Risk Aversion Estimates from Betting Markets*

František Kopřiva[†]
CERGE–EI[‡]

July 2009

Abstract

I analyze the risk preferences of bettors using data from the world largest betting exchange Betfair. The assumption of a constant bet size, commonly used in the current literature, would lead to an unrealistic model of bettor's decision making as a choice between high return - low variance and low return - high variance bet, automatically implying risk loving preferences of bettor. 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 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 about the average bet size to test the validity of EUT theory. 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.

*The author would like to thank Eva Hromadkova for helpful comments and editing of the paper. The financial support of the Grant Agency of the Charles University, Grant No. 34509 is gratefully acknowledged. The views expressed are those of the author and do not necessarily reflect the position of any of the affiliated institutions.

[†]Email: Frantisek.Kopriva@gmail.com

[‡]A joint workplace of the Center for Economic Research and Graduate Education, Charles University, Prague, and the Economics Institute of the Academy of Sciences of the Czech Republic. Address: CERGE–EI, P.O. Box 882, Politických vězňů 7, Prague 1, 111 21, Czech Republic

1 Introduction

Validity of behavioral theories of decision-making under uncertainty that assume the non-linear probability weighting functions, i.e. Rank-dependent expected utility theory (RDEU) and Cumulative Prospect theory (CPT), has been recently tested in numerous studies. Most of these studies are experimental; however, there exists an innovative strand of empirical literature on this topic that analyzes the price data from betting markets. Nevertheless, none of these empirical studies were using the information on the actual size of bet that bettor chooses facing particular odds, though it is actually one of his key decisions. In this paper, I use data from the world largest betting exchange Betfair to compute the bet sizes chosen by an average bettor at particular odds and I use this information to test the validity of Expected utility theory (EUT) versus its alternatives with non-linear probability weighting functions.

The choice of betting as the framework for the analysis of general risk attitudes of agents is natural, because it stands out as one of the few real world situations ideal for the analysis of decision-making under risk and uncertainty due to its specific feature - repeated occurrence of a large number of similar events. Betting markets are in literature also often compared to financial markets, although the opponents stress that in contrast to them the betting markets have almost nonexistent dynamics, and if there is some (e.g in-play betting markets) it is usually easy to analyze. However, the comparison is more than suitable for Betfair - due to the stunning betting volume and a design similar to the order-driven stock exchanges, Betfair's market microstructure resembles real financial markets.

To the best of my knowledge, all the previous studies with exception of Bradley (2003) estimate risk preferences at betting markets assuming that bettors place the same amount of money on different odds (i.e. that the bet size is constant irrespectively of the probability of the outcome). However, in reality bettors differentiate their bet size significantly with respect to the odds they face, which is also confirmed in our data. Also, from the theoretical point of view, the choice of bet size is actually the only stage in the decision-making process of bettor when he has a completely free choice, and is therefore the most informative of his risk preferences. It is very unlikely that, given the odds are fair, bettor will decide to bet the same amount on outcomes with significantly different probabilities of winning and, if he would, this would suggest that he is strongly risk averse which is in direct opposition to the results of previous studies dealing with the behavior of bettors.

The theoretical fallacy of the "constant bet size" approach to the estimation is further inflated by the specific properties of data used. Odds on betting markets often exhibit so called favorite longshot bias: usually, the bets on low probability events provide a lower expected return compared to the bets on high probability events. By using just the price data (odds) and implicitly assuming same bet size for both high and low probability events, the previous studies ended up with the analysis of situation where the bettor is indifferent between betting on low probability event with lower return and higher variance and betting on high probability event with higher return and lower variance.[1] Such a scenario unavoidably implies the conclusion that the average bettor is risk loving under EUT. On the other hand, after accounting for the bet size the situation translates into the choice familiar from the standard investment decision-making process where the representative bettor is indifferent between betting on low probability event with lower return and lower variance (given the average bet size on low probability events) and high probability event with higher variance (again, given the average bet size on high probability events).

[1]See Weitzman (1965), Ali (1977), Kanto et al. (1992), Hamid et al. (1996), Golec and Tamarkin (1998), Jullien and Salanie (2000).

Use of EUT implicitly introduces into estimation process the assumption that bettors weight probabilities linearly. This assumption have been questioned by above mentioned non-expected utility models like RDEU and CPT. The key assumption of these models is that people do not necessarily assign the probabilities linear weights and have their own particular weighting function (e.g. inverse-S-shaped weighting function). The rich dataset with information about the average bet size that covers the whole range of winning probabilities allows me to test validity of the EUT linear weighting function of bettors without any specific definition of the alternative. I follow the methodology of Golec and Tamarkin (1998) and draw two conditioned subsamples based on the occurrence of favorite in the event, using odds as a proxy for the objective probabilities of winning of particular outcomes. I thus create two subsamples with different composition of probabilities of winning of particular outcomes, on which I further test whether people assign different weights to different probabilities.

The paper is structured as follows: section 2 critically summarizes the existing literature on the risk preferences of bettors; section 3 outlines the methodology and estimation strategy. The data description is provided in section 4, section 5 presents and discusses the results and section 6 concludes the study.

2 Literature Review

There are several studies that estimate the risk preferences of agents using data from betting markets - mainly horse or greyhound racing. These papers rely on the assumption of a representative bettor, i.e. they are estimating the preferences of an average or marginal bettor. The early studies on this topic treat all events (races) as identical and thus are able to group them in different ways, e.g. by odds intervals or position of horse. Assuming that the bettors maximize their expected utility, Ali (1977) and Kanto et al. (1992) conclude that racetrack bettors are risk loving.

Golec and Tamarkin (1998) point out that all previous studies have considered just mean and variance (risk) of the bet. Their results suggest that when one accounts also for the skewness of the return distribution, the representative bettor turns out to be risk averse but skewness-loving. They also try to address the problem of racetrack betting data which consist of relatively few favorites (high probability results) compared to number of underdogs by conditioning their sample on the races with high probability winning horses.

Further advance in the field was introduced by Jullien and Salanie (2000) who designed a new methodology of estimating the preferences of a representative bettor without any grouping of data, because ,as they argue, the betting behavior may differ with different characteristics of the particular racetracks event. Using the data from UK horse betting markets they compare EUT, RDEU and CPT by assuming particular functional forms of utility function and weighting function. Under EUT they conclude that the representative bettor is risk loving. Nevertheless, they conclude that the CPT fits the data better and that the utility function of representative agent exhibits risk aversion over the losses and mild risk loving over gains.

Ghandi (2008) relaxes the assumption of representative bettor and assumes that the betting market is a pool of heterogeneous agents according to which horse they decide to bet on. He shows that the diversity of bettors sufficiently explains the usually observed price - return scenario. Further, he points out that EUT better explains the data as the estimated weighting function of behavioral alternatives actually coincides with the linear function. Therefore, he concludes that EUT outperforms its behavioral alternatives. He points out that the people betting on lowest odds are the most risk averse bettors. However, our data suggest that more risk averse bettors rather choose to bet low amounts on the underdogs.

The only study that does not neglect the bet size in the analysis of behavior of bettors is Bradley (2003). Nevertheless, as he does not have data on the bet size he performs his analysis by calculating what would be the optimal bet size of a representative bettor assuming CPT. By assuming that the only utility that a bettor has from bet is derived from expected return and variance he computes the optimal bet as an argument of the maximum weighted expected utility given the probabilities and odds. However, this methodology and presence of a favorite longshot bias in his data would ex-ante imply risk loving preferences of bettors under EUT. As Bradley (2003) is not able to separate the estimates of parameters of utility function and weighting function, he concludes that if the weighting function is linear the agent would be risk loving for gains and risk averse for losses.

All the above-mentioned studies analyze the risk preferences of bettors on horse racing markets. However, as already mentioned, the horse racing events usually consist of large number of horses with low probability of winning and low number of favorites. This may lead to a situation where we would estimate the risk preferences of bettors just over the bets with low probability of winning. As pointed out by Forrest and McHale (2007), the tennis betting markets possess a nice feature of having nearly complete distribution of events and bets over the whole probability range. Therefore, by using the data from tennis and soccer markets I can analyze the behavior of bettors over the complete set of probabilities.

3 Methodology

Contrary to the classic betting markets, the betting exchanges are much more similar to financial markets. The Betfair betting exchange is designed as an order driven market where the bettors can place either limit orders or market orders on a given outcome. When placing a limit order, the bettor has to decide whether to place back or lay bet and he has to stipulate the odds and volume of the bet.

On the other hand, when placing a market order bettor hits the odds already available on the market and chooses just the volume and side of the market. Every market on the exchange (for example on the winner of tennis, soccer match or horse race) consists of several outcomes with ex-ante objective probabilities of winning $p_1, \ldots, p_N$. Assume that the bettor decides to place a back bet (the outcome will happen) on the outcome 1 of volume one dollar at odds O_1. If the outcome 1 occurs, the bet brings profit $R_1 = O_1 - 1$ with probability p_1 and if the outcome 1 does not occur the bettor will lose one dollar. If the bettor places a lay bet (the outcome will not happen) on the outcome 1 of volume one dollar at odds O_1, the bet brings profit one dollar if the outcome 1 does not occur and loss $-R_1$ if the outcome 1 occurs. As we are analyzing the markets with one possible winner only, the probabilities $p_1, \ldots, p_N$ sum up to one. Thus, backing an outcome at odds O_1 is actually the same as laying all the other outcomes at respective odds $O_2, \ldots, O_N$ which can be also represented as a lay bet at compound odds $O_c = 1/(\frac{1}{O_2} + \cdots + \frac{1}{O_N}) = O_1/(O_1 - 1)$. Obviously, the conversion between lay and back odds is far less computationally demanding when there are just two or three outcomes.

I assume that on Betfair there exist three main types of bettors - bookmakers, traders and common bettors. The first type is endogenously emerged bookmakers who post large volume limit orders and just occasionally use market orders to balance their portfolios.[2] I further assume that these bettors are risk neutral as they try to balance their liabilities and earn profit from spread. Also, market odds are an outcome of competition between such bettors. The second type of bettors - traders place both limit and market orders, open and close their positions and earn money from the differences of the asset price during the time. These bettors are therefore also mostly placing large volume orders and the size of their bets is

[2]By analyzing the in-trade soccer markets, Gil and Levitt (2007) point out that the endogenously emerged market makers were on the one side of the trade for 65 percent when the markets were inplay.

balanced with respect to the odds, i.e., they are also acting as risk neutral bettor.

In the analysis of risk preferences, I am mostly interested in the behavior of the third type of bettors - common bettors, as this represents the risk attitude of the general population of bettors most accurately. I assume that a common bettor typically chooses just one outcome on which to bet and places mostly back market orders. I discuss the validity of this assumption and its implications for the results in Data section and then in more details in the Appendix.

My model follows the empirical methodology of Jullien and Salanie (2000). I represent the behavior of common bettors as a behavior of one representative agent - the average bettor. The average bettor is at first facing a decision whether to bet at all on a given market and then, given the odds, he chooses the amount to bet on a particular outcome. As my concern is not to analyze which markets attract the most bettors but rather to study the choices of the bettors on the market, I avoid modeling why people bet at all and focus on the comparison of choices of combinations of odds and bet sizes of the average bettor. I assume that he is aware of the objective probabilities of winning for all outcomes. Furthermore, under EUT for every two outcomes i, j on the market with given odds O_i and O_j and probabilities p_i and p_j, there exist the combination of bet sizes B_i and B_j such that average bettor with wealth M is indifferent between betting on these two outcomes, such that

$$p_i u(M+B_iR_i,\theta)+(1-p_i)u(M-B_i,\theta) = p_j u(M+B_jR_j,\theta)+(1-p_j)u(M-B_j,\theta).$$

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

$$p_i = \frac{1}{u(M+B_iR_i,\theta)-u(M-B_i,\theta)}\left(\frac{1+\sum_{j=1}^{N}\frac{u(M-B_j,\theta)}{u(M+B_jR_j,\theta)-u(M-B_j,\theta)}}{\sum_{j=1}^{N}\frac{1}{u(M+B_jR_j,\theta)-u(M-B_j,\theta)}} - u(M-B_i,\theta)\right)$$

Betfair is charging a fee of 2-5 % from net winnings, with lower fees for highly active and large customers. As my concern is to estimate the risk preferences of a representative common bettor, I have decided to include the 5% tax into my estimation procedure as the volume needed for lower percentage fees are quite large. Therefore, I decrease all the returns R_i by 5%.

As I do not observe any information about the wealth or income of average bettor, I am using the CARA utility function in the form $u(x, \theta) = \dfrac{1 - e^{-\theta x}}{\theta}$; otherwise my parameter estimates would be based either on the arbitral choice of the wealth M or the number of parameters will increase. The model is then estimated by Maximum Likelihood Estimation using the formula of probability of the winning outcome. The likelihood function is computed as a sum of logs of probabilities for ex-post winners from each match.

One of the key assumptions of alternative behavioral theories of decision-making under uncertainty (RDEU and CPT) is that probabilities enter the formula of expected utility in a non-linear form. In other words, the bettors have some non-linear probability weighting function. We can, however, test the validity of EUT without explicitly formalizing the alternative theories. If the assumption of a linear weighting function of EUT is right then the estimated parameters of risk aversion of average bettor should be the same no matter in which interval of probability range 0-1 are the objective probabilities of players/teams/horses. Therefore, I draw two subsamples - the one with strong favorites and the other without any favorite - separately for each sport. Under null hypotheses the EUT theory holds and therefore the estimates on the subsamples should not be statistically different from each other. If the results differ we can reject EUT in favor of theories with the non-linear weighting functions of probabilities.

4 Data

I am using aggregated historical data from the world's largest betting exchange Betfair for all tennis, soccer and horse racing winners markets between 2004 and 2008. For each outcome of every market and for each odds at which at least one bet was placed the data consist of 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 and a winning outcome. Although on Betfair one can also place bets during the matches, I am using data only on the bets that were placed before the start of the match or race, as I want to analyze the ex-ante attitude towards the risk rather than the reaction of bettor on the news from the ongoing match.

Recent studies dealing with the risk attitude of bettors (Jullien and Sallanie 2000, Ghandi 2008, etc.) were using the starting prices - the final odds at the start of the event from pari-mutuel markets or the final odds of bookmakers. However, there are always two final odds (back and lay) at betting exchange markets. Further, as the odds tend to fluctuate even before the start of the event, the final odds do not usually reflect the overall information about the market in equilibrium. Therefore, I decided to take the weighted average (by total volume matched) of odds at which some bets were 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 find a reasonably long time interval to encompass odds fluctuating around equilibrium, which would still be short enough to screen out large changes of odds signalizing that the market is not in equilibrium.[3] The different time lengths between soccer, tennis and horse racing reflect the different nature of the markets in these sports. The liquidity of Betfair markets varies

[3] I considered intervals in the range of 2 minutes - 10 hours interval before the start of the match and then I chose those intervals satisfying max 3% average range of imputed probabilities from odds.

tremendously, being as low as two bets with total volume matched £4 to as high as 42,421 bets and £9,496,375 volume matched. Due to the lack of the liquidity of some markets, I further restrict my analysis only on those matches with at least 20 bets placed on each outcome of the event. In case of horse racing, to assure that I have all the outcomes of an event I ruled out those events where the sum of imputed probabilities was lower than 0.98 and considered only events where the total number of outcomes (horses) was lower than 13.[4] All these steps restricted my analysis on 17,371 tennis match winner markets, 70,831 soccer match winner markets and 59,386 horse race winner markets.

For the further analysis of risk preferences of average bettor I use the average bet size computed as the volume matched over the number of bets from all odds at which some trade was done during the relevant time interval preceding the start of the event. The volume matched consists of the volume of both market and limit matched orders on the back and lay side, and number of bets is a sum of both back and lay matched orders. So, in fact, I use the average size of both back and lay matched orders. This average matched order size varies remarkably with odds, suggesting that bettors bet different amount on different odds, justifying the importance of including the bet size in analysis. The average bet size for all three sports are presented in Figure1.

However, 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 average back bet size is the same as the computed average bet size. Although this is a rather simplified approach, I impose several assumptions supported by empirical evidence that ensure that results of the estimation are reliable and correctly interpreted.

Concerning the behavior of the three types of bettors, I assume that bookmakers

[4]The races with more than 13 horses account for less than 8% of total number of races.

and traders are mostly acting as risk neutral agents and their orders are much larger in magnitude then the orders of common bettors. The bookmakers are balancing their liabilities and therefore act like a risk neutral bettor in our model. Similarly, the traders who are trying to make profit on the small changes of the odds during the time would most likely post either balanced bets on different outcomes or back and lay bets of similar size on one particular outcome and therefore also act like a risk neutral bettor.[5] On the other hand, the common bettors who just choose the outcome place mostly the back market orders.[6] As the two types of bettors are risk neutral, the estimates will be driven by risk preferences of the common bettor, and will be biased towards risk neutrality.

I performed the empirical check of my assumptions on analysis of market orders on 60 markets of 2006 soccer World Cup, where I could derive information on the number of backs and lays. According to these data, the share of "backers" on the market orders is larger than share of "layers". The share of back market orders ranges from 60% to 90% with an average 73% share of observed orders for 180 outcomes (3 outcomes per market) of match winner markets and ranges from 60% to 96% with an average 86% share of observed orders for 1020 outcomes (17 outcomes per market) of the correct score markets. Also, the average market lay orders are always remarkably higher then the average market back orders.

As I pointed out before, the usual characteristics of betting market data is so called favorite longshot bias. Smith et al. (2006) suggest that the favorite longshot bias should be lower on betting exchanges. Our data are in consonance with this as we do not see any statistically significant evidence of long shot bias on graphs of expected returns. Still, Figures 2 and 3 show, that when I plot the expected return against the imputed probability, that the odds on underdogs might be less

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

[6] In the Appendix I present formal proof that for the validity of results it is enough when the ratio of common bettors placing lay bets is smaller than 0.5.

favorable than odds on favorites for tennis and soccer match winner markets. So, by neglecting the bet size we would have similar problem as in Jullien and Salanie (2000), where according to the model the bettor is, given his preferences, indifferent between betting the same amount a on a high probability (low risk) event with higher return and low probability (high risk) event with lower return.

5 Results

At first I focus on the importance of accounting for the bet size in the analysis of risk preferences of bettors. As already discussed above, when one uses just the price data the estimates are driven mainly by the longshot bias. As the odds in our data are usually the outcome of a competition between endogenously emerged bookmakers, the bias is lower and not statistically significant when depicted only by odds or the imputed probability of winning (Figure 2 - 4). In the Table 1 I present the estimates of risk aversion for bettors assuming the constant bet size. The results are indicating risk loving preferences of an average bettor, which is a finding similar to the one of Jullien and Salanie (2000). In other words the results confirm that the favorite longshot bias is present, though in weaker magnitudes, also at all three types of markets at Betfair.

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

$Market$	θb	$Std.dev.$	$p-value$	95% CI_{lower}	95% CI_{upper}
Tennis	$-.036$	.0109	0.001	$-.0576$	$-.0150$
Soccer	$-.015$	.0046	0.001	$-.0244$	$-.0063$
Horse racing	$-.003$	.0006	0.000	$-.0042$	$-.0017$

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

Further, the estimated coefficient θb consist of both the parameter of risk aversion θ and the average bet b, which I have computed to be on average £20 for horse racing, £45 for soccer and £107 for tennis. This implies that my estimates of risk aversion parameter θ on different sports at Betfair are of comparable size, but all of them significantly smaller than the estimates of Jullien and Salanie (2000). One of the reasons for that is a higher competition of bookmakers at Betfair markets, but also the fact that in my procedure I discard all the events with less than 20 bets on any of the outcomes and therefore screen out low liquidity markets, i.e. those facing a lower competition between bookmakers.

Based on the facts presented above, the bet size is key information for the analysis of the behavior of bettors, as they do not usually bet the same amount on different odds. In the Table 2 I present the results of the estimation of the risk preference parameter of CARA utility function under EUT, estimated using the information about the bet size, on three types of Betfair markets - tennis, soccer and horse races. Contrary to the previous results, after accounting for the bet size the estimates for tennis and soccer markets are not statistically different from zero, suggesting that under EUT the average bettor on tennis and soccer markets is risk neutral.

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

$Market$	θ	$Std.dev.$	$p-value$	$95\%\ CI_{lower}$	$95\%\ CI_{upper}$
Tennis	$-.0003$	.0002	0.225	$-.0009$	.0002
Soccer	.0001	.0001	0.222	$-.0001$	$-.0003$
Horse racing	$-.0005$	.0001	0.000	$-.0006$	$-.0004$

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

Quite intriguing are the results for horse races where the estimated parameter implies even stronger risk loving preferences of the average bettor than when using just the price data. In other words, ratio of the amount placed on the more probable outcome to the amount placed on less probable outcome is higher than the corresponding ratio for risk neutral bettor. I provide closer insight on this behavior further in text when testing the validity EUT. Overall, there are significant differences in the estimates compared to the constant bet size model suggesting risk neutral behavior of bettors for soccer and tennis. Moreover, the differences between the markets on different sports on Betfair already here raise the question about the appropriateness of EUT.

In the second step I test the key difference of EUT in comparison to RDEU and CPT, namely that EUT assumes the bettors to have a linear probability weighting function. If the EUT model of the behavior of bettors is correct, I should obtain the same estimates of risk preferences on the whole range of probabilities. In order to show whether this is true, I draw two types of subsamples from data on each sport. The first type is a subsample with favorites where I condition the selection of events based on the presence of a strong favorite. Due to the different number of outcomes in the particular sport[7] I include the event into the sample only if there exist: a tennis player with odds lower than 1.25 in tennis (imputed probability of winning greater than 80%), a team with odds lower than 2.0 in soccer (imputed probability of winning greater than 50%) and a horse with odds lower than 3.0 in the horse race (imputed probability of winning greater than 33%). I use the odds as proxy for the objective probabilities of winning. The second type of subsamples consists of events without any favorite, i.e. I include the event into 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

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

odds greater than 4.0 in the horse races (imputed probability of winning lower than 25%). Under EUT, the risk preferences of representative bettor should not differ whether he is betting on the event with strong favorite or on the event without any high probable winning outcome. Therefore, by comparing the results between the two types of subsamples I can easily test whether average bettor has a linear weighting function of probabilities.

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

$Market$	θ	$Std.dev.$	$p-value$	$95\%\ CI_{lower}$	$95\%\ CI_{upper}$
All events	$-.0003$	.0003	0.225	$-.0009$	.0002
- with favorites	.0004	.0004	0.277	$-.0003$	.0011
- no favorites	$-.0013$	.0005	0.007	$-.0023$	$-.0004$

Note: number of observations used in the estimation: all events - $17,371$ obs., with favorites - $4,101$ obs., no favorites - $7,787$ obs.

Table 4: Soccer markets - Estimates of risk aversion parameter of CARA utility function on the subsamples defined by the presence of favorite, accounting for the bet size.

$Market$	θ	$Std.dev.$	$p-value$	$95\%\ CI_{lower}$	$95\%\ CI_{upper}$
All events	$-.0001$	.0001	0.222	$-.0001$	.0003
- with favorites	.0003	.0001	0.011	.0001	.0005
- no favorites	$-.0004$	.0002	0.030	$-.0008$	$-.0001$

Note: number of observations used in the estimation: all events - $70,831$ obs., with favorites - $31,287$ obs., no favorites - $23,162$ obs.

The results for tennis, soccer and horse races are presented in Tables 3 - 5. As you can see, for all three sports the estimates of risk aversion parameter for the subsamples with favorite and without favorite are significantly different from each other. Therefore, I can reject the null hypothesis of linear probability weighting function in favor of its non-linear counterparts. In the further discussion I will focus on a details of estimation for particular sports.

According to the results the bettors on the tennis and soccer markets bet much more on the slightly more probable outcomes than on the slightly less probable outcomes, so that the ratio of the amount placed on the more probable outcome to the amount placed on less probable outcome is higher than the corresponding ratio for the risk neutral bettor. This might suggest that people are overweighting small differences in probabilities, and that the outcomes serve to each other as some kind of reference point, similarly to 0-1 reference points in Tversky and Kahneman (1992). 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 less probable outcome is lower in comparison with risk neutral bettor. This might suggest that people either underweight large differences in probabilities or simply underweight the large probabilities near the reference point 1. Also, they might be restricted in the amount available for betting as the model of risk neutral bettor implies remarkably high bet amounts for the high probable winning outcomes. Nevertheless, in both cases we can reject the hypothesis that the average bettor at Betfair has a linear weighting function of probabilities. Even though the estimates might be biased due to the lack of information on the number of backs and lays, in the Data and Appendix section I show that for soccer and tennis the results would be even more significant.

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

$Market$	θ	$Std.dev.$	$p-value$	95% CI_{lower}	95% CI_{upper}
All events	$-.0005$	.0001	0.000	$-.0006$	$-.0004$
- with favorites	$-.0008$	.0001	0.000	$-.0009$	$-.0007$
- no favorites	$-.0002$	.0001	0.109	$-.0004$	.0001

Note: number of observations used in the estimation: all events - $59,386$ obs., with favorites - $27,516$ obs., no favorites - $14,359$ obs.

Also the results from horse racing markets support our observation that bettors do not weight probabilities linearly. However, as already implied by the results in the first step, in the case of horse races the behavior of bettors seems to follow a different pattern compared to tennis and soccer. They still slightly overweight the small differences between probabilities of winning of horses in events without any strong favorite. Nevertheless, they overweight the middle sized differences in probabilities between underdogs and favorites even more. The rationale for this result, which actually explains the findings concerning the risk loving preferences of horse race bettors from the first step, lies in the higher number of outcomes on the horse race market and thus lower absolute values of implied probabilities as well as their differences. In such market structure, contrary to the tennis and soccer markets, the implied probabilities never cross the threshold where the underweighting behavior of bettors prevails.

6 Conclusion

In this paper I make several contributions to the field of literature that analyzes the risk preferences of bettors. Using the extensive dataset from the world largest betting exchange Betfair I show that bettors bet different amounts on different odds and that the bet size is the key information about their behavior under risk. Based on this, I abandon the assumption of the constant bet size commonly used in the literature and provide the corrected estimates of the risk preferences of bettors, which, indeed, differ significantly from the previous studies.

However, this research has also broader implications in the context of general analysis of the behavior under uncertainty, particularly the discussion of appropriateness of Expected Utility Theory (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 might use more reference points than just generally understood 0 and 1, as the outcomes might serve each other as reference points.

References

[1] Ali, M.M. (1977). Probability and Utility Estimates for Racetrack Bettors. *The Journal of Political Economy*, 85, 803-815.

[2] Bradley, I. (2003). The representative bettor, bet size, and prospect theory. *Economics Letters*, 78, 409-413.

[3] Forrest, D., and McHale, I. (2007). Anyone for Tennis (Betting)? *The European Journal of Finance*, 13, 751-768.

[4] Gandhi, A. (2008). Rational Expectations at the Racetrack: Testing Expected Utility Using Prediction Market Prices. mimeo

[5] Gil, R., Levitt, S. D. (2007). Testing the Efficiency of Markets in the 2002 World Cup. *The Journal of Prediction Markets*, 1, 255-270.

[6] Golec, J., and Tamarkin, M. (1998). Bettors Love Skewness, Not Risk, at the Horse Track. *Journal of Political Economy*, 106, 205-225.

[7] Hamid, S. S., Prakash, A. J. and Smyser, M. W. (1996). Marginal Risk Aversion and Preferences in a Betting Market. *Applied Economics*, 28, 371-376.

[8] Jullien, B., and Salanie, B. (2000). Estimating Preferences under Risk: The Case of Racetrack Bettors. *Journal of Political Economy*, 108, 503-530.

[9] Kanto, A.J., Rosenqvist, G. and Suvas, A. (1992). On Utility Function Estimation of Racetrack Bettors. *Journal of Economic Psychology*, 13, 491-498.

[10] Smith, M. A., Paton D., Williams, L. V. (2006). Market Efficiency in Person-to-Person Betting. *Economica*, 73, 673-689

[11] Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty*, 5, 297-323.

[12] Weitzman, M. (1965). Utility Analysis and Group Behavior: An Empirical Study. *Journal of Political Economy*, 73, 18-26.

Figures

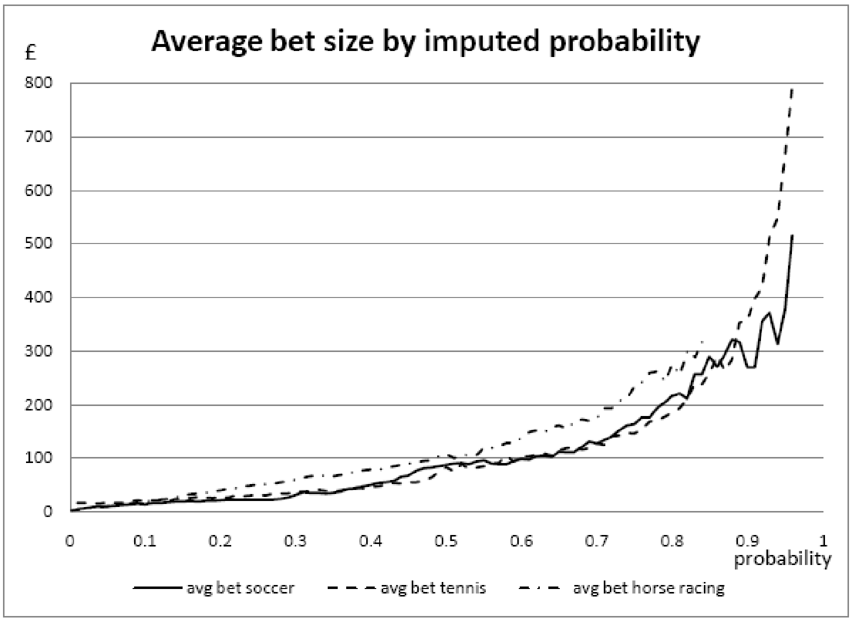

Figure 1: Average bet size (in £) given the imputed probability (1/odds) for all three sports markets.

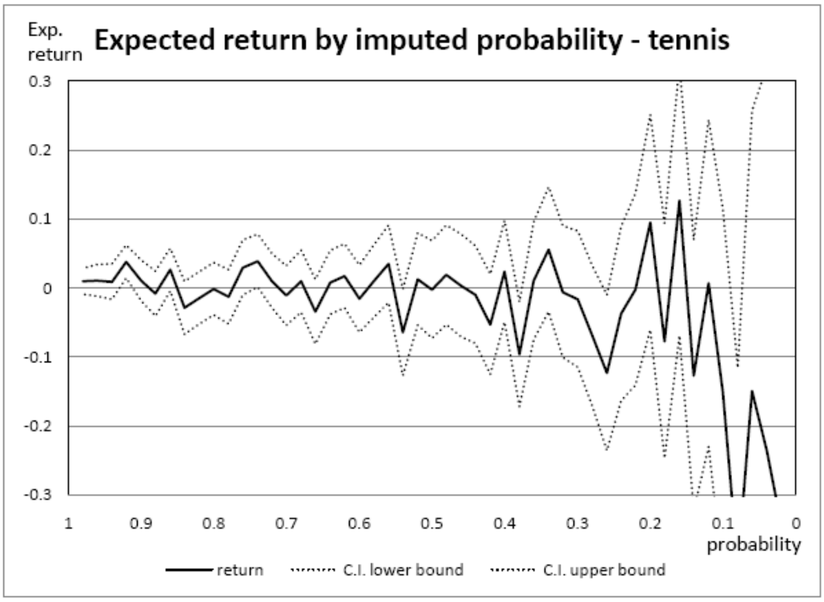

Figure 2: Expected return on given the imputed probability (1/odds) on tennis markets, with 95% confidence interval.

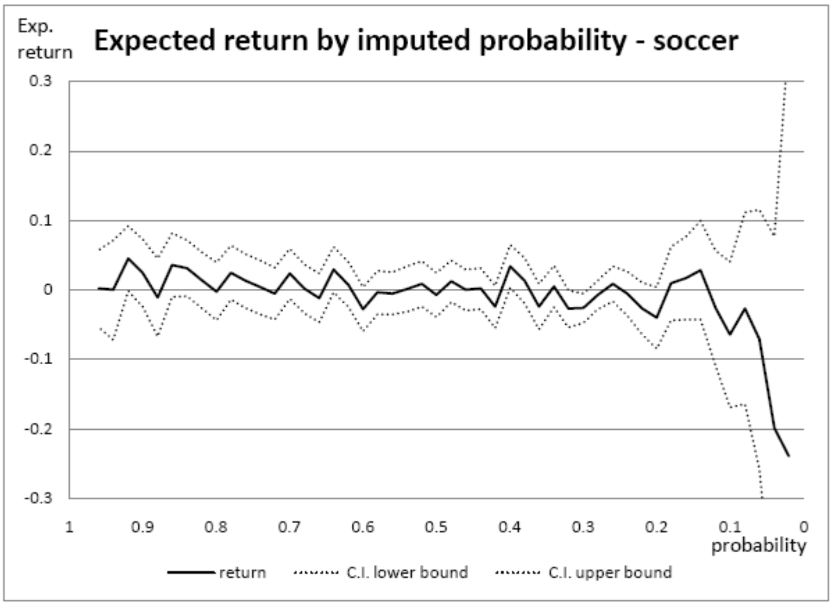

Figure 3: Expected return on given the imputed probability (1/odds) on soccer markets, with 95% confidence interval.

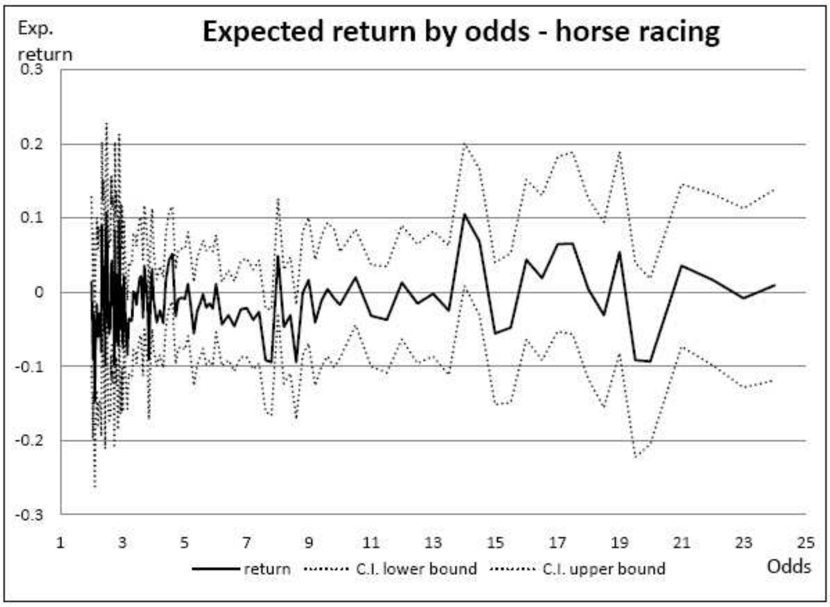

Figure 4: Expected return on given the odds on horse racing markets, with 95% confidence interval.

Appendix

In the model that is underlying my estimations I assume that all bettors are placing back bets, i.e. betting that a particular player will win, and that accepting of the bets is done mostly by bookmakers that stand outside the model. One could argue that this assumption is oversimplifying and that on real betting exchanges we can observe regular bettors on the both sides of the market. In my further analysis I show that not accounting for this fact biases the estimates towards risk neutral preferences and thus, the differences between subsamples would be even more significant.

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) we can compute corresponding number of backs (B_F and B_U) and lays (L_F and L_U) as

$$\begin{aligned} \#B_F &= mN_F \qquad \#L_F = (1-m)N_F \\ \#B_U &= mN_U \qquad \#L_U = (1-m)N_U \end{aligned}$$

Because there are only two players and we 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, we can express the cross-relations between the average back bet (B_F, B_U) and lay bet (L_F, L_U) on favorite and underdog, respectively, as

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

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

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

Solving for B_F and B_U gives

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

We are interested how the average back bet size changes with different proportion of backing bettors on the market. Taking derivatives of B_F and B_U with respect to m we get

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

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

where $B_F^a = \frac{VOL_F}{N_F}$ and $B_U^a = \frac{VOL_U}{N_U}$ 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.

If the results of my estimation suggest that the bettors are risk averse, the following inequalities hold[8]

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

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

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

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

[8]Within the utilized CARA utility framework, the ration of bets of risk netral bettor satisfies the condition $\frac{B_F}{B_U} = \frac{p_F}{p_U}$.

Combining the fact that $B_F < B_F^a$ and $B_U > B_U^a$ results in inequality

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

which means that use of the right average Back bet size would lead to even higher risk aversion estimate. Similarly, if we assume that the representative bettor is risk loving, we can reiterate previous analysis as follows

$$\begin{aligned}
\frac{B_F^a}{B_U^a} > \frac{p_F}{p_U} = \frac{\frac{1}{p_U}}{\frac{1}{p_F}} = \frac{O_U}{O_F} = (O_U - 1)& \\
(O_U - 1) < \frac{B_F^a}{B_U^a} \quad &\Rightarrow \quad \frac{\partial B_F}{\partial m} < 0 \Rightarrow B_F > B_F^a \\
\frac{B_U^a}{B_F^a} < \frac{p_U}{p_F} = (O_F - 1) \quad &= \\
(O_F - 1) > \frac{B_U^a}{B_F^a} \quad &\Rightarrow \quad \frac{\partial B_U}{\partial m} > 0 \Rightarrow B_U > B_U^a \\
\frac{B_F}{B_U} \quad &> \quad \frac{B_F^a}{B_U^a} < \frac{p_F}{p_U}
\end{aligned}$$

In both cases, use of average betting size computed under assumption that $m = 1$ biases the results towards risk-neutral preferences. Thus, we can conclude that our estimate of risk aversion is a lower bound of a real value.