CONSTANT BET SIZE? DON'T BET ON IT! TESTING EXPECTED UTILITY THEORY ON BETFAIR DATA

František Kopřiva

CERGE

EI

Charles University
Center for Economic Research and Graduate Education
Academy of Sciences of the Czech Republic
Economics Institute

Constant Bet Size? Don't Bet on It! Testing Expected Utility Theory on Betfair Data

František Kopřiva

Constant Bet Size? Don't Bet on It! Testing Expected Utility Theory on Betfair Data[*]

František Kopřiva[†]

CERGE–EI[‡]

Abstract

I analyze the risk preferences of bettors using data from the world's largest betting exchange, Betfair. The assumption of a constant bet size, commonly used in the current literature, leads to an unrealistic model of bettors' decision making as a choice between a high return - low variance and low return - high variance bet, automatically implying risk-loving preferences of bettors. However, the data show that bettors bet different amounts on different odds. Thus, simply by introducing the computed average bet size at given odds I transform the bettor's decision problem into a standard choice between low return - low variance and high return - high variance bets, and I am able to correctly estimate the risk attitudes of bettors. Results indicate that bettors on Betfair are either risk neutral (tennis and soccer markets) or slightly risk loving (horse racing market). I further use the information on the average bet size to test the validity of Expected utility theory (EUT). The results suggest that, when facing a number of outcomes with different winning probabilities, bettors tend to overweight small and underweight large differences in probabilities, which is in direct contradiction to the linear probability weighting function implied by EUT.

Keywords: decision making under risk, expected utility theory, betting exchanges
JEL classification: D01, D03, D81

[†]Email: frantisek.kopriva@cerge-ei.cz

[‡]CERGE-EI is a joint workplace of Charles University in Prague and the Economics Institute of the Academy of Sciences of the Czech Republic, Politických vězňů 7, 111 21 Prague, Czech Republic. Address: CERGE–EI, P.O. Box 882, Politických vězňů 7, Prague 1, 111 21, Czech Republic

Abstrakt

V článku se zabývám analýzou chování sázejících na největší online burze sázek - Betfair. Cílem této analýzy je bližší pochopení jednoho ze základních kamenů ekonomické teorie - rozhodování lidí v podmínkách nejistoty a jejich přístup k riziku. Sázení je jednou z mála situací v reálném životě, které se pro ověřování různých teorií o chování lidí v rámci rizika a nejistoty přímo vybízí. Všechny dosavadní studie na toto téma však opomíjely jeden podstatný fakt v rozhodování sázkařů - při výběru sázky hraje významnou roli nejen daný kurz, ale také výše sázky. Využitím dat z online burzy sázek nejen o výsledných kurzech na danou sázkovou příležitost ale i o, z dat vypočtených průměrně, vsazených částkách, tato studie velmi významně přispívá k analýze chování v rámci rizika a nejisoty. Studie, za použití nově navržené metodologie, testuje platnost jednoho z hlavních předpokladů EUT o racionálním přístupu k pravděpodobnostem. Moje výsledky naznačují, že sázkaři dávají při rozhodování v rámci rizika malým rozdílům v pravděpodobnostech blízko nule vyšší váhu než čistě racionálně smýšlející člověk a naopak velkým rozdílům v pravděpodobnostech výrazně nižší váhu než čistě racionálně smýšlející člověk.

1 Introduction

Expected utility theory (EUT) is considered to be one of the keystones of modern economic theory, yet its validity has been challenged by a large number of studies. The most prominent critique of EUT concerns the assumption that probability enters linearly into people's preferences over lotteries . As pointed out by Tversky and Kahneman (1992), the impact of the probability on the preferences over lotteries also depends on its distance from the so-called reference points - certainty and impossibility. This notion became the building block of *behavioral theories*[1] of decision making under risk and uncertainty, and led to the introduction of non-linear probability weighting functions.

A number of experiments document that behavioral theories are able to explain decision making under risk and uncertainty remarkably better than can EUT. There are, however, few empirical studies which assess the validity either of EUT or of behavioral theories in real situations. An innovative strand of empirical literature on this topic analyzes price data (odds) from *betting markets*[2]. These papers generally try to explain the existence of *favorite-long shot bias*[3], where bets on low probability outcomes of events have a lower expected return than bets on high probability outcomes; an observation which is not consistent with standard EUT under the classic risk-averse utility function assumption. To explain this inconsistency, two lines of argument have been used - either positing a risk-loving utility function under EUT, or introducing probability weighting functions in behavioral theories (Snowberg and Wolfers 2010).

The main drawback of previous studies on betting markets is, however, the

---

[1]See for example Tversky and Kahneman (1992) - Cumulative prospect theory (CPT); and Quiggin (1982)- Rank-dependent expected utility theory (RDEU).

[2]Weitzman (1965), Ali (1977), Kanto, Rosenqvist, and Suvas (1992), Hamid, Prakash, and Smyser (1996), Golec and Tamarkin (1998), Jullien and Salanie (2000), Bradley (2003), Gandhi (2008), Snowberg and Wolfers (2010).

[3]Favorite-long shot bias is one of the most prominent empirical regularities observed on betting data and was first noted by Griffith (1949) in horse racing betting markets.

absence of data on bet size. With the exception of Bradley (2003), they all posit an implicit assumption that bettors place the same amount of money on outcomes with different odds (i.e. that the bet size is constant irrespective of the probability of the outcome). As discussed in the next section, if we allow bettors to bet different amounts on outcomes with different odds, their decision problem is transformed into a standard choice between low return - low variance and high return - high variance bets. Thus, the existence of long shot bias can be consistent with the standard risk-averse utility function under EUT, and need not resort to behavioral theories for explanation.

I design a novel empirical test to assess the validity of EUT vs. behavioral theories using information on how much bettors bet on different outcomes of a particular event. Further, applying data from the world's largest betting exchange, Betfair, to a wide range of events (tennis, soccer and horse races), for which outcomes span the whole range of winning probabilities, allows me to analyze decision under risk and uncertainty under various scenarios. I draw conditioned subsamples based on the occurrence of a favorite in the event (i.e. event with/without a clear favorite),[4] using odds as a proxy for the objective probabilities of winning. These subsamples, and particularly the ratio of bets on different outcomes among events, provide rich information to test whether bettors weight probabilities linearly. As the conditioned subsamples fundamentally differ in their probability ranges of outcomes, finding substantial differences in how bettors assess the probabilities and determine the ratio of bet sizes on different outcomes in these conditioned subsamples strictly contradicts the linear probability weighting function assumption in EUT.

The paper is structured as follows. Section 2 analyzes the implications of the

---

[4]Conditioning on races with high-probability winning horses was used first by Golec and Tamarkin (1998) to address the problem of racetrack betting data, which consist of relatively few favorites (high-probability results) compared to the number of underdogs.

constant bet size assumption; section 3 outlines the methodology and estimation strategy. Data description is provided in section 4, section 5 presents and discusses the results and section 6 concludes

2   Constant Bet Size Assumption

In recent years, the emergence of literature that analyzes the behavior of bettors on betting markets has fostered great progress in the understanding of decision making under risk and uncertainty. Early studies analyzing the risk preferences of bettors treat all events (races) as identical, group them by different characteristics, e.g. by odds intervals or position of a horse in the race (see for example Ali (1977) and Kanto, Rosenqvist, and Suvas (1992)), and conduct their analysis on the aggregated values. A further advance in the field was introduced by Jullien and Salanie (2000) who design a new methodology which does not require aggregation because, as they argue, betting behavior may differ with different characteristics of the particular horse race event. Given the limited availability of data, however, these papers all rely on the assumption of a representative bettor and constant bet size, i.e. they estimate the preferences of an average or marginal bettor who is indifferent about betting the same amount on different outcomes of a particular event. ?) relaxed this assumption by assuming a pool of heterogeneous agents who differ in their preferences over the horses. Effectively, however, he estimates the behavior of several marginal bettors who are indifferent between betting on two outcomes instead of betting on all outcomes.

The underlying assumption of the previous studies implies that, under EUT, a marginal bettor facing an event with two outcomes is indifferent between betting on the favorite or on the underdog. Assume a horse race with just two horses, one favorite with a probability of winning p_F and decimal odds[5] O_F , and one underdog

---

[5]Bookmakers in Europe, the UK and the US have different standards of displaying odds.

with probability p_U ($p_U = 1 - p_F$) and odds O_U, where $p_U < p_F$ and $O_U > O_F$. The marginal bettor bets a constant amount B (i.e., bet size is constant). Then, the marginal bettor is indifferent between betting on the favorite vs. betting on the underdog if and only if

$$
\begin{aligned}
EU_F &= p_F u(M + (O_F - 1)B) + (1 - p_F)u(M - B) \\
&= u(M + (O_U - 1)B) + (1 - p_U)u(M - B) = EU_U \\
&\quad p_F u(M + (O_f - 1)B) - p_U u(M + (O_U - 1)B) = (p_F - p_U)u(M - B)
\end{aligned}
\tag{1}
$$

In the presence of long shot bias the return on the favorite is higher than that on the underdog:

$$
\begin{aligned}
p_F(M + (O_F - 1)B) + (1 - p_F)(M - B) &> p_U(M + (O_U - 1)B) + (1 - p_U)(M - B) \\
p_F(M + (O_F - 1)B) - p_U(M + (O_U - 1)B) &> (p_F - p_U)(M - B)
\end{aligned}
\tag{2}
$$

Without loss of generality we can assume that $M - B = 0$ and $u(0) = 0$. Thus,

$$
\frac{u(M + (O_U - 1)B)}{u(M + (O_F - 1)B)} = \frac{p_F}{p_U} > \frac{(M + (O_U - 1)B)}{(M + (O_F - 1)B)}
\tag{3}
$$

As $M + (O_U - 1)B > M + (O_F - 1)B$, the utility function has to be convex at least in some range of the interval $(M + (O_F - 1)B, M + (O_U - 1)B)$ - i.e., the marginal bettor has to exhibit risk-loving preferences.

Following the same line of thought, Snowberg and Wolfers (2010) point out that, in the presence of long shot bias and without assuming a non-linear probability weighting function, one has to allow for the risk-loving preferences of a marginal

---

Further in the text I use the European style of odds, also called decimal odds. Odds $O^E = 1.40$ imply that the bet will bring payoff $(O^E - 1)B = 0.40B$ if the outcome wins and $-B$ if the outcome loses. In the UK odds are usually displayed in the form $O^{UK} = x/y = 2/5$ s.t. $O^E = (x + y)/y$. In the US odds are displayed in the form $+X$ or $-X$ where the negative odds are for those bets where the payoff is lower than the bet. In our case the US odds are $O^{US} = -X = -250$; $O^E = 1 + 100/X$ if the US odds are negative and $O^E = 1 + X/100$ if the US odds are positive.

bettor under EUT. They recognize that using information on price data (odds) alone for simple bets (e.g., a bet on the winner of a horse race) is not sufficient to confirm the validity of EUT vs. behavioral theories. Instead, they compare the price data on simple bets - win bets and compound bets - exactas, trifectas and quinellas (exotic bets on the order of the first two or three horses and on the two horses to come first in the race in either order) and find evidence in favor of behavioral theories. Therefore, they conclude that the long shot bias is mainly driven by the misperception of probabilities rather than by the risk-loving preferences of rational bettors.

Nevertheless, these previous studies lack important information, namely how much people bet on different outcomes with different winning probabilities. The only study to account for bet size in the analysis of bettors' behavior is Bradley (2003). As he does not have data on bet size, he performs his analysis using the imputed optimal bet size of a representative bettor. By assuming that the only utility that a bettor has from a bet is derived from expected return and variance, he computes the optimal bet as an argument for the maximum weighted expected utility given the probabilities and odds. However, he still does not consider the main reason for including bet size, namely how it changes the estimates of the revealed risk preferences of the marginal bettor. If one allows the marginal bettor to bet amount B_F on the favorite and B_U on the underdog, the above stated key formula for identification of his risk preferences changes to

$$
\begin{aligned}
EU_F &= p_F u(M + (O_F - 1)B_F) + (1 - p_F)u(M - B_F) \\
&= p_U u(M + (O_U - 1)B_U) + (1 - p_U)u(M - B_U) = EU_U \qquad (4)
\end{aligned}
$$

Let me define the following function for the ensuing analysis

$$
\mathcal{F}(\theta, B_F, B_U, p_F, p_U, O_F, O_U, M) = EU_F - EU_U \qquad (5)
$$

where θ represents the risk preference parameter of the utility function. Contrary to the previous case, conditional on the ratio of the average bet sizes on the two outcomes, equation (4) may have no, one, or two solutions under standard utility assumptions. Henceforward, all examples are produced under EUT with the standard CARA utility function assumption $u(x, \theta) = (1 - e^{-\theta x})/\theta$, where $\theta > 0$ corresponds to risk-averse preferences.[6]

In the first step, I focus on the analysis of the fair odds case (i.e., no long shot bias present). Under EUT the risk-neutral bettor should be willing to bet any amount of money, which is unrealistic. Therefore, it is more reasonable to define a risk-neutral bettor in terms of the bet size ratio as the limit case between a risk-loving and risk-averse bettor (see Figure 1 below).

Figure 1: Bet size ratio with fair odds
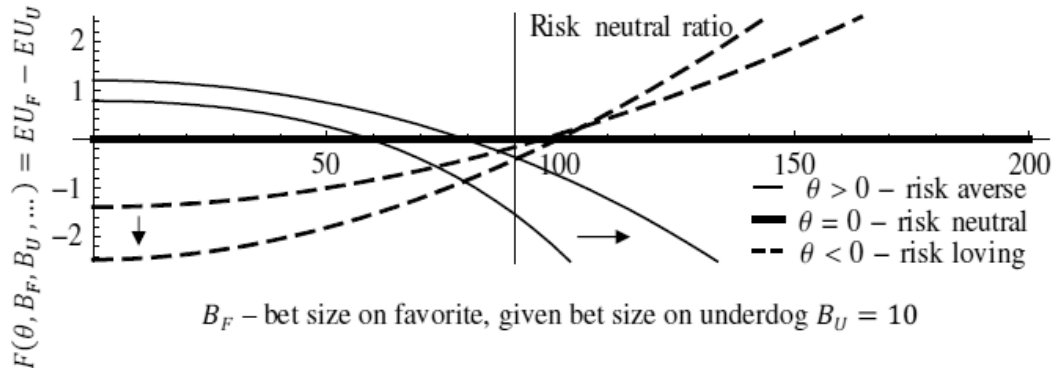
Note: The figure depicts the difference of expected utility of betting on the favorite and betting on the underdog under fair odds, with the bet size on the favorite on the horizontal axis, for different risk aversity parameters θ of the CARA utility function. The difference is illustrated in an example, where the probability of the favorite winning is 90% and the probability of the underdog winning is 10%. The arrows illustrate the shift towards less risk-averse values of parameter θ.

Mathematically, the risk-neutral bettor would choose, under fair odds, the ratio:

---

[6]One would get similar results with the standard CRRA utility function $u(x, \rho) = x^{(1-\rho)}/(1-\rho)$.

$B_F/B_U = \lim_{\theta \to 0}(B_F/B_{(}U)$ such that $F(\theta, B_F, B_{(}U), \ldots) = 0)$, with solution

$$B_F/B_U = \frac{p_F}{\sqrt{p_F(1 - p_F)}} \Big/ \frac{p_U}{\sqrt{p_U(1 - p_U)}}$$

In the case with two outcomes, the previous expression boils down to

$$B_F/B_U = p_F/p_U \text{ as } p_F + p_U = 1$$

In the example depicted in Figure 1, this corresponds to the $B_F = 90$ (bet ratio 9:1). This analysis can be used to derive the risk preferences of the bettor from the ratio of bets that I would observe in the data without long shot bias. In Figure 2, I distinguish two possible cases of bet ratios B_F/B_U : 1.) $B_F/B_U \leq 1$: This case corresponds to the constant bet size assumption (thick line in Figure 2). In this case, equation 4 has one closed solution in $\theta = 0$ and one limit solution for $\theta \to \infty$. However, the above analysis implies that the ratio of those bets consistent with the behavior of a risk-neutral bettor is definitely higher than 1. Thus, I pick the limit solution $\theta \to \infty$ as the correct one and interpret the risk preferences of the marginal bettor who bets constant amounts on both outcomes as extremely risk averse.[7] 2.) $B_F/B_U > 1$: In line with the previous argument, the risk preferences continuously shift from extremely risk averse to extremely risk loving. Based on the bet size ratio we can distinguish two cases: if the bet size ratio is lower than that chosen by the risk-neutral bettor, then the marginal bettor is risk averse (Figure 2, dashed line); whereas if the bet size ratio is higher, we can infer that the marginal bettor has risk-loving preferences (Figure 2, dot-and-dashed line), with a risk-neutral ratio of bets between them (Figure 2, thin line).

In the second step, the analysis is generalized for the presence of long shot bias.

---

[7]This behavior may also be interpreted as a decision to bet a certain amount of money but an unwillingness to bet/lose any more money than that. Such behavior will be more consistent with immense risk-averse preferences than with risk neutral preferences.

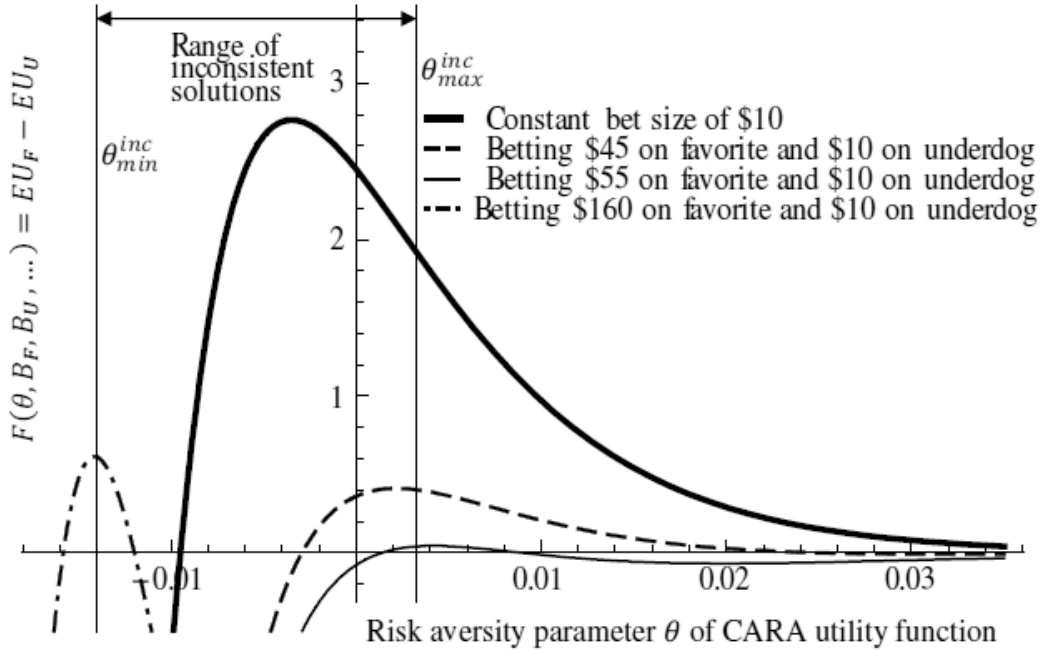Figure 2: Eliciting risk preferences from bet size ratio under fair odds
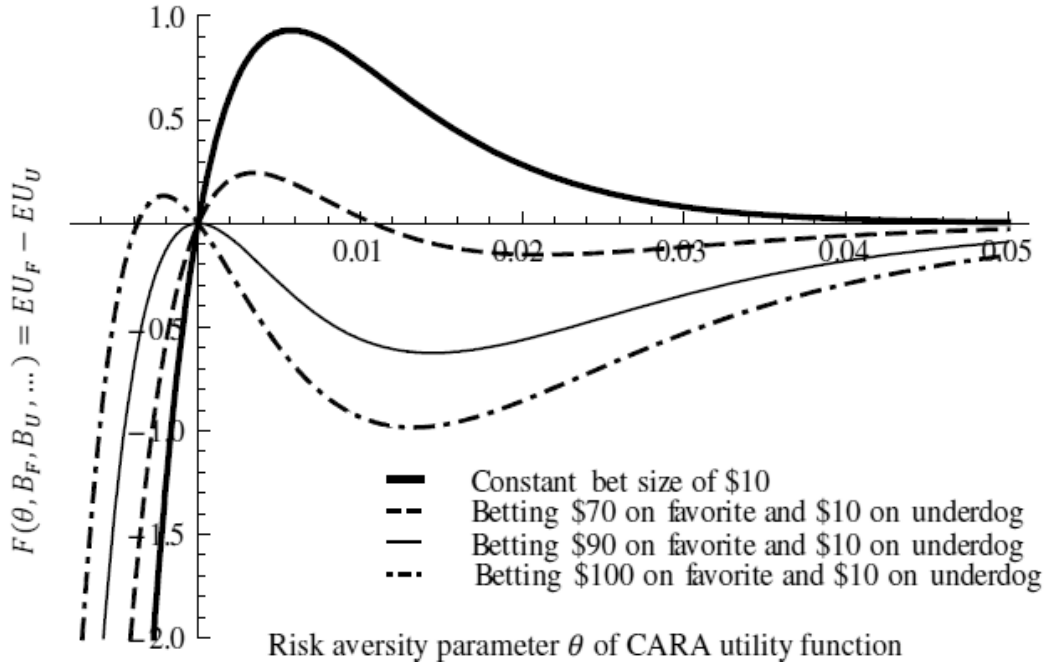
Note: The figure depicts the difference of expected utility of betting on the favorite and betting on the underdog under fair odds with the risk aversity parameter θ of the CARA utility function on the horizontal axis, given the bet size on the favorite and on the underdog. The difference is illustrated in an example, where the probability of the favorite winning is 90%, and the probability of the underdog winning is 10%.

Similar to the case with fair odds, the main assumption is that the ratio B_F/B_U , consistent with the indifference of the marginal bettor, increases with decreasing risk aversion. Thus, if I find a solution that does not satisfy this assumption I consider it to be inconsistent.

As presented in Figure 3, there exists a range $(\theta_{min}^{inc}, \theta_{max}^{inc})$ of risk parameter theta, which corresponds to inconsistent solutions (roots) of equation 4.[8] In this range, the lower risk aversion is connected to lower bet size ratio, contradicting the basic assumption. To better illustrate the inconsistency, Figure 4 depicts 6 lines corresponding with decreasing θ from strongly risk-averse (line 1) to strongly

_____

[8]One can show that if equation 4 has one inconsistent solution, it has a consistent solution (including the limit solution $\theta \to \infty$).

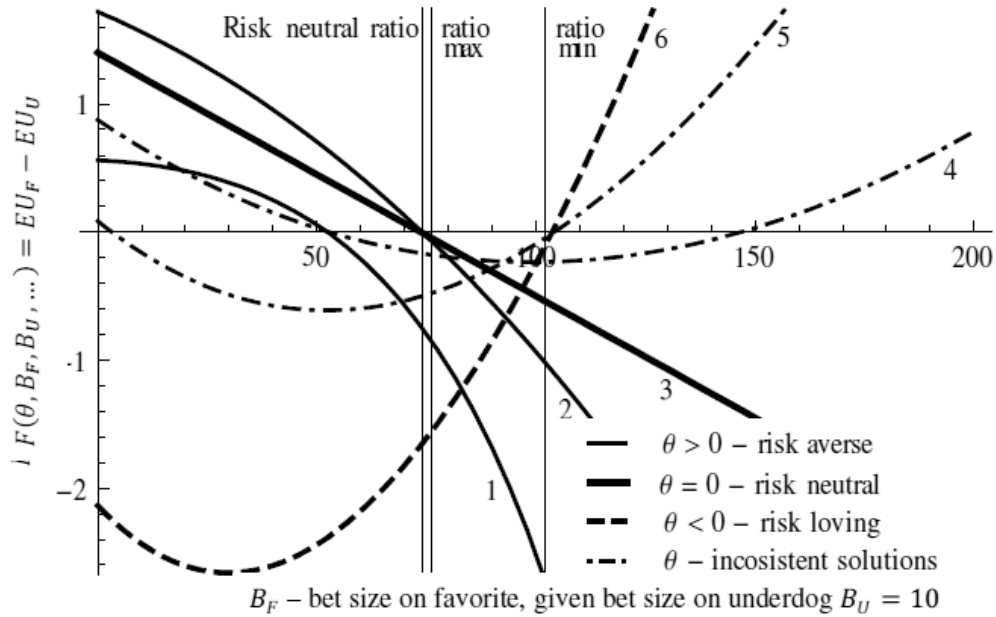Figure 3: Eliciting risk preferences from bet size ratio under long shot odds

Note: The figure depicts the difference of expected utility of betting on the favorite and betting on the underdog under long shot bias with the risk aversity parameter θ of the CARA utility function on the horizontal axis, given the bet size on the favorite and on the underdog. Long shot bias is present, i.e., odds on the favorite are 1.05 (with the probability of the favorite winning 90%), and odds on the underdog are 7 (probability of winning 10%), leading to an expected return of -0.06% on the favorite and -0.3% on the underdog.

risk-loving (line 6) preferences. Moving away from strongly risk-averse preferences, the bet size ratio increases until it reaches its maximum (line 2), marked as ratio max in the figure for $\theta = \theta_{max}^{inc}$ (from the previous Figure 3).

Decreasing θ even further, the bet size ratio starts to decrease (lines 3, 4 and 5 in Figure 3.4). These lines correspond to inconsistent values of θ within the range $(\theta_{min}^{inc}, \theta_{max}^{inc})$. Finally, starting at θ_{min}^{inc} (representing risk-loving preferences), the bet size ratio starts increasing again ($\theta < \theta_{min}^{inc}$; line 6). At this point the bet size ratio has reached its minimum for risk-loving preferences, marked as ratio min in Figure 4. For the ratio of bets between (ratio max; ratio min), no solution to equation 4 exists.

11

Figure 4: Bet size ratio under long shot bias

Note: The figure depicts the difference of expected utility of betting on the favorite
and betting on the underdog under long shot bias with the bet size on the favorite
on the horizontal axis, given the risk aversity parameter θ of the CARA utility
function. The difference is illustrated in an example, where the probability of the
favorite winning is 90% and the probability of the underdog winning is 10%. Lines
are numbered from most risk averse (1) to least risk averse/most risk loving (6).

It should be stressed that the solution for constant bet size lies within the range
of inconsistent solutions with $\theta < 0$ (Figure 3, thick dark line). This has led authors
who held to the constant bet assumption to the erroneous conclusion that bettors
have to exhibit risk-loving preferences under EUT.

3   Methodology

The microstructure of Betfair as a typical *betting exchange* differs from the
classic betting markets, and in certain respects it is more like financial markets.
Every market on the exchange (for example the event on the winner of a horse
race) consists of several outcomes with ex-ante objective probabilities of happening

$p_1, \ldots, p_N$. For every outcome of the event bettors have two options: To place a bet that the outcome will happen - the *back bet* in Betfair terminology; or to place a bet that the outcome will not happen - the *lay bet*. The Betfair betting exchange is designed as an order-driven market where bettors can place either limit orders or market orders. *Market order* means that the bettor just chooses a side (buy or sell on classic markets, back or lay on Betfair), a particular outcome, and a bet size. The bet is then matched at the best possible price available on the market. *Limit order* means that the bettor is not satisfied with any market odds available at the moment and chooses not only the side, outcome, and volume, but also the odds at which he is willing to bet. The bet then waits on the market until it is matched by some other bettor. Therefore, when placing a limit order, the bettor has to decide whether to place a back or lay bet and has to stipulate the odds and bet size. On the other hand, when placing a market order, the bettor hits the odds already available on the market and chooses just the bet size and side of the market.

Assume that the bettor decides to place a back bet (the outcome will happen) on outcome 1 of volume one dollar at odds O_1. With probability p_1, outcome 1 occurs and the bet yields profit $(1 - \tau)R_1 = (1 - \tau)(O_1 - 1)$, where τ is the commission (2-5%) that Betfair charges on the net winnings. If outcome 1 does not occur the bettor will lose one dollar. If the bettor places a lay bet (the outcome will not happen) on outcome 1 of volume one dollar at odds O_1, the bet yields profit $(1 - \tau)$ dollars if outcome 1 does not occur and loss $-R_1$ if outcome 1 occurs. As I focus on those markets with one possible winner, the probabilities $p_1, \ldots, p_N$ sum up to one. Thus, backing an outcome at odds is actually the same as laying all the other outcomes at respective odds.

Generally, bettors may have different prior beliefs about the underlying probabilities of winning of the outcomes. However, they update their beliefs using market

prices. Therefore, in equilibrium, all bettors can use the odds to infer the true underlying probabilities of the outcomes. Further, bettors can be divided into three main categories - common bettors, bookmakers, and traders. I assume that the majority of Betfair customers (more than 2 million people) may be characterized as *common bettors*, who typically bet only on one outcome and mostly place back market orders. I discuss this particular assumption and its implications on the results in the Appendix.

The other two types of bettors - bookmakers and traders - are professional bettors who use the Betfair markets for making a profit. I assume that bookmakers post mostly large volume limit orders and only occasionally use market orders to balance their portfolios.[9] I consider them to be risk neutral, as they basically try to balance their liabilities and earn a profit from the spread. The third type of bettor, traders, are similar to bookmakers. Their main concern, however, is not to make money from the spread but to identify arbitrage opportunities. Therefore, they place both limit and market orders, open and close their positions, and earn their profit from the differences of the asset price over time. These bettors are, therefore, usually placing large volume orders and the size of their bets is balanced with respect to the odds, i.e., they are also acting as risk-neutral bettors.

In September 2008 Betfair introduced a new policy of "premium charges", requiring customers who consistently win to pay at least 20% of their total profits in commission or other charges. Although this rule was aimed at bookmakers and traders, Betfair claimed that it affected less than 0.5% of its customers. Since, according to this statement, bookmakers and traders make up less than 0.5% of all bettors, I direct my attention to the majority - common bettors - when analyzing the risk attitude of the general population of bettors.

---

[9]In analyzing the in-trade soccer markets, Gil and Levitt (2007) point out that the endogenously emerged market makers were on one side of the trade for 65 percent when the markets were inplay, i.e. betting during the running event.

Due to the differences in market microstructure and in the behavior of bettors on different sports markets I analyze the tennis, soccer and horse race events separately. My empirical methodology follows the seminal paper of Jullien and Salanie (2000). For each event the common bettors face the following successive decisions:

1. The bettor decides whether or not to bet;

2. Conditional on the characteristics of the event, outcomes, and subjective probabilities of winning, the bettor decides how much he would be willing to bet on every outcome;

3. After observing the odds the bettor chooses which outcome in the event he will bet on.

The decisions in the first and second steps depend on both the event/outcome parameters and the personal characteristics of each bettor. All bettors have their own motives for betting and as no information about their personal characteristics is known, I do not model this decision. Further, I assume that the decisions of common bettors can be represented by the behavior of a representative agent - *marginal bettor* - with initial wealth M. The marginal bettor is able to anticipate from the equilibrium odds the true probability of winning of particular outcomes in the event. Furthermore, under EUT, for every two outcomes i, j on the market with given odds O_i and O_j, probabilities p_i and p_j, and average bet sizes B_i and B_j, the marginal bettor with given utility function and risk preference parameter θ is indifferent between betting on these two outcomes, such that

$$p_i u(M+B_i(1-\tau)R_i,\theta)+(1-p_i)u(M-B_i,\theta) = p_j u(M+B_j(1-\tau)R_j,\theta)+(1-p_j)u(M-B_j,\theta).$$

As the probabilities sum up to one, I obtain the analytical solution for probabilities in the form

$$p_i = \cfrac{\cfrac{1 + \displaystyle\sum_{j=1}^{N} \cfrac{u(M - B_j, \theta)}{u(M + B_j(1 - \tau)R_j, \theta) - u(M - B_j, \theta)}}{\displaystyle\sum_{j=1}^{N} \cfrac{1}{u(M + B_j(1 - \tau)R_j, \theta) - u(M - B_j, \theta)}} - u(M - B_i, \theta)}{u(M + B_i(1 - \tau)R_i, \theta) - u(M - B_i, \theta)} \qquad (6)$$

As I do not observe any information about the wealth or income of the marginal bettor, I use the CARA utility function in the form $u(x, \theta) = (1 - e^{-\theta x})/\theta$; otherwise the parameter estimates would be based either on the arbitrary choice of wealth M, or would have to be estimated as an additional parameter. Each p_i is uniquely defined by the set of B_i's, R_i's, and θ. Therefore, similarly to Jullien and Salanie (2000), θ is estimated by Maximum Likelihood Estimation using formula (6) of the probability of the winning outcome. The likelihood function is then a sum of logs of probabilities for ex-post winners p_W from each match:

$$L(\theta) = \sum_{c=1}^{C} \log p_w(R_1^c, \ldots, R_N^c, B_1^c, \ldots, B_N^c, \theta)$$

One of the key assumptions of alternative behavioral theories of decision making under uncertainty is that probabilities enter the formula of expected utility in a non-linear form. In other words, bettors have a non-linear probability weighting function. It is possible, however, to test the validity of EUT without explicitly formalizing the alternative theories. If the assumption of a linear weighting function of EUT is correct, then the estimated risk aversion parameters of the marginal bettor should be the same regardless of the winning probabilities of players/teams/horses. Therefore, for each sport I draw two subsamples: one with the presence of strong favorites (and large differences in winning probabilities between outcomes) and the other without a favorite (and small differences in winning probabilities between outcomes). Under the null hypothesis, EUT holds and therefore the estimates on

the subsamples should not be statistically different from each other. If the results differ, EUT can be rejected in favor of theories with non-linear weighting functions of probabilities.

4   Data

I use aggregated historical data from the world's largest betting exchange, Betfair, for all tennis, soccer and horse race winners' markets between June 2004 and December 2008. All the studies described in Section 2 analyze the risk preferences of bettors only on horse race markets. However, horse race events usually consist of a large number of outcomes (horses) with a low probability of winning and only a few outcomes with a high probability of winning. This could lead to a situation in which I would estimate the risk preferences of bettors just on those bets with a low probability of winning. As pointed out by Forrest and McHale (2007), however, the tennis betting markets possess the nice feature of having a nearly complete distribution of events with outcomes over the whole probability range. Using data from the tennis and soccer markets, then, allows me to analyze the behavior of bettors facing the complete set of probabilities.

For each outcome on every market and for each odds at which at least one bet was placed, the data from Betfair include information on: the number of bets placed, total volume matched, date and time of the first and last matched bet on given odds, scheduled and actual start of the event, indicator of inplay bets[10] and indicator of the winning outcome. Although on Betfair one can also place bets during the matches, I only use data on those bets that were placed before the start of the match or race, so as to analyze the ex-ante risk attitudes rather than the reaction of bettors to news from the ongoing match.

Recent studies on the risk attitude of bettors (e.g. (Jullien and Salanie 2000);

---

[10]Bets placed after the event has started

(Gandhi 2008)) have employed starting prices - the odds valid at the start of the event. At betting exchange markets there are, however, always two values of odds - back and lay. Moreover, the odds tend to fluctuate even before the start of the event; using just the final value of odds would result in loss of information about the volume matched and the number of bets placed at odds slightly different than the final odds.[11] I therefore use the weighted average of odds (by volume matched) at which at least one bet was placed during the last two hours preceding the start of the match for soccer and tennis, and during the last five minutes preceding the start of the race for horse racing. The aim was to determine a time interval reasonably long enough to encompass small fluctuations of odds around equilibrium, yet still short enough to screen out large changes of odds signaling that the market is not in equilibrium.[12] The different lengths of time intervals for soccer, tennis, and horse racing reflect the different microstructure of the markets in these sports. Due to the lower number of soccer and tennis markets, as well as the lower number of outcomes on these markets and longer time intervals between these events, the odds on soccer and tennis markets do not often exhibit large fluctuations before the start of the event.

The liquidity of Betfair markets varies tremendously, being as low as two bets with £4 volume matched to as high as 42,421 bets and £9,496,375 volume matched. Due to the lack of liquidity, I further restrict the analysis to those markets at which at least 20 bets have been placed on each outcome of the event. In the case of tennis and soccer matches the number of outcomes is given, yet for horse races the number of outcomes differs for each race. Thus, to ensure that all the outcomes of horse race events are accounted for, I ruled out those events where the sum of imputed probabilities was lower than 0.98 and considered only those events where the total

---

[11]I effectively treat these small fluctuations as if the market was already in equilibrium.

[12]I considered intervals in the range of 2 minutes - 10 hours before the start of the match. I chose the longest interval in which the average fluctuation of probability representation of odds (i.e. imputed probability equal to the inverse value of odds) was still lower than 3%.

number of outcomes (horses) was lower than or equal to 13.[13] All these steps restricted the analysis to 17,371 tennis match winner markets, 70,831 soccer match winner markets, and 59,386 horse race winner markets.

For further analysis of the risk preferences of marginal bettors, I use the *average bet size* computed as the volume matched over the number of bets from all odds at which at least one bet was made during the relevant time interval preceding the start of the event. The volume matched encompasses both the volume of market and limit matched orders on the back and lay side, and the number of bets is the sum of both back and lay bets. So, in fact, I use the average size of both back and lay bets. Average bet size varies remarkably with odds, suggesting that bettors bet different amounts on different odds, and justifying the importance of including the bet size in analysis. The average bet sizes for all three sports are presented in Figure 8 in the Appendix.

With the available data I am not able to distinguish between the average back bet size and the average lay bet size as I do not have information on the number of back or lay orders. Thus, in further estimation I assume that the computed average bet size corresponds to the average back bet size. In the Appendix I provide a mathematical proof that under plausible assumptions on the behavior of bettors this approach delivers reliable and correctly interpreted estimation results.

As pointed out before, a usual characteristic of betting market data is the so-called favorite-long shot bias. Smith, Paton, and Williams (2006) suggest that favorite-long shot bias should be lower on betting exchanges. My data are consonant with this, as they exhibit smaller long shot bias on horse race markets (see Figure 5). Still, Figures 6 and 7 in the Appendix show the presence of fairly strong long shot bias on the tennis and soccer match winner markets.

---

[13]Races with more than 13 horses account for less than 8% of the total number of races

Figure 5: Expected return per dollar bet on horse races at Betfair
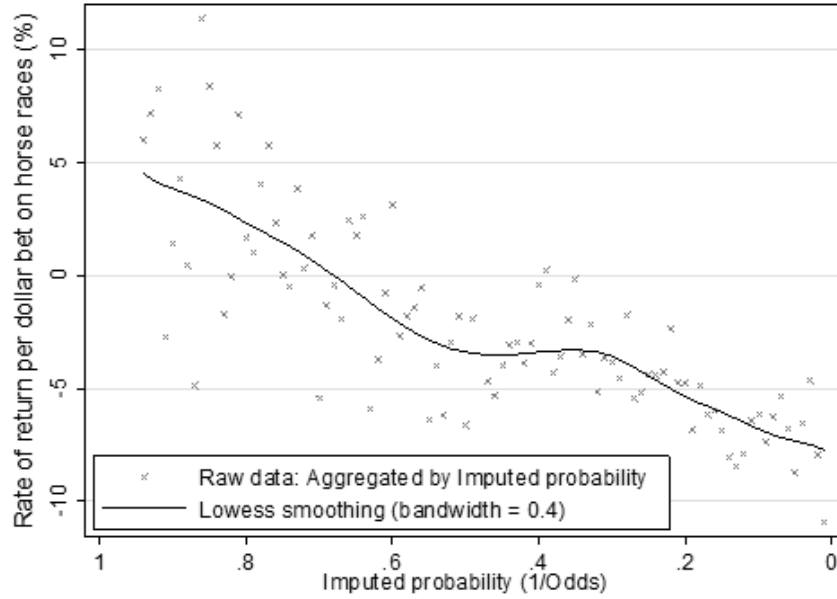
Note: number of observations used: 59,386 horse races.

5   Results

In the first part, I focused on the importance of accounting for bet size in the analysis of risk preferences of bettors. As discussed above, when we use just price data, the estimates are driven solely by the long shot bias. In Table 1 the estimates of risk aversion for bettors, assuming constant bet size, are presented. The results indicate that the marginal bettor has risk-loving preferences, a finding similar to that of Jullien and Salanie (2000).

Further, the estimated coefficient θB consists of both the parameter of risk aversion θ and the average bet size B, which is £20 for horse racing, £45 for soccer, and £107 for tennis. This implies that the estimates of the risk aversion parameter θ on different sports at Betfair are of comparable size, but all of them are significantly smaller than the estimates of Jullien and Salanie (2000). One reason may be the higher competition among bookmakers at Betfair markets, but also that as part of

Table 1: Estimates of risk aversion parameter of CARA utility function, assuming constant bet size.

Market	θB	S.D.	p-value	$95\%CI_{lower}$	$95\%CI_{upper}$
Tennis	-0.036	0.011	0.001	-0.058	-0.015
Soccer	-0.015	0.005	0.001	-0.024	-0.006
Horse races	-0.003	0.001	0.000	-0.004	-0.002

Note: number of observations used in the estimation: tennis - 17,371 obs., soccer - 70, 831 obs., horse races - 59,386 obs.

the data cleaning procedure I discarded all events with fewer than 20 bets on any of the outcomes and therefore screened out low liquidity markets, i.e. ones facing lower competition among bookmakers.

As explained in Section 2, bet size is key to the analysis of bettors' behavior, as bettors do not usually bet the same amount on different odds. Indeed, accounting for bet size dramatically changes the results for all sport types, as presented in Table 2. These differences between the markets on different sports raise questions about the appropriateness of EUT.

Table 2: Estimates of risk aversion parameter of CARA utility function, accounting for different bet size.

Market	θ	S.D.	p-value	$95\%CI_{lower}$	$95\%CI_{upper}$
Tennis	-0.0003	0.0002	0.225	-0.0009	0.0002
Soccer	0.0001	0.0001	0.222	-0.0001	0.0003
Horse races	-0.0005	0.0001	0.000	-0.0006	-0.0004

Note: number of observations used in the estimation: tennis - 17,371 obs., soccer - 70, 831 obs., horse races - 59,386 obs.

In the second step I test the key difference between EUT and behavioral theories, namely the assumption that bettors have a linear probability weighting function. If the EUT model of bettors' behavior is correct, we should obtain the same estimates of risk preferences over the whole range of probabilities. Therefore, I draw two types of subsamples from the data on each sport. The first type is a subsample

with favorites, where I condition the selection of events on the presence of a strong favorite. Due to the different number of outcomes in the particular sport[14], I include the event in the sample only if there are: (a) a tennis player with odds lower than 1.25 in tennis (imputed probability of winning greater than 80%); (b) a team with odds lower than 2.0 in soccer (imputed probability of winning greater than 50%); and (c) a horse with odds lower than 3.0 in the horse race (imputed probability of winning greater than 33%). I use the odds as a proxy for the objective probabilities of winning. The second type of subsample consists of events without any favorite, i.e. I include the event in the sample only if both players have odds greater than 1.5 for tennis (imputed probability of winning lower than 66%); if all outcomes have odds greater than 2.3 in soccer (imputed probability of winning lower than 43%); and if all horses in the race have odds greater than 4.0 in the horse races (imputed probability of winning lower than 25%). Under EUT, the risk preferences of the representative bettor should not differ regardless of whether he is betting on an event with a strong favorite or on an event without large differences in the winning probabilities of outcomes. Therefore, by comparing the results of the two types of subsamples I can easily test whether a marginal bettor has a linear weighting function of probabilities.

Table 3: Tennis markets - Estimates of risk aversion parameter of CARA utility function on subsamples defined by the presence of favorite, accounting for bet size

Market	θ	S.D.	p-value	$95\% CI_{lower}$	$95\% CI_{upper}$
All events	-0.0003	0.0002	0.225	-0.0009	0.0002
-with favorites	0.0004	0.0004	0.277	-0.0003	0.0011
- no favorites	-0.0013	0.0005	0.007	-0.0023	-0.0004

Note: number of observations used in the estimation: tennis - 17,371 obs., soccer - 70, 831 obs., horse races - 59,386 obs.

---

[14]There are two players for tennis, three outcomes for soccer and usually more than six outcomes for horse races leading to significant differences in the objective probabilities of winning between the outcomes in these sports.

Table 4: Estimates of risk aversion parameter of CARA utility function, accounting for different bet size.

Market	θ	S.D.	p-value	$95\% CI_{lower}$	$95\% CI_{upper}$
All events	0.0001	0.0001	0.222	-0.0001	0.0003
- with favorites	0.0003	0.0001	0.011	-0.0001	0.0005
- no favorites	-0.0004	0.0002	0.030	-0.0008	-0.0001

Note: number of observations used in the estimation: tennis - 17,371 obs., soccer - 70, 831 obs., horse races - 59,386 obs.

The results for tennis, soccer and horse races are presented in Tables 3-5. Estimates of the risk aversion parameter for the subsamples with a favorite and without a favorite are significantly different from each other for all three sports. I can therefore reject the null hypothesis of a linear probability weighting function in favor of its non-linear counterparts. Details of the estimation for particular sports are discussed below.

Results for tennis and soccer indicate that the ratio of the bets on outcomes with small differences in probabilities is higher than the ratio consistent with the behavior of risk-neutral bettors. This might suggest that people overweight small differences in probabilities. On the other hand, the opposite is true on markets with strong favorites, where the ratio of the amount placed on the more probable outcome to the amount placed on the less probable outcome is lower in comparison with risk-neutral bettors. This might suggest that people either underweight large differences in probabilities or simply underweight the large probabilities near the reference point 1. Another possible explanation is that bettors have restrictions on their maximum bet size; that is, when the model of risk-neutral bettors implies remarkably high bets for high probable outcomes, the maximum bet size may function as a binding constraint, resulting in a significantly lower bet ratio of bets on events with strong favorites than on events without any favorites. In both cases, however, I can reject the hypothesis that the marginal bettor at Betfair has a linear

weighting function of probabilities.

These results bring further insight to the theories of Tversky and Kahneman (1992). They assume that people underweight large probabilities and overweight small probabilities, i.e., that zero and certainty serve as reference points from which people offset their perception of probabilities. My results suggest that even particular outcomes serve each other as reference points, which leads to observed over-weighting of small differences in probabilities and underweighting of large differences in probabilities.

Table 5: Horse race markets - Estimates of risk aversion parameter of CARA utility function on subsamples defined by the presence of a favorite, accounting for bet size

Market	θ	S.D.	p-value	$95\%CI_{lower}$	$95\%CI_{upper}$
All events	-0.0005	0.0001	0.000	-0.0006	0.0004
- with favorites	-0.0008	0.0001	0.000	-0.0009	-0.0007
- no favorites	-0.0002	0.0001	0.109	-0.0004	0.0001

Note: number of observations used in the estimation: tennis - 17,371 obs., soccer - 70, 831 obs., horse races - 59,386 obs.

The results from horse racing markets also support the observation that bettors do not weight probabilities linearly. However, as suggested by the results in the first step, in the case of horse races the behavior of bettors seems to follow a different pattern than in tennis or soccer. Bettors still slightly overweight the small differences between probabilities of winning of horses in events without any strong favorite, yet they overweight the middle-sized differences in probabilities between underdogs and favorites even more. The rationale for this result lies in the higher number of outcomes on the horse race market and thus the lower absolute values of implied probabilities as well as their differences. In such a market structure, unlike the tennis and soccer markets, the implied probabilities never cross the threshold where the underweighting behavior of bettors prevails.

6 Conclusion

This paper makes several contributions to the literature on decision making under risk and uncertainty. Using an extensive dataset from the world's largest betting exchange, Betfair, I show that bettors bet different amounts on different odds, and that bet size is key to explaining their attitude towards risk. I abandon the assumption of constant bet size commonly used in the literature and provide corrected estimates of the risk preferences of bettors which, indeed, differ significantly from previous studies.

This research also has broader implications for the general analysis of behavior under uncertainty, particularly for discussions regarding the validity of EUT. My results suggest that, when facing a number of outcomes with different winning probabilities, bettors tend to overweight small and underweight large differences in probabilities, which is in direct contradiction to the linear probability weighting function implied by EUT. These findings can be presented as a refinement on Tversky and Kahneman (1992), who report the same behavior of agents with respect to absolute values of probabilities. My results also support the theory of reference points in decision making under uncertainty. However, they indicate that people may use more reference points than the generally accepted 0 and 1, as the outcomes might serve as each other's reference points.

References

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

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

Forrest, D., and I. McHale. 2007. "Anyone for tennis (betting)?" *European Journal of Finance* 13 (8): 751–768.

Gandhi, A. 2008. Rational expectations at the racetrack: Testing expected utility using prediction market prices. Manuscript Univ. Wisconsin-Madison.

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

Golec, J., and M. Tamarkin. 1998. "Bettors love skewness, not risk, at the horse track." *Journal of Political Economy* 106 (1): 205–225.

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

Hamid, S., A.J. Prakash, and M. W. Smyser. 1996. "Marginal risk aversion and preferences in a betting market." *Applied Economics* 28 (3): 371–376.

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

Kanto, A. J., G. Rosenqvist, and A. Suvas. 1992. "On utility function estimation of racetrack bettors." *Journal of Economic Psychology* 13 (3): 491–498.

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

Smith, M. A., D. Paton, and L. V. Williams. 2006. "Market efficiency in person-to-person betting." *Economica* 73 (292): 673–689.

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

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

Weitzman, M. 1965. "Utility analysis and group behavior: An empirical study." *Journal of Political Economy* 73 (1): 18–26.

Appendix

I assume that bookmakers and traders act as risk-neutral agents and that their orders are larger than the orders of common bettors.[15] On the other hand, common bettors who just choose the outcome mostly place back market orders. As the first two types of bettors are risk neutral, the estimates will be driven by the risk preferences of the common bettors, and will be biased towards risk neutrality. I assume that all common bettors are placing back bets, i.e., betting that a particular player will win, and that matching of the bets is done mostly by bookmakers who stand outside the model. Nevertheless, on real betting exchanges the common bettors can be observed on both sides of the market. In the further text I analyze the effect of the simplifying assumption on the validity of the results.

Let us assume that proportion m of all bets are backs and $1-m$ are lays. Given the total number of bets on a favorite (N_F) and an underdog (N_U) I can compute the corresponding number of backs $(B_F$ and $B_U)$ and lays $(L_F$ and $L_U)$ as

$$\#B_F = mN_F;\ L_F = (1-m)N_F$$

$$\#B_U = mN_U;\ L_U = (1-m)N_U.$$

Because there are only two players and I assume that odds O_F and O_U are fair, it holds that $O_F = \frac{O_U}{O_U-1}$ and $R_F = \frac{1}{R_U}$. Thus, I can express the cross-relations between the average back bet (B_F, B_U) and lay bet (L_F, L_U) on the favorite and

---

[15] On Betfair, the volume of a lay order is defined not as the liability of a lay bettor, but as his profit which equals the stake of the bettor on the back side of the trade.

the underdog, respectively, as

$$L_F = B_U(O_U - 1) \quad L_U = B_F(O_F - 1)$$

Total matched volumes on the favorite and the underdog (VOL_F, VOL_U) are equal to

$$
\begin{aligned}
VOL_F = VOL_{BF} + VOL_{LF} &= \#B_F B_F + \#L_F L_F = m N_F B_F + (1 - m) N_F L_F = \\
&= m N_F B_F + (1 - m) N_F B_U (O_U - 1) \\
VOL_U = VOL_{BU} + VOL_{LU} &= \#B_U B_U + \#L_U L_U = m N_U B_U + (1 - m) N_U L_U = \\
&= m N_U B_U + (1 - m) N_U B_F (O_F - 1)
\end{aligned}
$$

Solving for B_F and B_U gives

$$
\begin{aligned}
B_F &= \frac{(1 - m) N_F (O_U - 1) VOL_U - m N_U VOL_F}{N_F N_U (1 - 2m)} \\
B_U &= \frac{(1 - m) N_U (O_F - 1) VOL_F - m N_F VOL_U}{N_F N_U (1 - 2m)}
\end{aligned}
$$

I am interested in how the average back bet size changes with a different proportion of backing bettors on the market. Taking derivatives of B_F and B_U with respect to m I get

$$
\begin{aligned}
\frac{\partial B_F}{\partial m} &= \frac{N_F(O_U - 1)VOL_U - N_U VOL_F}{N_F N_U (1 - 2m)^2} \\
\frac{\partial B_U}{\partial m} &= \frac{(N_U(O_F - 1)VOL_F - N_F VOL_U}{N_F N_U (1 - 2m)^2} \\
\frac{\partial B_F}{\partial m} > 0 \Leftrightarrow (O_U - 1) &> \frac{N_U}{N_F} \frac{VOL_F}{VOL_U} = \frac{\frac{VOL_F}{N_F}}{\frac{VOL_U}{N_U}} = \frac{B_F^{comp}}{B_U^{comp}} \\
\frac{\partial B_U}{\partial m} > 0 \Leftrightarrow (O_F - 1) &> \frac{N_F}{N_U} \frac{VOL_U}{VOL_F} = \frac{\frac{VOL_U}{N_U}}{\frac{VOL_F}{N_F}} = \frac{B_U^{comp}}{B_F^{comp}}
\end{aligned}
$$

where $B_F^{comp} = VOL_F/N_F$ and $B_U^{comp} = VOL_U/N_U$ denote the average back bet

sizes under the assumption that m = 1, i.e., that all bettors are backing, which I used in my estimates. B_F and B_U are continuous on the range of $m \in (0.5; 1\rangle$. Therefore if $m > 0.5$ and the results of my estimation suggest that the bettors are risk averse, the following inequalities hold[16]:

$$\frac{B_F^{comp}}{B_U^{comp}} < \frac{p_F}{p_U} = \frac{1 - p_U}{p_U} = \quad (O_U - 1)$$

$$(O_U - 1) \quad > \frac{B_F^{comp}}{B_U^{comp}} \Rightarrow \frac{\partial B_F}{\partial m} > 0 \Rightarrow_{if\ m > 0.5} B_F < B_F^{comp}$$

$$\frac{B_U^{comp}}{B_F^{comp}} > \frac{p_U}{p_F} = \frac{1 - p_F}{p_F} = \quad (O_F - 1)$$

$$(O_F - 1) \quad < \frac{B_U^{comp}}{B_F^{comp}} \Rightarrow \frac{\partial B_U}{\partial m} < 0 \Rightarrow_{if\ m > 0.5} B_U > B_U^{comp}$$

Combining the fact that $B_F < B_F^{comp}$ and $B_U > B_U^{comp}$ results in inequality

$$\frac{B_F}{B_U} < \frac{B_F^{comp}}{B_U^{comp}} < \frac{p_F}{p_U}$$

This means that use of the right average back bet size would lead to an even higher risk aversion estimate. Similarly, if the results suggest that the marginal bettor is risk loving, I can reiterate the previous analysis as follows:

$$\frac{B_F^{comp}}{B_U^{comp}} > \frac{p_F}{p_U} = \frac{1 - p_U}{p_U} = \quad (O_U - 1)$$

$$(O_U - 1) \quad < \frac{B_F^{comp}}{B_U^{comp}} \Rightarrow \frac{\partial B_F}{\partial m} < 0 \Rightarrow_{if\ m > 0.5} B_F > B_F^{comp}$$

$$\frac{B_U^{comp}}{B_F^{comp}} < \frac{p_U}{p_F} = \frac{1 - p_F}{p_F} = \quad (O_F - 1)$$

$$(O_F - 1) \quad > \frac{B_U^{comp}}{B_F^{comp}} \Rightarrow \frac{\partial B_U}{\partial m} > 0 \Rightarrow_{if\ m > 0.5} B_U > B_U^{comp}$$

$$\frac{B_F}{B_U} > \frac{B_F^{comp}}{B_U^{comp}} > \frac{p_F}{p_U}$$

---

[16]Within the utilized CARA utility framework, the ratio of bets of a risk-neutral bettor satisfies the condition $B_F/B_U = p_F/p_U$.

In both cases, use of average betting size computed under the assumption that m = 1 biases the results towards risk-neutral preferences. Thus, as long as the real proportion of common bettors on the back side of the market is higher than 0.5, it is reasonable to conclude that my estimate of risk aversion/risk love is a lower/upper bound of a real value.

I have also performed an empirical check of my assumptions through the analysis of bets on 60 markets of the 2006 soccer World Cup for which I have available information on the number of back and lay bets. According to this data, the share of "backers" on the market orders is larger than the share of "layers". The share of back bets ranges from 60% to 90% with an average 73% share of all observed bets for 180 outcomes (3 outcomes per market) of match winner markets, and ranges from 60% to 96% with an average 86% share of all observed bets for 1020 outcomes (17 outcomes per market) of the correct score markets. The average lay bet sizes are always remarkably larger than the average back bet sizes.

Figure 6: Expected return per dollar bet on tennis at Betfair
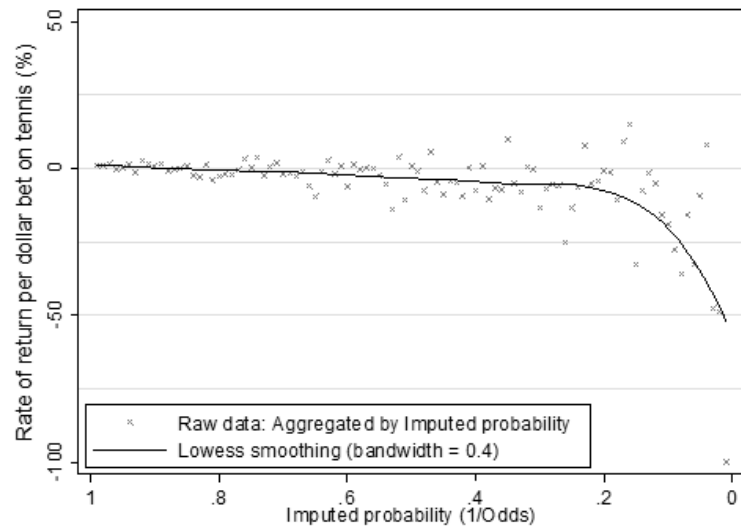
Note: number of observations used: 17,371.

Figure 7: Expected return per dollar bet on soccer at Betfair
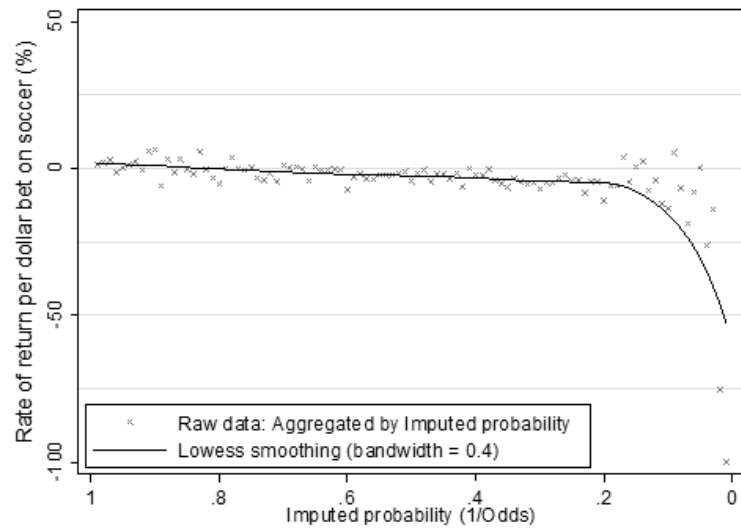
Note: number of observations used: 70,831.

Figure 8: Average bet size at Betfair with respect to imputed probability
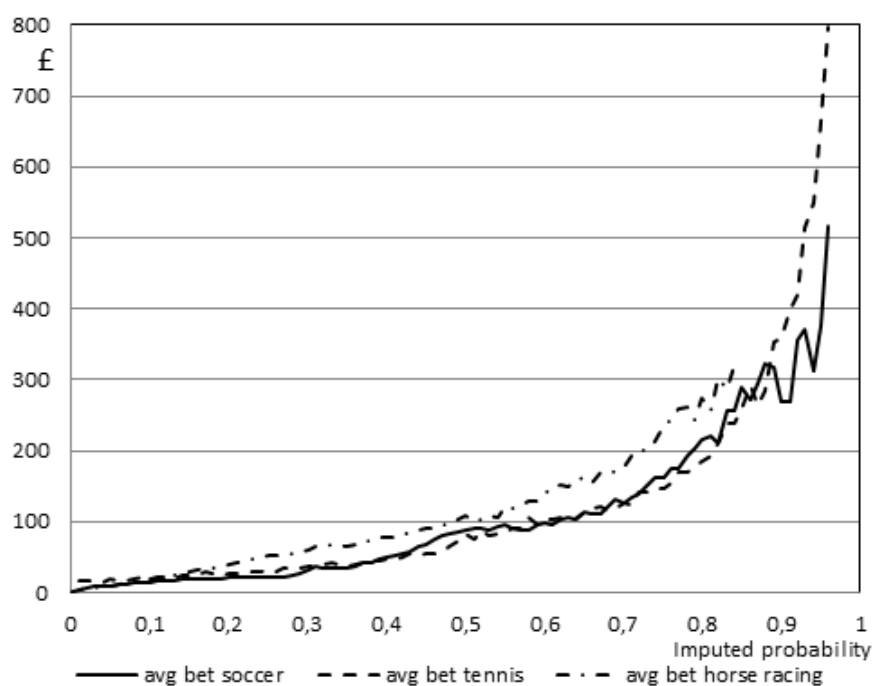
Note: number of observations used: tennis - 17, 371 obs., soccer - 70, 831 obs., horse races - 59, 386 obs.