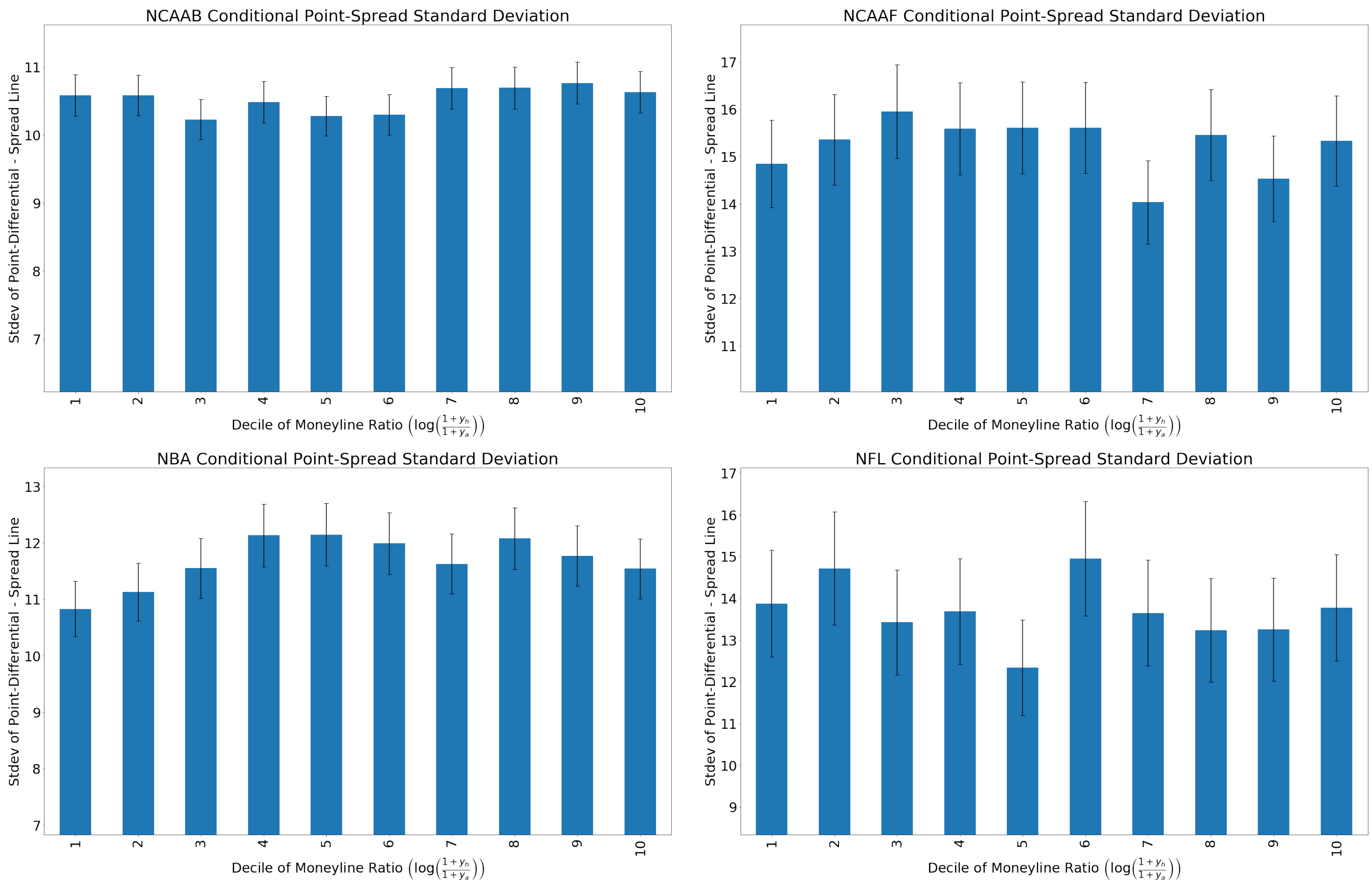


Table A.1: Relationship between Moneyline and Spread Lines

The table returns results from the regession 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	43964	8784	16784	3686	73218

Figure A.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.

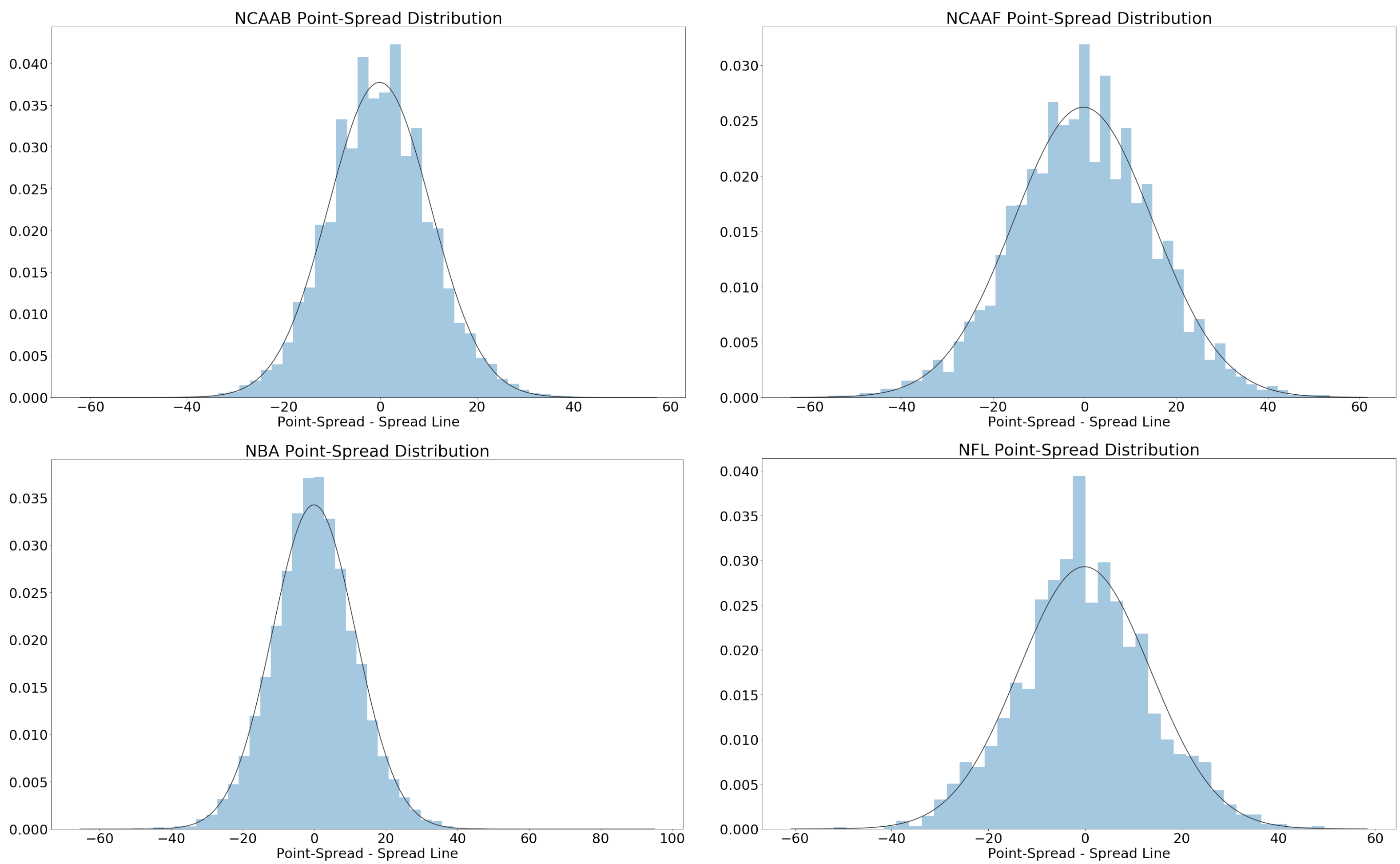


Figure A.4: Accuracy of Over/Under Contracts by Sport

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 bars correspond with plus/minus two standard errors relative to the mean number of points scored in a bin.

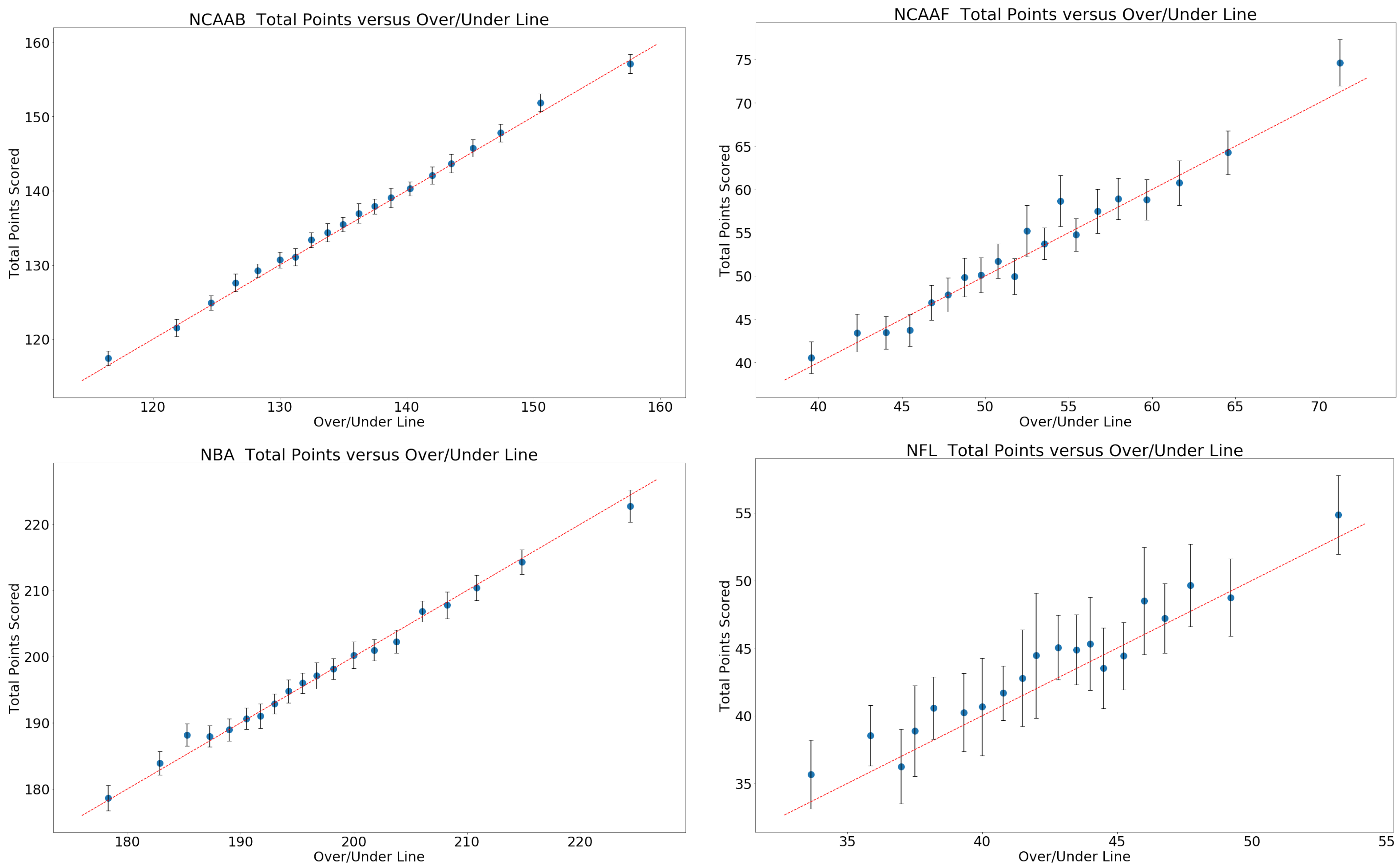
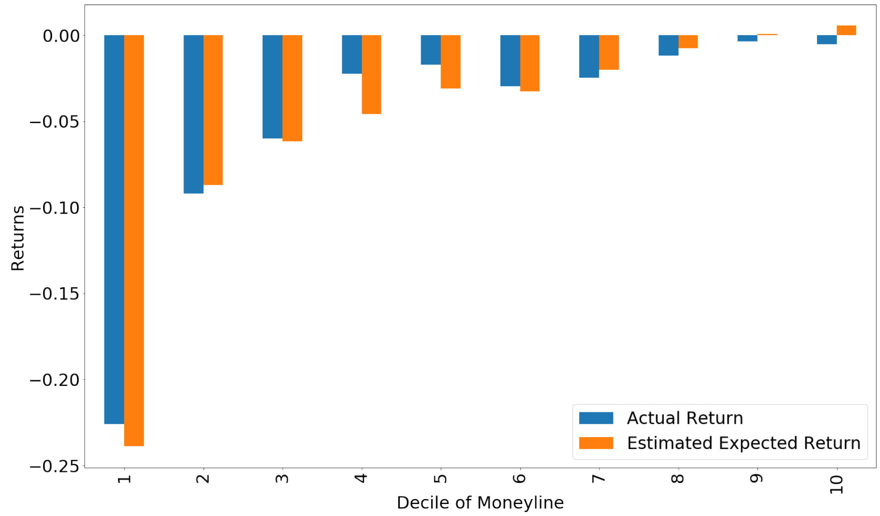


Figure A.5: Moneyline Contract Expected Returns v. Realized Returns



A.2 Non-Parametric Estimation of Win Probabilities

Here we describe the procedure to estimate the win probabilities for Moneyline contracts used in Section 3.2.

Within each sport, we first run a non-parametric regression of contract returns on the closing Moneyline of each contract and the closing Spread line of the game, running regressions separately for contracts betting on the home team and contracts betting on 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 each contract. To validate the accuracy of the estimates, we 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.5. The expected returns match up with the realized returns of contracts within each decile and capture the patterns observed in the data.

We take the Moneyline for each contract and convert into the payoff associated with winning a $1 on the game, which we denote as x. Denoting the expected return as $E(R)$, we can write the expected return of each contract in terms of x and the probability of the contract paying off, p, which is the value we are interested in estimating.

$$E(R) = px + (1-p)(-1) \tag{25}$$

Re-writing Equation (25) in terms of p, we get $p = \frac{E(R)+1}{x+1}$. We use this expression to estimate win probabilities for each contract from estimated expected returns. To assess the estimated win probabilities, we categorize each contract into 50 equally-spaced bins within each sport 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.6. In each sport, we find that the estimated and realized win probabilities match up along a 45 degree line, with an especially tight relationship in the sports where we have more observations.

FIGURE A.6: ESTIMATED WIN PROBABILITIES VS. ACTUAL WIN PERCENTAGES

The figure plots a scatterplot estimated win probabilities against win percentages, with one panel for each sport. The methodology to estimate win probabilities is described in Appendix A.2. For each sport, 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.

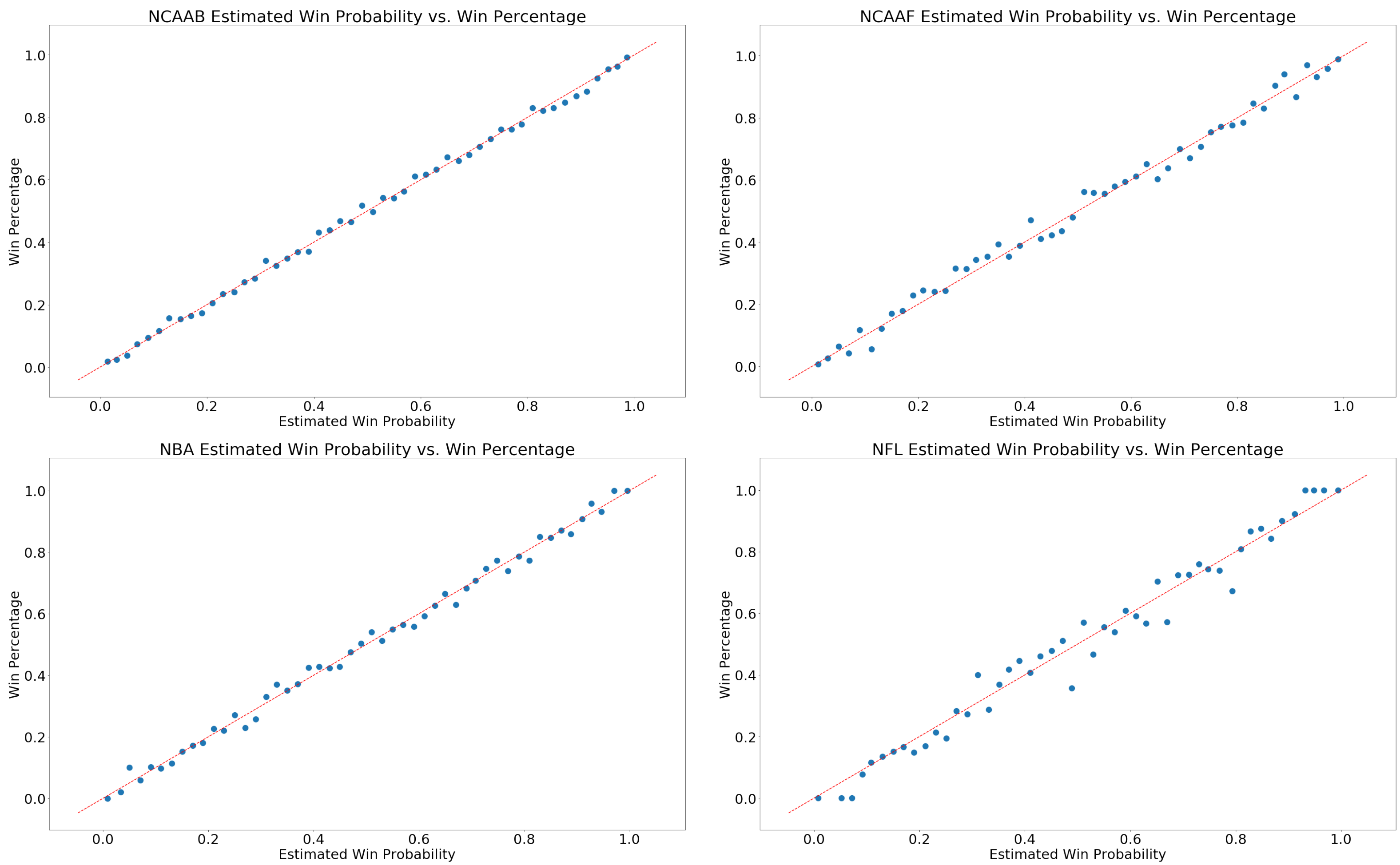


Table A.2: Index Option Sample

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

A.3 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 A.2.

Listed index options are European options, while listed equity options are American. Implied volatilites for European options are calculated using the Black-Scholes formula. Implied volatilities for American options are calculated by OptionMetrics using the Cox et al. (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 A.7: Cross-Section of Betting Returns with Opening Lines

The figure reproduces first Panel of Figure 1, the first Panel of Figure 2, and Figure 3, using opening lines to sort contracts and calculate returns rather than the closing lines.

]bigskip

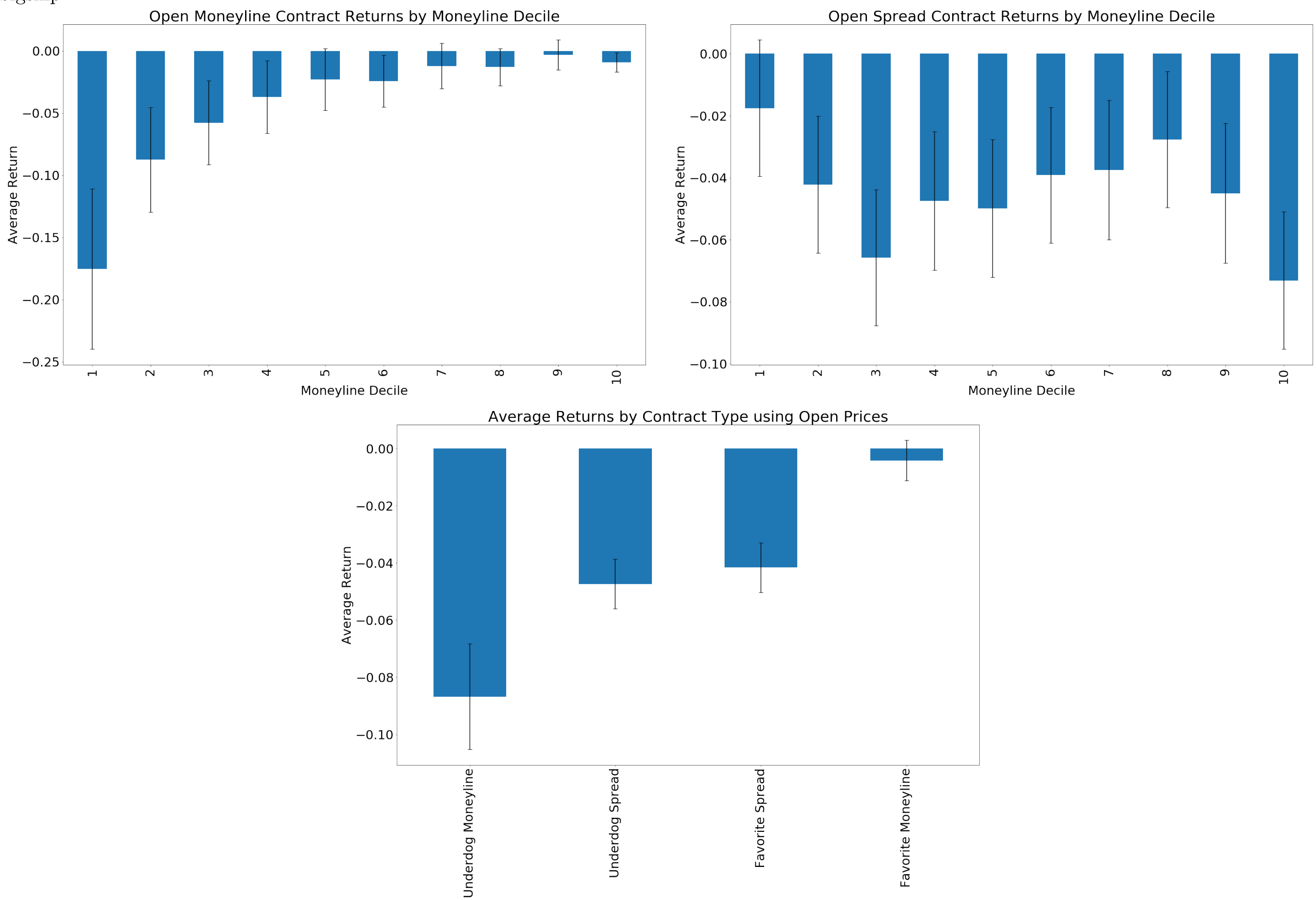
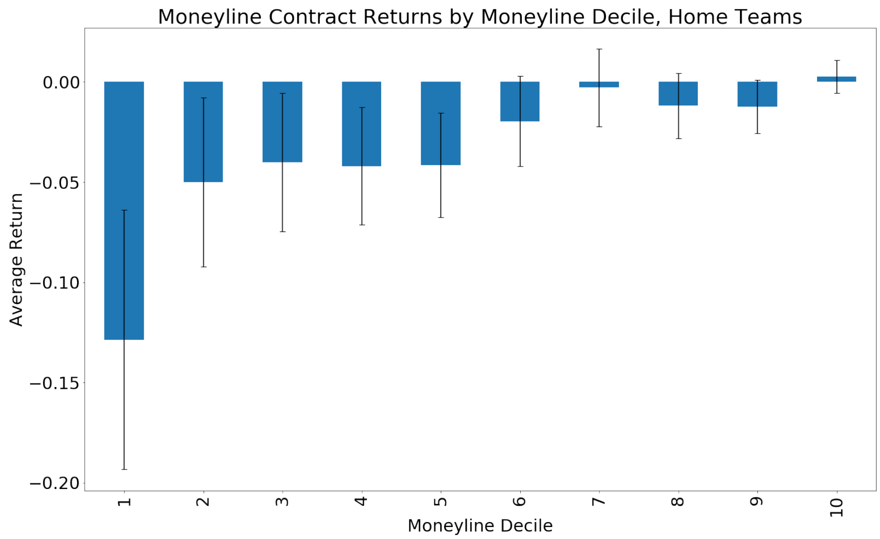


Figure A.8: Average Moneyline Contract Return Across Games

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 bars correspond with two standard errors above and below the average return for a particular contract type. 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.



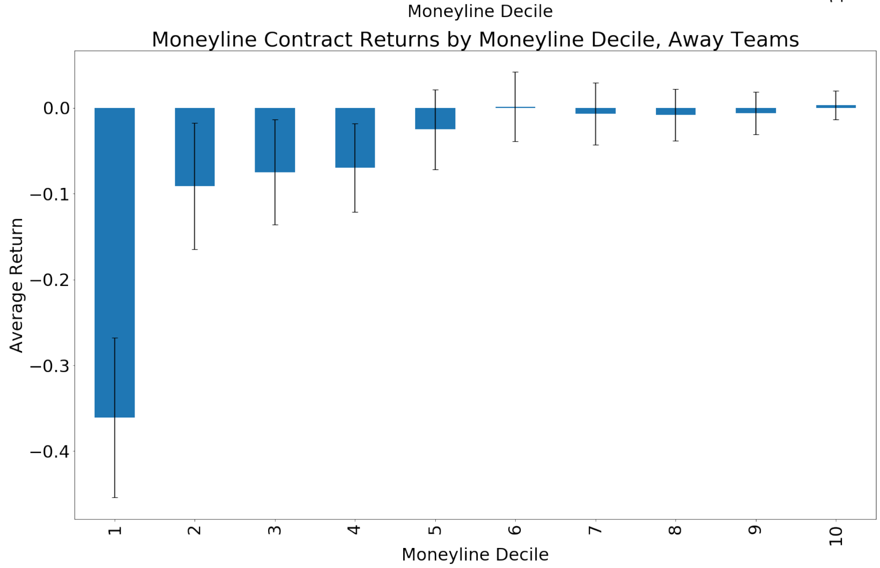
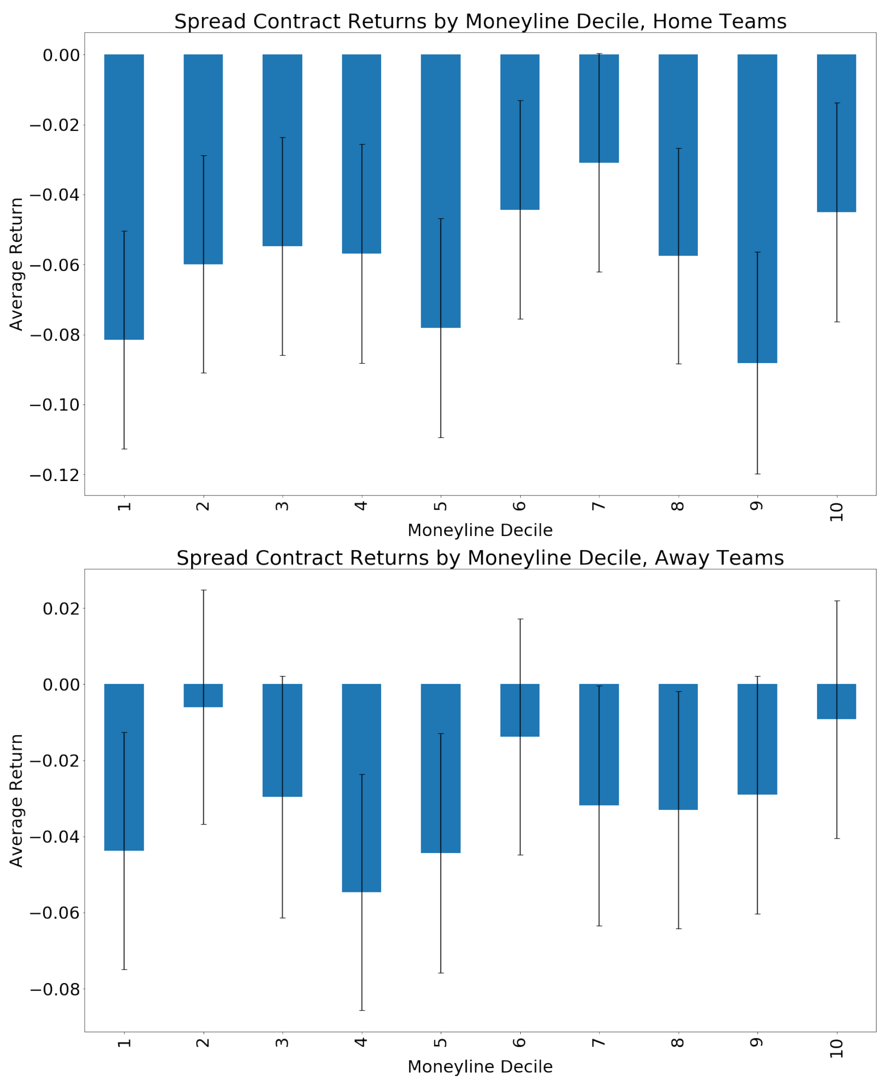


Figure A.9: Average Spread Contract Return Across Games

The figure plots the average returns of Spread 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 bars correspond with two standard errors above and below the average return for a particular contract type. 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.



A.4 Bet Sizes and Odds

We provide additional, supplemental evidence for the relationship between the estimated average bet size and odds presented in the text.

First, to calculate the estimated bet size ratios used in our calculation, we use the ratio of the number of bets placed on the two teams which is provided with our dataset. To understand how the proportion of bets placed on both teams varies across games, we construct the bet ratio for each game, defined as the ratio of the number of bets placed on the home team to the number of bets placed on the away team. In Figure A.10, we sort all games into 20 equally sized bins based on the odds ratio (the ratio of the odds the home team wins to the odds the away team wins) and plot a binned scatter plot of the average log odds ratio versus the average log bet ratio in each bin. The figure reveals that the bet ratio is, by-and-large, decreasing in the odds ratio. This is consistent with more bets being placed on the home team the more favored the home team is. However, the pattern isn't strictly monotonic - as the odds ratio becomes extremely negative (the home team is nearly assured victory), the proportion of bets placed on the home team *decrease* and the proportion of bets placed on the underdog away team actually *increase*. In Figure A.11, we plot the same binned scatterplot separately for each sport. The same patterns largely persist across sports.

Next, we turn our attention to the bet size ratios, which capture the size of the average bets placed on games. In Figure A.12, we replicate the binned scatterplot of the log odds ratio vs. the log bet size ratio presented in the main text, with one scatterplot for each sport. The figure reveals that the estimated relationship between the odds ratio and the bet size ratio is consistent across the sports in our sample, with bet sizes decreasing the less favored a team is.

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 setting with no bookmakers. We consider a sample of betting data for win bets on soccer matches. 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. The data come from Betfair, which is the world's largest betting exchange. For our sample, 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.

In Figure A.13, we replicate the binned scatterplot of the log bet size ratio on the log odds ratio in Figure 6 for the Betfair sample. The figure reveals a qualitatively and quantitatively similar pattern to what we find in our sample, with larger bet sizes on the favorite.

Figure A.10: Odds vs Estimated Bet Sizes By Sport

The figure plots a scatter plot of the log ratio of number of bets placed on the home team relative to the number of bets placed on the away team (the "bet ratio") versus the log ratio of the odds on the home team to the odds on the away team (the "odds ratio"). Each game is sorted into one of twenty equally sized bins based on the odds ratio of the game. Each point corresponds with the average log bet size ratio and average log odds ratio within a bin. The bars correspond with plus or minus two standard errors above and below the mean.

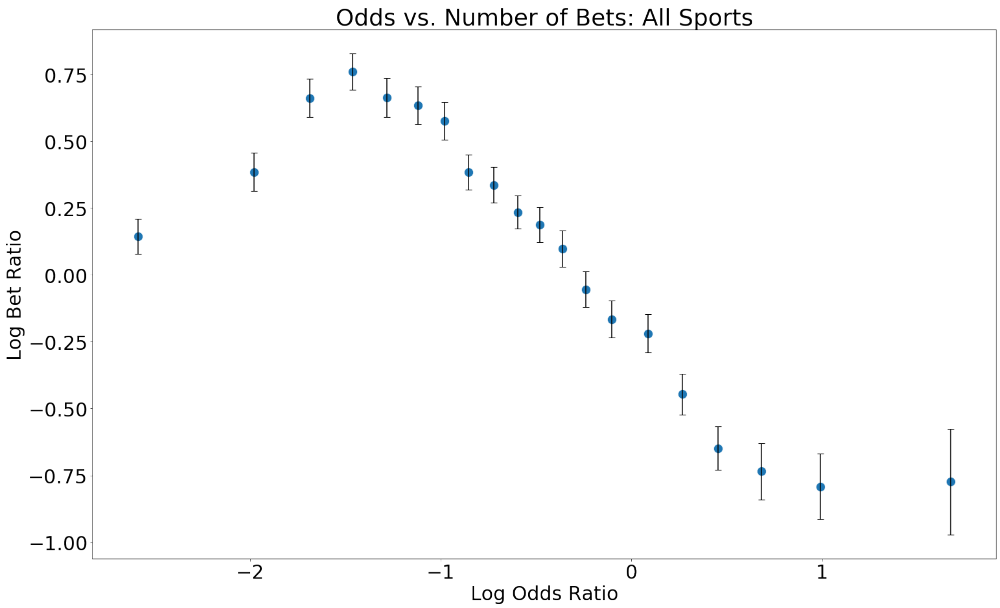


Figure A.11: Odds vs Estimated Bet Sizes By Sport

The figure replicates Figure A.10 for each sport in our sample. Each panel plots a scatter plot of the log ratio of the number of bets placed on the home team relative to the number bets placed on the away team (the "bet ratio") versus the log ratio of the odds on the home team to the odds on the away team (the "odds ratio"). Each game is sorted into one of twenty equally sized bins based on the odds ratio of the game. Each point corresponds with the average log bet ratio and average log odds ratio within a bin. The bars correspond with plus/minus two standard errors above and below the mean for each bin.

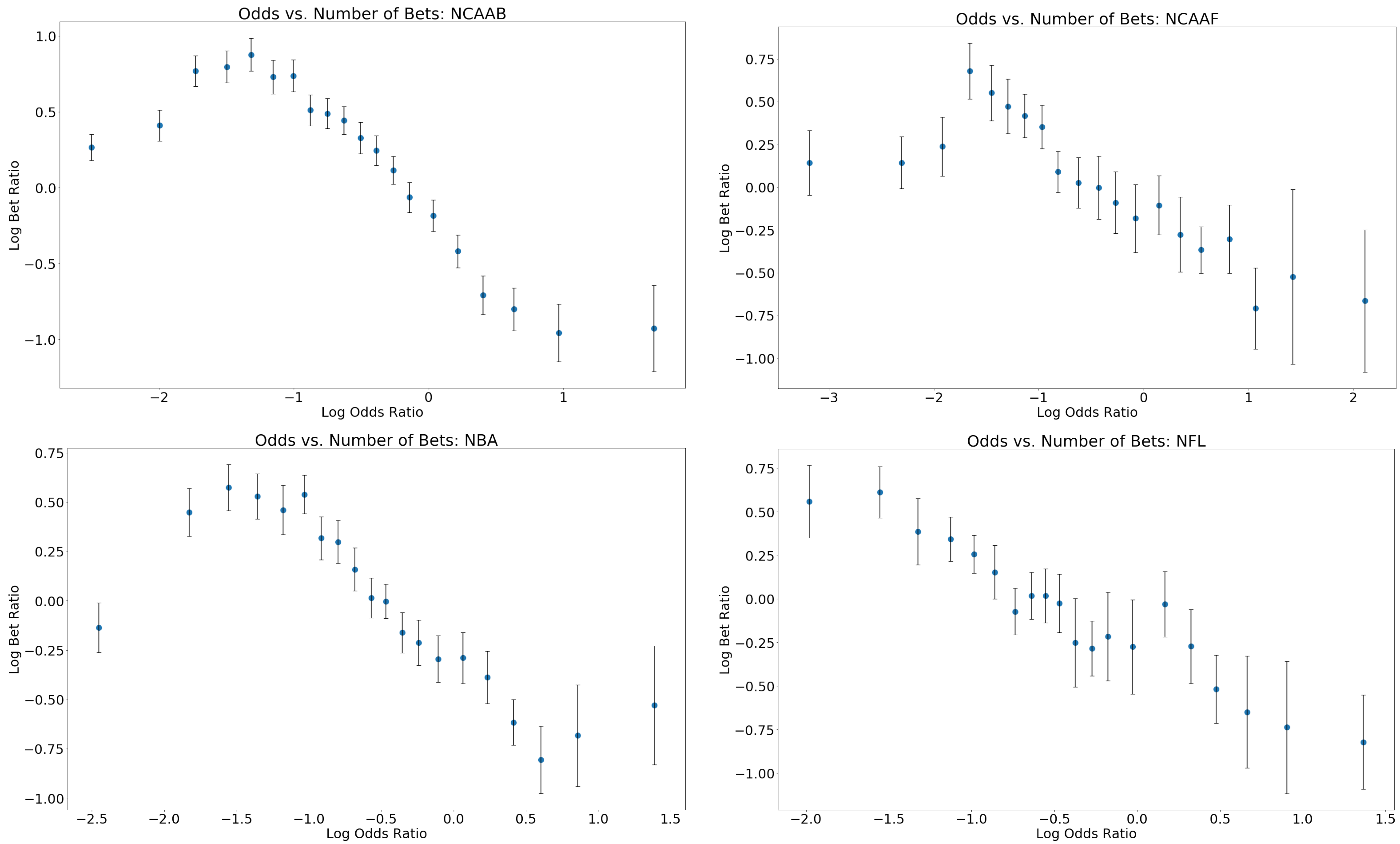


Figure A.12: Odds vs Estimated Bet Sizes By Sport

The figure replicates Figure 6 for each sport in our sample. Each panel plots a scatter plot of the log ratio of 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") versus the log ratio of the odds on the home team to the odds on the away team (the "odds ratio"). Each game is sorted into one of twenty equally sized bins based on the odds ratio of the game. Each point corresponds with the average log bet size ratio and average log odds ratio within a bin. The bars correspond with plus/minus two standard errors above and below the mean for each bin.

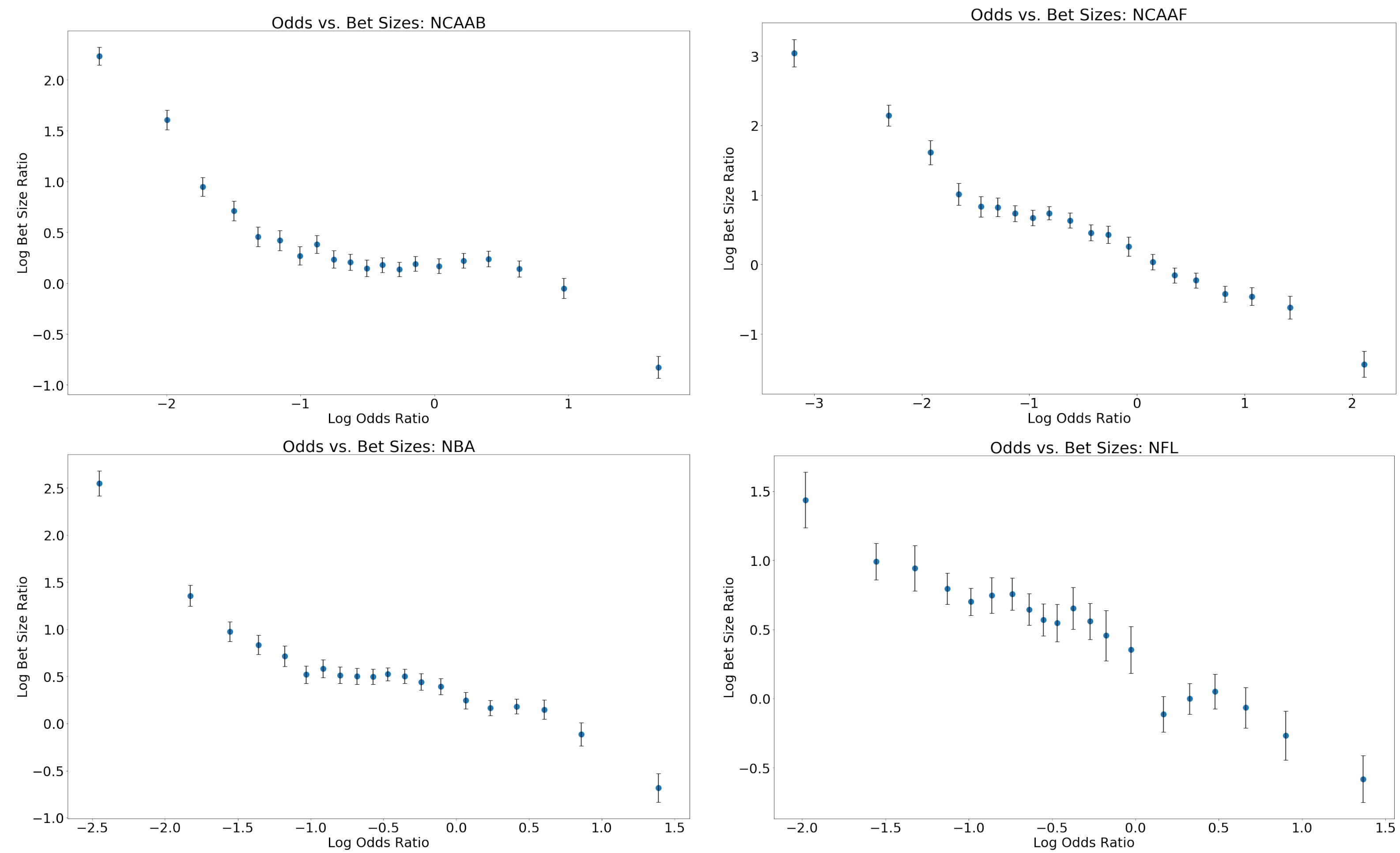
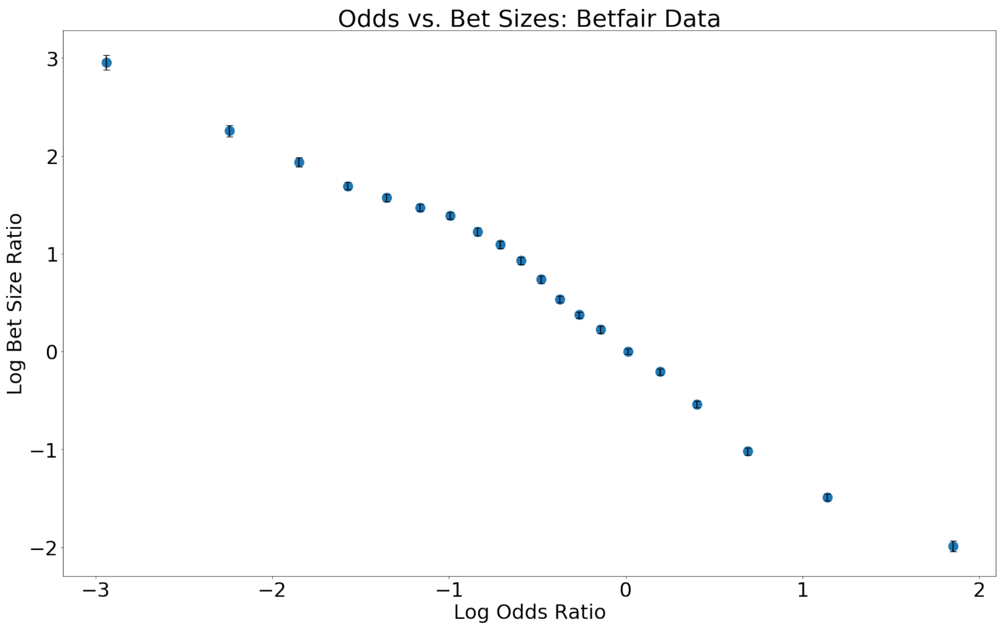


Figure A.13: Odds and Average Bet Sizes: Betfair Sample

The figure plots a scatter plot of the log ratio of the 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") versus the log ratio of the odds on the home team to the odds on the away team (the "odds ratio"). Each game is sorted into one of twenty equally sized bins based on the odds ratio of the game. Each point corresponds with the average log bet size ratio and average log odds ratio within a bin. The bars correspond with plus/minus two standard errors above and below the mean. The sample of games is a sample of major league soccer games from 2006-2011, with data coming from Betfair.



A.5 CPT Estimates Across Sports

Table A.3: CPT Parameter Estimates by Sport using Average Bet Sizes

The table displays Maximum Likelihood Estimates for Cumulative Prospect Theory preference parameters found by maximizing Equation (12), as in Panel A of Table 3. Here, we report separate results for each sport. The MLE estimates of parameters are calculated assuming that bettors are indifferent between betting the average amount wagered on the home and away team in each game. The first five columns present parameters estimates of γ, the probability weighting parameter, and λ, the loss-aversion parameter, for a given α, the diminishing sensitivity parameter. The last four columns of each panel present MLE estimates of α and λ for a given value of γ. Standard erorrs are reported in parentheses.

		Fixed α Estimates					Fixed γ Estimates		
	Fixed α	0.50	0.70	0.88	1.00	Fixed γ	0.65	0.90	1.00
NCAAB	γ	0.471	0.659	0.828	0.940	α	0.691	0.957	1.063
		(0.009)	(0.012)	(0.015)	(0.018)		(0.013)	(0.018)	(0.020)
	λ	0.966	0.953	0.941	0.934	λ	0.953	0.936	0.929
		(0.027)	(0.029)	(0.031)	(0.032)		(0.029)	(0.031)	(0.033)
NCAAF	γ	0.485	0.678	0.852	0.968	α	0.671	0.929	1.033
		(0.017)	(0.024)	(0.031)	(0.035)		(0.024)	(0.033)	(0.037)
	λ	0.908	0.893	0.879	0.871	λ	0.895	0.876	0.868
		(0.076)	(0.079)	(0.082)	(0.084)		(0.078)	(0.082)	(0.084)
NBA	γ	0.511	0.715	0.898	1.021	α	0.637	0.882	0.980
		(0.015)	(0.021)	(0.027)	(0.030)		(0.019)	(0.026)	(0.029)
	λ	0.967	0.954	0.943	0.936	λ	0.959	0.943	0.937
		(0.052)	(0.054)	(0.056)	(0.058)		(0.053)	(0.056)	(0.057)
NFL	γ	0.556	0.778	0.979	1.112	α	0.585	0.810	0.900
		(0.040)	(0.056)	(0.071)	(0.081)		(0.042)	(0.059)	(0.065)
	λ	0.750	0.733	0.717	0.706	λ	0.743	0.724	0.715
		(0.128)	(0.129)	(0.130)	(0.131)		(0.127)	(0.128)	(0.128)