Behavioural biases never walk alone: An empirical analysis of the effect of overconfidence on probabilities.

Isabel Abinzano, Luis Muga and Rafael Santamaria

Universidad Publica de Navarra

(SPAIN)

October 2014

Abstract:

This paper presents evidence of the impact of overconfidence bias in asset prices drawn from a study based on data from tennis betting exchanges. A series of betting strategies in tournaments with a clear-cut favourite are shown to yield significant economic returns. The impact of overconfidence bias on betting odds increases with trading volume, media coverage, and levels of disagreement between overconfident and Cumulative Prospect Theory bettors. Just as in traditional financial markets, arbitrage limits are shown to be a necessary condition for the impact of behavioural biases on prices.

Keywords: Overconfidence, betting exchanges, anomalies, behavioural finance

JEL code: G02, G14

ACKNOWLEDGEMENTS

This paper has received financial support from the Spanish Ministry of Economy and Competitiveness (ECO2012-35946-C02-01).

Corresponding author: Luis Muga. Dpto. Gestión de Empresas. Universidad Publica de Navarra. Campus de Arrosadía s/n 31.006 Pamplona (Spain). Ph. 34-948166079. Fax. 34-948169404. E-mail: rafael@unavarra.es

Behavioural biases never walk alone: An empirical analysis of the effect of overconfidence on probabilities.

Abstract:

This paper presents evidence of the impact of overconfidence bias in asset prices drawn from a study based on data from tennis betting exchanges. A series of betting strategies in tournaments with a clear-cut favourite are shown to yield significant economic returns. The impact of overconfidence bias on betting odds increases with trading volume, media coverage, and levels of disagreement between overconfident and Cumulative Prospect Theory bettors. Just as in traditional financial markets, arbitrage limits are shown to be a necessary condition for the impact of behavioural biases on prices.

1.- Introduction

In recent years, betting markets have proved a useful "testing ground" for the validity of behavioural finance theories (see Smith et al., 2006 or Hetherington, 2006). Betting markets a much more suitable context in which to investigate potential mispricing[1], despite the problems typical of information markets where strategies are traded based on the probability of given outcomes, and the odds reflect the market's estimate of that probability. This leaves the odds vulnerable to the various cognitive biases that may appear in punters' probability estimates, the commonest being those described by the Prospect Theory (Kahneman and Tversky, 1979). Under these conditions, the presence of traders operating under bounded rationality may cause prices to diverge from their fundamentals and thus skew the odds in betting markets.

According to the tenets of the Cumulative Prospect Theory (CPT), market agents operate on the basis of probability transformations, whereby low probabilities are over-weighted and high probabilities are under-weighted. Given that the probability of the favourite winning each match is a high probability outcome, it will be underweighted. This should mean that the return on a strategy based on systematically betting on the favourite winning the set of matches that make up the tournament (a composite event) will be non-negative, or even positive if the market is not fully arbitraged. In this CPT context, this paper attempts to show empirical evidence of the price impact of overconfidence and self-attribution biases, described in theoretical models such as that of Daniel et al. (1998), using data drawn from tennis betting markets[2]. According to this model a run of good news in conjunction with the presence of overconfident investors results in asset overpricing. Furthermore, overconfidence bias tends to arise in combination with self-attribution bias, whereby traders ignore any information contradictory to their priors while incorporating

[1]Betting exchanges offer the final value of the underlying asset and have lower limits to arbitrage, due to the absence of short-sales constraints.

[2] Tennis betting markets were selected as our context of analysis because they allow gamblers to bet on a player either to win or to lose the tournament (a simple event) or all of the rounds (a composite event).

any confirmatory information into their decisions. Fundamentals, however, eventually prevail and prices tend towards equilibrium.

Using this argument and drawing a parallel with the tennis gambling market, we use the favourite to win a given tournament as an analogy for a stock whose price is driven by a run of good news. We define the favourite as the player whose estimated probability of winning the tournament according to betting market odds is greater than that of any other prior to its commencement[3]. If the model predictions are correct, gamblers will over-estimate the favourite's chances of winning the tournament prior to its commencement, thus reducing the odds on that outcome. However, as the real information emerges, round by round throughout the tournament, the odds will adjust and should increase.

The paper is organised into 6 sections. Section 2 contains a review of the literature and its relationship with the main characteristics of the betting market analysed. Section 3 presents the theoretical framework and the hypotheses to be tested. Section 4 describes the data base. Section 5 shows the results of the various betting strategies employed. Section 6 describes the robustness check; and the 7th and final section contains the main conclusions.

2.- Behavioural Finance and betting exchanges

Previous research on sport betting markets has focused on the one hand on the so called "Favourite-long shot bias" (FLB) in bookmaker-based markets (see for example Jullien and Salanié, 2008). This bias basically reflects a tendency for higher rewards to accrue from bets placed on favourites, versus lower rewards for bets placed on less likely outcomes (long shots), given the odds offered by the bookmakers. Although this phenomenon might be compatible with the predictions of Cumulative Prospect Theory, as shown by Snowberg and Wolfers (2010), it has also been attributed to causes such as information acquisition costs and the presence of uninformed punters (Hurley and McDonough, 1995), who tend to place too much money on very unlikely outcomes and too little money on more likely ones. This type of behaviour exerts a pressure on bookmakers that eventually skews the odds. Koch and Shing (2008) show that it is the pattern of the odds set by bookmakers in the United Kingdom that tends to accentuate the presence of Favourite-long shot bias. Moreover, Lahvicka (2013), shows evidence of FLB in bookmaker tennis betting markets. This result is explained by bookmakers protecting themselves against better informed bettors.

On the other hand, some recent papers have tested different arbitrage strategies in this type of bookmaker betting markets. Vlastakis et al. (2009) find simple betting strategies that provide

[3] The requirement for bettors to consider a player the favourite is a run of good news. In the event of being associated with a sufficiently significant run of bad news, such a player will automatically cease to be considered the favourite (their odds will raise in the betting market).

significant positive returns using data from European Football Betting Markets or Ashiya (2013), who shows arbitrage opportunities in the Japanese Racetrack Betting Market.

Recently, the popularity of the Internet has given rise to another type of betting market, where, the odds are set not by the bookmakers but by the punters' supply and demand, thus following a process similar to that of an order-driven financial market. That is, the investors' buy and sell orders build the order book which organises the transactions. Thus, the spread is not determined by the ask and bid price set by the bookmaker, but by the best ask and bid prices placed on the order book at that moment. This type of market is known as a "betting exchange". This form of market organisation has some advantages when compared with bookmaker-based markets in terms of price setting. Firstly, Franck, Verbeek and Nüesch (2010) show that betting exchanges are better predictors of future events. These same authors (Franck, Verbeek and Nüesch, 2013) signal the presence of arbitrage opportunities between these two types of markets due to mispricing in the bookmaker-based market. Smith et al. (2006) further report that pricing is less vulnerable to FLB in betting exchanges than in bookmakers markets, a finding which they attribute to lower transaction costs. They also find liquidity to have a major impact, in that, the more liquid the market, the lower the degree of FLB. "Betting exchanges" lend themselves particularly well to the testing of hypotheses about investor behaviour, because the practical absence of arbitrage constraints, which are even lower than in many financial markets, allows arbitrageurs to exploit any mispricing due to noise trading.

Under these premises, betting exchanges offer an ideal setting in which to test for behavioural biases among agents. Taking investor behaviour to be adequately described by the assumptions of the Cumulative Prospect Theory (Tversky and Kahneman, 1992), the aim in the case in hand is to test for a possible additional effect of punters' overconfidence bias deriving from a run of good news about a potential outcome, (the favourite winning the tournament). Confirmation of such an effect would help to reinforce the existing indirect evidence about the impact of investors' overconfidence bias on stock prices in financial markets. The presence of investors trading under the effect of this behavioural bias has been suggested as the source of the momentum effect (Cooper et al., 2004), and of the presence of speculative bubbles in stock prices (Hong et al., 2006).

The phenomenon of overconfidence bias is not unique to financial markets. Various papers have reported its presence in information markets and betting markets. Berg and Rietz (2010) show that overconfidence bias has an impact on long term price formation in the Iowa Electronic Market for different events and Arkes (2011) finds evidence that over-optimistic punters bias prices in betting markets for NBA basketball games.

As already mentioned, our choice of market for the analysis proposed in this paper is the tennis betting exchange. The choice of this sport stems from the fact that the competitions are organized as single elimination tournaments. The market has a sufficient degree of liquidity, which is important, because Smith et al. (2006) have shown liquid markets to be less vulnerable to FLB[4], and Borghesi et al. (2009) have shown that behavioural biases have a greater impact on prices in betting markets with low liquidity. The fact that tennis also attracts wide media coverage allows us to assume that punters will be sufficiently well informed as to avoid being affected by the aforementioned FLB, of which media coverage is the main driver (Hurley and McDonough, 1995). Finally, the punters are not as strongly identified with the players as they tend to be in other sports which have teams representing a city or a country[5].

3.- Theoretical framework and hypotheses

The CPT (Tversky and Kahneman, 1992) is a modified version of the Prospect Theory (Kahneman and Tversky, 1979) which analyses uncertain as well as risky prospects with any number of outcomes, applies the probability weighting function to the cumulative probability distribution and not to the probability density function[6], and allows different weighting functions for gains and for losses. Formally, let (x_i, p_i) be a risky prospect, $(x_{-m}, p_{-m}; \ldots; x_{-1}, p_{-1}; x_0, p_0; x_1, p_1; \ldots; x_n, p_n)$ with $x_i > x_j$ if $i>j$ where x_i is the outcome (positive or negative)that results if event A_i occurs and $p_i=P(A_i)$ is the probability of event A_i occurring. Under Cumulative Prospect Theory, the value of the gamble is $\sum_{i=-m}^{n} \pi_i v(x_i)$ where $\pi_i = \pi_i^+$ if $i > 0$ and $\pi_i = \pi_i^-$ if $i < 0$, with

$$\pi_i^+ = w^+(p_i + \ldots + p_n) - w^+(p_{i+1} + \ldots + p_n) \text{ if } 0 \le i < n \text{ ; } \pi_n^+ = w^+(p_n) \text{ if } i = n \qquad (1)$$

$$\pi_i^- = w^+(p_{-m} + \ldots + p_i) - w^-(p_{-m} + \ldots + p_{i-1}) \text{ if } -m < i \le 0 \text{ and } \pi_{-m}^- = w^-(p_{-m}) \text{ if } i = -m \text{ .} \qquad (2)$$

The functions w^+ and w^- are strictly increasing from the unit interval into itself, satisfying $w^+(0) = w^-(0) = 0$ and $w^+(1) = w^-(1) = 1$. In this modified version, Tversky and Kahneman (1992) propose specific functional forms for $v(x)$ and π:

$$v(x) = x^\alpha \text{ if } x \ge 0 \text{ ; } v(x) = -\lambda(-x)^\beta \text{ if } x < 0 \qquad (3)$$

$$w^+ = \frac{p^\gamma}{(p^\lambda + (1-p)^\gamma)^{1/\gamma}} \text{ and } w^- = \frac{p^\delta}{(p^\delta + (1-p)^\delta)^{1/\delta}} \qquad (4)$$

where $\alpha \in (0,1)$, $\beta \in (0,1)$, $\gamma \in (0,1)$, $\delta \in (0,1)$ and $\lambda > 1$.

---

[4]The top seeds in the ATP (Association of Tennis Profesionals) ranking have to play in the four Grand Slam tournaments (the Australian Open, Roland Garros, Wimbledon, and the US Open) and the 9 Masters Series Circuit tournaments, which, in the sample period, were (Indian Wells, Miami, Rome, Monte Carlo, Hamburg, Canada, Cincinnati, Madrid and Paris Bercy). This choice guarantees a minimum level of market liquidity.

[5] Tetlock (2004) claims that one of the factors motivating people to take part in this type of market, namely, their identification with one of the teams or players in the sporting event, could skew the results and prove difficult to control. This assumption of a lower degree of national identification does not apply in other tennis events, such as the Davis Cup, which are not considered in this paper.

[6] This ensures that Cumulative Prospect Theory does not violate first-order stochastic dominance, a weakness of the original Prospect Theory.

These functions capture the key elements of the theory. Thus, the function *v(x)* is concave over gains, convex over losses, and exhibits a greater sensitivity to losses than to gains. The degree of sensitivity to losses is determined by λ, the coefficient of loss aversion. The weighting functions, w^{+} and w^{-}, include a nonlinear transformation of the probability scale, which over weights small probabilities and under weights moderate and high probabilities. This form provides a good approximation to both the aggregate and the individual data for probabilities in the range between 0.05 and 0.95 (see Tversky and Kahneman, 1992, pg.310).

It is known that the probability of a composite event $A_j = (A_{j,1} \cap A_{j,2} \cap ... \cap A_{j,N})$ can be obtained by multiplying the probabilities of the N simple events $A_{j,n}$ of which it is composed, such that

$$P_j = \prod_{n=1}^{N} P_{j,n}^{c} \text{; with } P_{j,n}^{c} = P_{j,n|(j,1 \cap j,2 \cap ... \cap j,n-1)} \; \forall n > 1 \quad (5)$$

where P_j is the probability of a composite event A_j and $P_{j,n}^{c}$ is the conditional probability of the simple event $A_{j,n}$.

If the simple event and the composite event can be perfectly arbitraged[7], assuming the presence of sophisticated investors, it is reasonable to suppose that

$$w(P_j) = \prod_{n=1}^{N} w(P_{j,n}^{c}) \quad (6)$$

In the case in hand, the "Win Tournament" (W_T) outcome is considered a simple event on which it is possible to bet at odds which allow us to infer the associated probability. If, for the sake of homogeneity, we assume that this probability is estimated at a point in time t_0 immediately prior to the start of the tournament,

$$w(P(W_T \mid \Omega_{t_0})) = \prod_{n=1}^{N} w(P(W_n^{c} \mid \Omega_{t_0})) \text{ with} \quad (7)$$

$$P(W_n^{c}) = P(W_n \mid W_1 \cap W_2 \cap ... \cap W_{n-1}) \text{; } \forall n > 1 \text{ and N=rounds of the tournament} \quad (8)$$

where $w(P(W_n^{c} \mid \Omega_{t_0}))$ is the probability of winning the n[th] round, given the knowledge that an event $(W_1 \cap W_2 \cap ... \cap W_{n-1})$ has already occurred, as far as can be inferred from the market odds for that bet, conditional to the information available at time t_0, and $w(P(W_t \mid \Omega_{t_0}))$ is the probability of winning the tournament conditional to the set of information available at time t_0, Ω_{t_0}. Note that the "Win Tournament" (W_T) outcome is considered a simple event and this outcome can also be obtained from the product of the probabilities of the simple events that make up the composite event "Win Tournament", that is, "Win all rounds of the tournament".

[7] In absence of perfect arbitrage there is no reason to fulfill this relationship. Nevertheless, the presence of sophisticated investors and perfect arbitrage between both events (simple event and the composite event) lead to the fulfillment of this relationship because, in other case, it is possible to obtain abnormal positive returns without assuming any risk.

Given the characteristics of the competition, however, we cannot obtain information about $P(W_n \mid \Omega_{t_0})$ for any n>1, given that we do not know which players will get through from the first to the second round. Thus, the probability of the composite event "win all the rounds of the tournament" is not given by the simple events as in expression (7) but by the following expression:

$$w(P(W_T \mid \Omega_{t_N})) = \prod_{n=1}^{N} w(P(W_n^c \mid \Omega_{t_n})) \quad (9)$$

the value of which may be greater or less than $w(P(W_T \mid \Omega_{t_0}))$ the difference being random. The fact that $w(P(W_T \mid \Omega_{t_N}))$ is obtained on the basis of a more complete set of information than $w(P(W_T \mid \Omega_{t_0}))$ neither increases nor decreases its value, since this will depend on the player's progress, which will increase or decrease his probability of winning throughout the consecutive rounds of the tournament.

3.1.- Testable Hypotheses

Bettor overconfidence bias can affect the favourite's estimated probability of winning the tournament or winning all of the rounds. Prior to the tournament's commencement, the favourite will be the primary driver of overconfidence. A player who has, at some point during the run-up to the tournament, held the position of favourite, will very likely have lost that position by the time the tournament commences, if he or she is hit by a run of bad news. Thus, it takes a run of good news for the market to consider a player the tournament favourite.

Daniel et al. (1998) developed a model in which investors' overconfidence bias is strongly apparent in a market when there is a run of good news about a particular firm. This bias may also arise in combination with self-attribution bias, whereby investors ignore any information that contradicts their priors. The combined effect of these two biases leads to an overreaction of stock prices, such that, if the market has a high proportion of this type of investor and there is a run of good news, a stock price will appear overvalued. The model also predicts that prices will revert to their fundamentals in the long term, since information eventually prevails. In the same vein, Vergin (2001) shows that investors have a stronger tendency to overreact to good news than to bad news and therefore the effect will be more notable in the wake of a run of good news, as the above argument suggests.

Transporting this argument to the context of tennis tournaments, we must assume that, prior to the commencement of the tournament, given a build-up of good news about the favourite to win and in the presence of the biases described in Daniel et al. (1998), the betting market's estimate of the favourites' probability of winning will be higher (and thus the odds lower) than under perfect incorporation of information without mispricing. It should be noted that, according to the tenets of the CPT, very high objective probability will be associated with lower transformed probability. Our main point, however, is that overconfidence bias will result in a reduction in the underestimation of objective probability, that is, in relative terms, it will be overestimated with respect to the transformed probability estimate in a CPT model with no overconfidence bias. Given that, in this case, overconfidence bias runs counter to the probability transformation effect described by the CPT, it is not easy to predict the final effect on the betting market odds, since they will depend on the degree of probability transformation which, in turn, depends on the objective probability estimate and the level of overconfidence bias among the punters.

The extended period prior to the tournament for gamblers to bet on the winner of the tournament, especially in the case of Grand Slam[8] events, allows the build up of good news about the favourites, which skews the punters' estimates of the favourites' chances of winning. Further, as already mentioned, Berg and Rietz (2010) find that the presence of overconfident traders has a greater effect on prices over the long term. Later, as the tournament proceeds, information pertaining to the favourite's real likelihood of winning will gradually reveal itself, thus relieving the downward pressure on the odds on the various matches and reducing the degree of overconfidence bias in the information incorporated into the odds. This suggests the following hypothesis:

H_1: *The probability of the favourite winning the tournament (a simple event) is higher than the product of his probabilities of winning all rounds in the tournament including the final (a composite event),* or, in mathematical terms:

$$w(P(W_T \mid \Omega_{t_0})) > \prod_{n=1}^{N} w(P(W_n^c \mid \Omega_{t_n})) \qquad (10)$$

Given that the potential spread between these two probabilities could be exploited through a pseudo-arbitrage strategy (pure arbitrage being unfeasible), the presence of overconfident punters will increase arbitrage constraints due to noise trader risk, which will allow the spread to persist for a period before reverting to equilibrium, as predicted by the model presented in De Long et al. (1990).

It seems reasonable to assume that a minimum of media attention is required for news to reach the market. Indeed, Tetlock (2004) and Avery and Chevalier (1999) report price overreaction to news due to media coverage. This leads us to the following hypothesis:

H_2: *The underestimation of the probability of the favourite winning the tournament (a simple event)* $w(P(W_T \mid \Omega_{t_0}))$ *will be lower in markets with higher media coverage.*

[8] Betting on the winner of a tournament opens in advance of the start of the tournament, the period varying from 100 days in the case of Grand Slam tournaments, to 6 days in that of Masters Series tournaments. Betting on the outcome of the first round commences 3 days in advance, after which there are two days to place a bet on a Grand Slam match and 1 to place a bet on a Masters Series match.

This implies that the spread $w(P(W_T \mid \Omega_{t_0})) - \prod_{n=1}^{N} w(W_n^c \mid \Omega_{t_n}))$ will widen as media coverage increases, because high media attention conveys the good news about the favourite to a greater number of bettors.

Thirdly, the literature relating to the presence in financial markets of traders with behavioural biases claims that such biases will be stronger in individual traders than in institutional traders[9]. Given that traditional bookmakers cover their positions by operating through "betting exchanges", their presence as sophisticated traders can have a significant impact on prices, which enables us to propose the following hypothesis:

H_3: *The presence of informed traders in the market will mitigate the reduction in the underestimation of the probability of the favourite winning the tournament [the simple event* $w(P(W_T \mid \Omega_{t_0}))$].

Finally, the finance literature has found a positive relationship between mispricing caused by traders' behavioural biases and total volume traded. For predicting this relationship, Hong and Stein (2007) presented a family of models that might be termed "Disagreement models". Scheinkman and Xiong (2003) develop a model which relates overconfidence bias in stock market investors to the emergence of speculative bubbles associated with high trading volume; and Hong, Scheinkman and Xiong (2006) demonstrate this effect empirically. The disagreement in the case that concerns us is between gamblers with high vs. low levels of overconfidence and self-attribution bias: in this context, the greater the level of volume traded in the market the greater the mispricing of the odds. This leads us to fourth and last hypothesis:

H_4: *The reduction in the underestimation of the probability of the favourite winning the tournament [the simple event* $w(P(W_T \mid \Omega_{t_0}))$] *will be greater in markets with higher trading volume.*

<u>4.- The database</u>

We analyse data from the "Betfair"[10] betting exchange for the period running from May 2004 to December 2008. The reason for this choice of sample period is that it lends itself particularly well to the testing of the above hypotheses on investor overconfidence because it featured the dominance on the ATP[11] circuit of two outstanding favourites: Roger Federer (on grass and hard courts) and Rafael Nadal (on clay). Since then, with the emergence of Novak Djokowic and Andy Murray, among others, the number of potential favourites to win tournaments has increased,

[9] Frazzini (2006), for example shows that the disposition effect is stronger in individual investors than in mutual fund managers.

[10] Betfair is the largest online betting company in the UK and the largest betting exchange in the world with over four million subscribers.

[11] Some examples of runs of good news about favorites during this period (2004-2008) include: Roger Federer having won the highest number of consecutive matches in tennis history on hard courts and grass (56 on hard court, 65 on grass, and 26 against one of the top 10 seeds) and Rafael Nadal having achieved the highest number of consecutive matches won (81) on clay.

thus reducing each player's individual chance of winning. During the sample period, the women's circuit (WTA TOUR) does not fulfil the necessary conditions for testing the impact of the presence of overconfident bettors on the odds, given the lack of any clear favourite[12].

The analysis uses data on ATP tournaments in which the top-seeded players must participate and cannot withdraw except in the event of injury[13]. This guarantees a minimum level of market liquidity (bid-ask spread and depth).

For each event, the Betfair database includes both pre-event (PE) and in-play (IP) betting data. We haven't however used the odds on the outcome of the tournament once the event has started or the odds on individual matches after their commencement (IP), because these would capture not only gamblers' prior beliefs but also information regarding the results of matches and the progress of the tournament. Furthermore, Docherty and Easton (2012) using golf tournament betting data drawn from the Betfair company's database, show that once the competition has started, the odds under react to information emerging from the proceedings. Williams (2010) reach similar conclusions using NBA betting data.

Table 1 gives a summary of each of the tournaments considered in this paper, including the number of entries per tournament in the database. After cleaning and filtering, we are left with a dataset that includes 281,952 bets on the favourite to win Grand Slam events, 110,852 bets in the alternative market for these events, and 155,521 bets on Masters Series events placed prior to their commencement (PE)[14].

From this database, we extracted the betting odds, the number of bets placed at those odds, the betting volume in pounds sterling, the first and last time each odds were taken, event identifiers and the direction of the bet. These data enable the construction of odds variables such as the volume-weighted average, the closing odds before the start of each match and the tournament, the odds at which most of the money was bet, the standard deviation, the skewness coefficient, the average bet size, the total bets taken per event, and the total duration of the betting period for a given outcome (in number of days).

Given the aims of the paper, the said data refer only to the favourite, that is, the player with the lowest odds before the tournament starts. As indicated in the preceding sections, this player must, by variable construction, be the one that generates the accumulation of bettors acting

[12] Focusing on the main events of the WTA circuit between the 2004 and 2008 seasons, we can see that up to 9 different players succeed in winning at least one Grand Slam title.

[13] As noted, these are Grand Slam and Masters Series tournaments. The analysis does not include the Hamburg (2006), Madrid (2004) or Paris (2004) tournaments, where the average weighted odds on the favourite to win were higher than 8. Paris (2006) was also excluded because the favourite, Roger Federer, withdrew before the tournament started.

[14] In only 9 of the 18 women's Grand Slam tournaments included in our database it is possible to implement the strategy proposed in the paper to detect the presence of overconfident bettors. Proof of this comes from the fact that in 3 cases out of 9, in which the strategy can be implemented, the favourite changed over the course of the tournament.

under overconfidence bias. First, we extracted data pertaining to the market for bets on the favourite to win in each tournament, then the data pertaining to the market for bets on the favourite to win each of the rounds of each tournament. For every match, the Betfair Company will, in addition, take bets on either the favourite or his rival winning (or losing). This alternative market proves particularly useful for testing whether markets with the same information but different types of gamblers generate the same odds.

5.- Results

5.1.- Returns of the pseudo-arbitrage strategies

The probability of the favourite winning each match is a high probability outcome. In line with CPT tenets, this probability will be underweighted[15]. As in Andrikogiannopoulou and Papakonstantinou (2013)[16], in this analysis, we assume that individuals evaluate bets using the probabilities implied by their odds (Q_j)[17].

The information to be drawn from the weighted average odds shows that, overall, strategies based on betting on the favourite winning every match leading up to and including the final yield positive, albeit not significant, returns (26.41%, p=0.18). Similar findings are obtained for betting on the favourite winning the tournament (15.1% p=0.35). These findings provide evidence of the effect of probability transformation in the case of high probabilities. Nevertheless, the presence of sophisticated bettors should prevent simple probability transformation from making the probability of the simple event significantly different from that of the composite event. The presence of bettors operating under overconfidence and self-attribution biases, however, might explain why the probability of the simple event is less underestimated than that of the composite event. The presence of behaviourally-biased investors in a market should increase noise trader risk and turn informed traders away, thus increasing arbitrage constraints and leaving the way open for temporary mispricing, as postulated by De Long et al. (1990).

We explore the possibility of mispricing by selecting an ex-ante investment strategy. It is based on betting that the favourite will lose the tournament[18], assuming the effect that the presence of overconfident bettors will have on the estimated probability of this occurring, and staking the potential winnings of this bet on the composite event, that the favourite will win each of the rounds. This strategy can lead to three possible outcomes. If the player loses the tournament, the bettor will win the bet placed on that event and lose the bet placed on the match in which the

---

[15]In the 2006 Australian Open, for instance, Roger Federer's average estimated probability of winning a match was 0.9264 (implied probability) and his probability of winning the tournament was 0.6572.

[16] These authors motivate the assumption that subjective beliefs can be approximated with the probabilities implied by the prices due to the betting markets are quite efficient and several online tools help individuals convert the quoted prices into the implied probabilities. They also shown that there are small but significant deviations of the subjective from implied probabilities, but their results are not affected.

[17] We analyse data from betting exchanges where there is no bookmaker, so implied probabilities do not have to take into account the broker commission. $IP_j=1/Q_j$, where IP_j is the implied probability on outcome j and Q_j is the resulting market odds on outcome j.

[18]As mentioned earlier, one of the advantages of betting exchanges is that they offer both odds in favour and odds against a given outcome.

player is knocked out of the competition, leaving the bettor's winnings at 0. If the player wins the tournament and the assumption of overconfidence bias is fulfilled $w(P(W_T \mid \Omega_{t_0})) > \prod_{n=1}^{N} w(P(W_n^c \mid \Omega_{t_n}))$, the bettor will receive a positive return. Conversely, if the above assumption is not fulfilled, the returns will be negative or null. In mathematical terms, this strategy takes the following form:

$$R_{k,N} = \frac{\prod_{n=1}^{N} Q(W_n \mid \Omega_{t_n}) - Q(W_T \mid \Omega_{t_0})}{Q(W_T \mid \Omega_{t_0})} \tag{11}$$

where $R_{k,N}$ is the return to strategy k, assuming that it involves N simple events and that that the strategy can be fully implemented. $Q(W_T \mid \Omega_{t_0})$ is the observed market odds on winning the tournament conditional to the set of information available at time t_0 , Ω_{t_0} and $Q(W_n \mid \Omega_{t_n})$ the observed market odds on winning the n^{th} round conditional to the set of information available at time t_n , Ω_{t_n} .

Table 2 gives the results for all those tournaments - Grand Slam (GS) and Masters Series (MS) - in which it was possible to implement the above strategy after transaction costs[19]. The strategy was implemented at three different odds. The first was the volume-weighted average odds for each event. The second was the odds with the highest betting volume and the third was the closing odds prior to each event start, assuming these last to capture all the pre-event information[20]. Note, however, that, in a study of Italian football betting markets, Innocenti et al. (2012) reported gamblers displaying various behavioural biases arriving on the betting scene shortly before match commencement.

Using volume-weighted average odds, we found a non-significative average return across these strategies of 3.58% for each event. The return was 6.80% for the strategy using the odds with the highest betting volume and 10.27% for the one using the closing odds which are significantly different from 0 using the t-statistic. We also ran a non-parametric test (the binomial test). Using the average weighted odds, the strategy proved fully implementable in 36 cases, yielding positive returns in 19 of them and negative returns in 17 (p=0.3089). The returns using the odds with the highest betting volume were positive in 24 cases and negative in 12 (p=0.0144), and with the closing odds there were 21 positive and 10 negative (p=0.0147)[21]. The strategy has proven economically significant in two cases out of the three considered.

This finding is consistent with the presence of overconfident bettors, whose influence on odds is greater than that of transaction costs. In the absence of overconfident bettors, the difference

---

[19] The costs set by Betfair during the sample period range between 2% and 5% of the winnings per bet placed, depending on the total bets of the individual bettor. Given our trading-intensive strategies, we assume the lowest level of transaction costs. Betfair does not apply transaction costs to losing bets.

[20] The raw returns of the strategies are all significantly different from 0 using the t-statistics. The results are available from the authors upon request.

[21]Only 31 strategies were possible with the closing odds, because of missing values in the data provided by Betfair for 2008.

between the estimated probability of the simple event (the favourite winning the tournament) and that of the composite event (the favourite winning all the rounds of the tournament) should differ only randomly. In short, the results provide evidence to support hypothesis H1. That is, the odds on the favourite winning the tournament are low in comparison to the odds on his winning each round, by which time real information is emerging.

The results can be explained by the fact that the events do not take place simultaneously. Contrary to the rational hypothesis that the absence of behavioural biases that would skew probability estimates is purely due to chance, the overconfidence bias in the case in hand is unequally distributed over the two events. The simple event shows the effect of an accumulation of overconfidence in the favourite's chances of winning the tournament, which is driven by a fairly long run of good news. This probability estimate, which is overestimated with respect to that found in the absence of such a bias, cannot be fully arbitraged because it does not offer pure arbitrage opportunities. This, together with the impact of noise traders, increases arbitrage constraints, thus allowing temporary divergence of prices from their equilibrium values. The start of the tournament sends new information to the market, which gradually adjusts the impact of the aforementioned bias on the odds and reduces its effect on the estimated probabilities of the simple events that make up the composite event (win all the rounds of the tournament).

5.2.- Media coverage: Grand Slam Vs Masters Series

If overconfidence is the main explanation for the findings in relation to H1, this bias should increase with media coverage of the event, in line with the findings reported in studies such as those of Tetlock (2004) and Avery and Chevalier (1999). We test this hypothesis (H2) by splitting the sample into two subsamples: betting markets for GS tournaments, which receive greater media coverage, and MS tournaments. Table 2 also gives the results after deducting transaction costs yielded by implementing the strategies separately for each subsample. It can be seen that, in the case of GS tournaments, both the t statistic and binomial test values indicate significant returns to the described strategies for the three types of odds considered. Specifically, strategies based on volume-weighted average odds in GS events gave estimated average returns of 5.32%, those based on the odds with the highest betting volume 9.62%, and those based on the closing odds 17.64%. The same strategy yields lower returns when used in MS tournaments, however. Specifically, average returns were 1.84% (6 events with positive returns and 12 with negative returns) for the strategy based on the volume-weighted average odds, 3.98% (8 positive and 10 negative) for that based on the odds with the highest betting volume; and 3.98% (8 positive and 8 negative) for that based on the closing odds. Neither of the significance tests performed showed these returns to be significant.

These findings provide evidence to show that the greater media coverage given to GS tournaments increases the imbalance between the odds, and thereby the returns to the strategies, as stated in Hypothesis 2. Testing for differences in strategy returns between tournament types, the t statistics reveal no differences in terms of average returns, irrespective of the type of odds considered and whether or not transaction costs are included. Nevertheless, the results of the non-parametric Mann-Whitney U test for the odds with the highest betting volume and the closing odds, both before and after deducting transaction costs, enable us to reject the hypothesis of equal returns at the 10% level of confidence. The low number of strategies and the lower dependence of the fulfilment of the hypotheses on the distribution recommend the use of the non-parametric test.

The results, therefore, provide support for H2, and additional evidence to demonstrate that the probabilities of the simple event ("win the tournament") and the composite event ("win all the rounds of the tournament") are not random outcomes, are not uniform across all tournaments, and cannot be attributed to the impact of probability overestimation associated with conjunctive events. Indeed, the difference appears to be related to variables associated with the level of overconfidence bias among bettors.

To further explore the potential factors to explain the difference between the GS and MS tournaments, Table 3 shows the principal characteristics of the betting market for the winner of the tournament and the p-values of the t statistic for the difference of means between the two types of tournament. From this table, it can be seen that there is a significantly larger number of different betting odds in GS tournaments, although the standard deviation and skewness coefficient for the two types of tournament show no significant differences. The main differences are to be found in the market-volume variables, which show that GS tournaments attract a significantly higher betting volume, both in monetary terms and in numbers of bets, as was expected given the aforementioned higher media coverage of GS events. Finally, the average bet size is also shown to be significantly larger in GS tournaments. Since this variable serves us as a proxy for investor type, its higher value suggests a greater presence of institutional bettors. Despite the intuition that greater media coverage should increase the number of small bettors in the market, thus reducing the average bet, the results show that this sort of reasoning overlooks an important additional factor relating to bets made by bookmakers in order to hedge the odds offered to their customers. Higher betting volume resulting from greater media coverage drives bookmakers to make more use of the betting exchange to exploit its higher liquidity, thus increasing average bet size. This effect is much greater than that of the reduction in average bet size resulting from the influx of individual bettors, which explains the observed final result[22].

[22]The results of this characterization for each round in the tournaments are available from the authors upon request. They show that betting volume, both in monetary terms and in number of bets, the average bet and the duration of the betting period are always greater in Grand Slam tournaments, in most cases to a significant degree.

Finally, as noted earlier, the duration of the betting period, that is, the number of days that the market operates prior to the start of the competition, is considerably longer in GS betting markets. This could facilitate odds mispricing resulting from a run of good news about the favourite in combination with bettors' overconfidence[23].

5.3.-Arbitrage limits: The "alternative" market

The results of the previous section have shown that market characteristics such as investor type and trading volume could be having an impact on final odds prices. Obviously, the real opportunities for arbitrage between betting markets will also have an important impact on them. One of the issues raised previously is that the strategy for betting on the simple event "win tournament" is not strictly equivalent to the product of the simple events of which it is composed, since these are not all available at the starting point, t_0. The strategy does not, therefore, constitute pure arbitrage, in the strict sense of self-financed, risk-free strategies, implemented simultaneously at the same point in time, t_0. It is, therefore, worth investigating whether the above results can be attributed to arbitrage limits in the market.

Thus, the markets considered enable us to obtain a fairly accurate assessment of the issue raised, since one of the features of these markets is that it is possible, in every match, to bet in one of two "alternative markets". That is, the bet can be placed in the betting market for event "i against j" (player i against player j), where it is possible to bet on player i winning (or losing) or, in the betting market for event "j against i" (player j against player i) where it is possible to bet on player j losing (or winning). Thus, a bet in the "main" market[24]on the favourite winning a specific match should be equivalent to a bet in the "alternative" market on his opponent losing. For the same information to be incorporated into the odds in both markets, the following relationship must be fulfilled

$$QA(W_n/\Omega_{t_n}) = \frac{1}{(Q(W_n \mid \Omega_{t_n}) - 1)} + 1 \qquad (12)$$

where $Q(W_n \mid \Omega_{t_n})$ is the odds in any main market associated with event n (in the case in hand, this would be the n^{th} match in the tournament, denoted by i against j), conditional to the set of information available at t_n and $QA(W_n \mid \Omega_{t_n})$, the theoretical odds in the alternative betting market for that event (that is, the odds on the n^{th} round in the tournament, denoted by j against i).

---

[23] There are also significant characteristic differences between the men's and the women's Grand Slam betting markets. In particular, the average odds in women's tournaments are 2.88 versus 1.92 in men's tournaments, and the average number of different odds is 123.55 versus 34.89, respectively. These data help to illustrate the higher level of uncertainty and the absence of a clear favourite in the women's circuit during the period of interest The results are available from the authors upon request.

[24] The "main" market is considered to be the one where player i is the favourite.

The characteristics of the main and alternative markets differ significantly for the GS tournaments[25]. In particular, the number of different odds for each round is greater in the alternative market, except in the case of the final round. Betting volume is significantly lower in all rounds of the tournament in the alternative market, where the number of bets placed is also lower, and significantly so in five of the seven rounds. This effect may be due to bookmakers and informed bettors seeking to exploit the more liquid market, with a view to preventing their transactions from having a negative impact on the odds. Finally, the average bet size in all rounds is also significantly lower in this market.

The findings for the above-mentioned variables appear to reveal the presence of less informed bettors in the alternative market, which could lead to some forms of mispricing resulting from their potential behavioural biases. There are no significant differences between the volume-weighted average odds evolving in the two types of market considered. Furthermore, strategies implemented using odds drawn from the alternative market yield significant positive returns for all three types of odds considered (see Table 4). Similar findings are obtained for the main market. In fact, there is no significant difference in returns between the two markets, either with the t statistic or the Mann-Whitney U-value[26].

These findings indicate that the presence of behaviourally biased agents is not sufficient to influence prices in a given market, and that limits to arbitrage are also necessary, (Shleifer, 2000). In this case, information on the odds in both markets is simultaneously available to any agent, providing opportunities for arbitrage, which will equalize prices between both markets. The existence of this arbitrage is a guarantee that the market under consideration is efficient enough to show that the remainder of the findings of this paper are not due to price setting problems. It should be noted, however, that the virtually total lack of arbitrage constraints has not cancelled out the effect of overconfidence bias or that of the overall underestimation of probabilities in the odds in both markets. It has merely equalized these effects by preventing economic gains from any informational imbalance between these complementary markets (main and alternative). These results suggest that the returns of the initial strategies are due, not to pure arbitrage, but to pseudo-arbitrage operations, which incorporate the noise trader risk effect, whereby prices may diverge temporarily from their fundamentals.

5.4.- Multivariate Analysis

The results of the preceding sections have revealed the existence of several explanatory variables for the difference in odds between the simple event "win the tournament" and the composite event "win all rounds of the tournament". Specifically, they reveal a return spread between strategies

[25] Grand Slam tournaments were selected because of the statistical significance of the strategies and the fact that they all have the same number of rounds (seven), thus facilitating the comparison of the results. The results are available from the authors upon request.

implemented in GS vs MS tournaments. Market characteristic differences may, however, be masking the main reason for the return spread, thus concealing the explanatory variables for the strategy returns. In addition to whether the tournament is GS or MS, average bet size (as a proxy for the dominant investor type in the market) and betting volume are both highly important factors. Likewise, the presence or otherwise of a clear-cut favourite in the tournament could also affect the overestimation a player's probability of winning the final. For a joint analysis of the explanatory power of these variables, we propose the following regression:

$$R_k = \alpha + \beta_1 . GS + \beta_2 Fav + \beta_3 \ln(AvgBet) + \beta_4 \ln(Vol) + \varepsilon_k \tag{13}$$

where R_k is the return of the strategy k, GS is a dummy variable that takes a value of 1 for a GS tournament, and 0 otherwise, Fav is another dummy variable that takes a value of 1 when the volume-weighted average odds are less than 2 (in other words, the betting market gives the favourite a more than 50% chance of winning the tournament) and 0 otherwise, ln(AvgBet) is the natural logarithm of the average bet size, and ln(VOL) is the natural logarithm of the total betting volume.

The results of this analysis will enable us to determine whether, as the previous results suggest, there is a positive relationship between the volume traded in a market and trading strategy returns, which would be consistent with H4, and with the family of behavioural finance models known as "disagreement models" described by Hong and Stein (2007). They will also enable us to jointly test H2, which postulates a relationship between media coverage and strategy returns and H3, which predicts a negative link between the presence of sophisticated traders and strategy returns.

We ran the regression for the returns of the strategies using the volume-weighted average odds, the odds with the highest betting volume, and the closing odds, raw and net of transaction costs. OLS estimation is used with White's (1980) standard errors. The results are displayed in Table 5. The coefficient on the GS variable proves non-significant in all cases, while the coefficient on the FAV variable is positive at levels of significance close to 10%, when the dependent variable is calculated on the basis of the odds with the highest betting volume and the volume-weighted average odds, which suggests that the stronger the favourite, the higher the overestimation of the probability of the simple event, win the tournament, due to stronger impact from the overconfidence bias. The coefficient on the variable ln(AvgBet) shows a consistently negative sign with 10% significance when the dependent variable is constructed based on the closing odds. Given that higher average bet size is a proxy for a higher proportion of sophisticated trading, these results support H3.

---

[26]The strategy returns net of transaction costs, which are very similar to those shown in Table 5, are available from the authors upon request.

Finally, the only variable that proves significant in all the estimations is ln(Vol). A significant positive relationship can be observed between betting volume and strategy returns, as also found in traditional financial markets and as modelled in studies such as that of Scheinkman and Xiong (2003), which relate overconfidence bias with the emergence of speculative bubbles. This result provides evidence to support H4.

It also suggests that differences in the results found between GS and MS tournaments are basically due to trading volume, rather than the simple effect deriving from the type of tournament. This obviously does not rule out the potential effect of media coverage, since the two variables are closely related; that is, trading volume increases with media coverage.

In summary, the analyses performed so far reveal relative overestimation of the probability of the simple event, the favourite wins the tournament, with respect to the probability of the composite event, the favourite wins all rounds, the causal factor being overconfidence bias in the betting market, as also predicted by theoretical models of financial markets. Overconfidence bias is also closely related to higher trading volume, since it creates a divergence of opinion from traders who systematically underestimate probabilities. Finally, it shows a potential, albeit weaker, relationship with the level of media coverage.

Somewhat weaker evidence has been found for a negative relationship between the presence of sophisticated trading and overestimation of the favourite's probability of winning the tournament, suggesting that the former tends to reduce the latter[27]. Given that, as already noted, the strategy is not susceptible to pure arbitrage, the activity of noise traders increases arbitrage limits and, thereby, the likelihood of positive returns to the described strategy.

6.- Robustness test: Boostrap analysis

We have computed a test that uses the bootstrap method to simulate strategy returns, given the low number of tournaments included in the sample, to robust our results. This procedure involves extracting, with replacement, equally sized samples from the initial sample of tournaments in which the strategy is possible. Average strategy returns for all tournaments included in the bootstrap sample are then calculated. We then repeat the process 1000 times and calculate the average of the distribution of bootstrap means, and the level of significance of the null hypothesis of average returns of the strategy for the whole set of tournaments is equal to 0.

This procedure is performed for the GS and MS tournaments separately, and the results are given in Table 6. They show that the average returns to the strategies for the GS tournaments are positive and significant, both raw and net of transaction costs, whereas for the MS

[27] It is important to note an additional explanation, namely that sophisticated traders may also be affected, albeit less strongly, by overconfidence bias, which would also reduce their arbitrage activity.

tournaments, they are in no case significantly different from zero at the standard levels of significance. Thus, these results confirm those of the traditional tests performed earlier.

7.- Conclusions

A run of good news about a given asset leads to investor overconfidence that can cause its price to diverge from its equilibrium value. The price will revert to equilibrium in the long term as it incorporates real information (Daniel et al., 1998). This paper has tested this hypothesis in the particular context of a betting exchange, over a selected period of time offering the desired characteristic of the presence of a run of good news about a given event: the favourite to win a tournament. Our strategy is, nevertheless, a pseudo-arbitrage strategy, which cannot be considered risk-free.

The initial results confirm the hypothesis of downward pressure on the odds on favourites, contrary to the predictions of FLB. Thus, positive returns are obtained from betting against the favourite winning the tournament (the simple event) and then betting in favour of his winning each individual round (the composite event), as information is revealed to the market. This finding, however, is significant only for GS tournaments, which are characterized by greater media coverage and higher betting volume. This enables us to confirm that our results cannot be considered entirely a function of probability transformation as per CPT principles, nor can they be attributed to the overestimation of probabilities of conjunctive events, since not only do both these phenomena have the opposite effect from that described, they are also common to both GS and MS tournaments.

As is the case in traditional finance markets, observed mispricing is positively related with trading volume, as predicted by the family of Disagreement Models proposed by behavioural finance theories (Hong and Stein, 2007), which strengthens the hypothesis of overconfidence bias as the causal factor. It has also been shown that events attracting heavy media coverage yield higher returns, although the statistical significance of the difference fades after controlling for trading volume, suggesting that the latter captures all the impact of media coverage. A negative, although not consistently significant, relationship is found between average bet size, a proxy for sophisticated bettors, and strategy returns. This could be due to the increase in arbitrage limits caused by higher noise trader risk allowing temporary persistence of mispricing. Another contributing factor is potential, albeit less significant, overconfidence bias in sophisticated traders.

Finally, comparison of the results for the main and alternative betting markets have shown that proper advantage is taken of pure arbitrage opportunities, thus precluding the possibility of making significant returns. These findings indicate that the presence of behaviourally biased agents is not sufficient to influence prices in a given market, and that limits to arbitrage are also necessary.

The results confirm that betting exchanges are reasonably efficient, given their low transaction costs and the presence of active traders exploiting arbitrage opportunities. This makes them a suitable testing ground for the analysis of investor's behaviour that are difficult to isolate in traditional financial markets, as suggested by Smith et al. (2006) or Hetherington (2006).

REFERENCES

Andrikogiannopoulou A., Papakonstantinou F. (2013) "Estimating heterogeneous risk preferences from a panel of real-world betting choices" Swiss Finance Institute Research Paper No. 13-53. Available at SSRN: http://ssrn.com/abstract=2239194 or http://dx.doi.org/10.2139/ssrn.2239194

Arkes, J., (2011) "Do Gamblers Correctly Price Momentum in NBA Betting Markets?" *Journal of Prediction Markets*, 5(1), 31-50.

Ashiya, M. (2013) "Lock! Risk free arbitrage in the Japanese racetrack betting market" *Journal of Sports Economics,* Forthcoming doi: 10.1177/1527002513493630

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

Bar-Hillel, M. (1973) "On the Subjective Probability of Compound Events" *Organizational Behavior and Human Performance,* 9, 396 – 406.

Berg, J. E., Rietz, T. A., L (2010) "Longshots, Overconfidence and Efficiency on the Iowa Electronic Market" Available at SSRN: http://ssrn.com/abstract=1645062 or http://dx.doi.org/10.2139/ssrn.1645062.

Borghesi, R., Paul, R., Weinbach, A. P., (2009) "Market Frictions and Overpriced Favorites: Evidence from Arena Football" *Applied Economics Letters*, 16(9) 903-906.

Cooper, M.J. Gutierrez, R.C. Hameed, A. (2004) "Market states and momentum" *Journal of Finance*, 59,1345-1365.

De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J. (1990) "Noise trader risk in Financial markets" *Journal of Political Economy*, 98, 703 – 738.

Daniel, K. Hirshleifer, D. Subrahmanyam, A. (1998) "Investor psychology and security market under and overreactions" *Journal of Finance*, 53, 1839 – 1885.

Docherty P. A., Easton S. A., (2012) "Market efficiency and continuous information arrival: Evidence from prediction markets" *Applied Economics*, 44, 2461-2471.

Franck, E.P., Verbeek, E., Nüesch, S., (2010) "Prediction Accuracy of Different Market Structures - Bookmakers Versus a Betting Exchange" *International Journal of Forecasting*, 26, (3) . 448-459.

Franck, E.P., Verbeek, E., Nüesch, S., (2013) "Inter-market arbitrage in betting" *Economica*, 80, 300 - 325.

Frazzini, A. (2006) "The disposition effect and under-reaction to news" *Journal of Finance*, 61, 2017 – 2046.

Hetherington, A. (2006) "Betting against efficiency: Behavioral Finance in an NFL Gambling Exchange" Working paper. Available at SSRN: http://ssrn.com/abstract=881514.

Hong, H. Scheinkman, J. Xiong, W. (2006) "Asset float and speculative bubbles" *Journal of Finance*, 61, 1073-1117.

Hong, H. Stein, J.C., (2007) "Disagreement and the Stock Market" *Journal of Economic Perspectives*, 21, 109 – 128.

Hurley, W., McDonough, L. (1995) "A note on the Hayek Hypothesis and the favourite-longshot bias in pari-mutuel betting" *American Economic Review*, 85, 949 – 955.

Innocenti, A., Nannicini, T., Ricciuti, R., (2012) "The Importance of Betting Early" LABSI Working Paper No. 37/2012. Available at SSRN: http://ssrn.com/abstract=1999459.

Jullien, B. Salanié, B. (2008) "Empirical evidence on the preferences of racetrack bettors" en Efficiency of Sport and Lottery Markets: Handbooks in Finance Series (Hausch, D.B., and Ziemba, W.T. eds; Elesevier).

Kahneman, D., Tversky, A. (1979), "Prospect Theory: An Analysis of Decision under Risk" *Econometrica*, 47, 263 – 291.

Koch, A.K. Shing, H.F. (2008) "Bookmaker and pari-mutuel betting: Is a (reverse) favourite-longshot bias built-in?" *The Journal of Prediction Markets*, 2, 29-52.

Kruglanski, A.W., Freund, T., 2003 "The freezing and unfreezing of lay-inferences: Effects of Impresional Primacy, Ethnic Stereotyping, and Numerical Anchoring", in Kruglanski, A.W., Higgins, E.T. (Eds.) *Social Psychology. A general reader.* Psychology Press.

Lahvicka, J. (2013) "What causes the favourite-longshot bias? Further evidence from tennis" Available at SSRN: http://ssrn.com/abstract=2287335 or http://dx.doi.org/10.2139/ssrn.2287335

Scheinkman, J., Xiong, W. (2003) "Overconfidence and Speculative Bubbles" *Journal of Political Economy*, 111, 1183 – 1219.

Shleifer, A. (2000) "Inefficient markets: An introduction to behavioral finance" Clarendon Lectures in Economics, Oxford University Press.

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

Snowberg, E., Wolfers, J. (2010) "Explaining the Favorite-Long Shot Bias: Is it Risk-Love or Misperceptions?" *Journal of Political Economy*, 118(4), 723-746.

Tetlock, P.C. (2004) "How efficient are information markets? Evidence from an online exchange" University of Texas Working Paper.

Tversky, A.., Kahneman, D. (1974) "Judgment under uncertainty. Heuristics and biases" *Science*, 185, 1124 – 1131.

Tversky, A., Kahneman, D. (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty" *Journal of Risk and Uncertainty*, 5, 297 – 323.

Vergin, R.C. (2001) "Overreaction in the NFL point spread market" *Applied Financial Economics*, 11, 497 – 509.

Vlastakis, N., Dotsis, G., Markellos, R.N. (2009) "How efficient is the European Football Betting Market? Evidence from arbitrage and trading strategies" *Journal of Forecasting,* 28, 426 – 444.

White, H. (1980) "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity" *Econometrica*, 48(4), 817-838.

Williams, J. (2010) "Momentum and Sports Betting" Available at SSRN: http://ssrn.com/abstract=1553150 or http://dx.doi.org/10.2139/ssrn.1553150.

TABLE 1: Sample description

This table describes the study sample. It includes the number of different odds offered in each tournament, and the number of bets. It also shows lists the tournament favourites, that is, the players with the lowest odds in each competition, and the actually winners. Panel A displays the Grand Slam data and Panel B the Masters Series data. The Grand Slam data include bets both for and against the favourite winning the tournament, that is, both the main and alternative markets. * in the Masters Series data denotes that the data for that year were not included in the study.

PANEL A: Grand Slam

Tournament	Year	Favourite	Winner	Favourite Market		Alternative Market	
				Number of odds	Number of bets	Number of odds	Number of bets
Australian Open	2005	Roger Federer	Marat Safin	93	8507	76	2875
	2006	Roger Federer	Roger Federer	73	14458	75	5080
	2007	Roger Federer	Roger Federer	54	10554	83	3686
	2008	Roger Federer	Novak Djokovic	116	14279	113	4046
Roland Garros	2005	Roger Federer	Rafael Nadal	95	12587	65	3050
	2006	Rafael Nadal	Rafael Nadal	35	6386	66	3401
	2007	Rafael Nadal	Rafael Nadal	104	20438	72	8152
	2008	Rafael Nadal	Rafael Nadal	137	22708	97	8449
US Open	2004	Roger Federer	Roger Federer	77	8373	89	3165
	2005	Roger Federer	Roger Federer	65	10125	84	3834
	2006	Roger Federer	Roger Federer	83	17826	200	8958
	2007	Roger Federer	Roger Federer	49	11653	88	6054
	2008	Roger Federer	Roger Federer	212	22232	134	12496
Wimbledon	2004	Roger Federer	Roger Federer	111	20587	75	3718
	2005	Roger Federer	Roger Federer	70	16347	67	5122
	2006	Roger Federer	Roger Federer	26	7400	74	6817
	2007	Roger Federer	Roger Federer	32	15439	60	7221
	2008	Roger Federer	Rafael Nadal	135	42053	112	14728
TOTAL				*1567*	*281952*	*1630*	*110852*

PANEL B: Master Series					
Tournament	**Year**	**Favourite**	**Winner**	**Number of odds**	**Number of bets**
Canada	2004	Roger Federer	Roger Federer	73	2788
	2005	Rafael Nadal	Rafael Nadal	95	5536
	2006	Roger Federer	Roger Federer	37	4403
	2007	Roger Federer	Novak Djokovic	35	4330
	2008	Roger Federer	Rafael Nadal	38	1147
Cincinnati	2004	Roger Federer	Andre Agassi	30	1320
	2005	Roger Federer	Roger Federer	47	3934
	2006	Roger Federer	Andy Roddick	19	1978
	2007	Roger Federer	Roger Federer	55	3899
	2008	Roger Federer	Andy Murray	86	3311
Hamburg	2005	Roger Federer	Roger Federer	83	4974
	2006	Ivan Ljubicic	Tommy Robredo	*	*
	2007	Rafael Nadal	Roger Federer	55	5906
	2008	Rafael Nadal	Rafael Nadal	167	11756
Indian Wells	2005	Roger Federer	Roger Federer	57	3645
	2006	Roger Federer	Roger Federer	27	3542
	2007	Roger Federer	Rafael Nadal	18	1299
	2008	Roger Federer	Novak Djokovic	63	2945
Madrid	2004	Tim Henman	Marat Safin	*	*
	2005	Rafael Nadal	Rafael Nadal	95	5143
	2006	Roger Federer	Roger Federer	45	3376
	2007	Roger Federer	David Nalbandian	31	3884
	2008	Roger Federer	Andy Murray	127	8204
Miami	2005	Roger Federer	Roger Federer	65	5777
	2006	Roger Federer	Roger Federer	31	4271
	2007	Roger Federer	Novak Djokovic	23	2787
	2008	Roger Federer	Nikolay Davydenko	122	5654
Montecarlo	2005	Roger Federer	Rafael Nadal	49	3650
	2006	Rafael Nadal	Rafael Nadal	30	3391
	2007	Rafael Nadal	Rafael Nadal	51	5307
	2008	Rafael Nadal	Rafael Nadal	149	9378
Paris	2004	Andy Roddick	Marat Safin	37	1300
	2005	Andy Roddick	Tomas Berdych	*	*
	2006	Roger Federer	Nicolay Davydenko	*	*
	2007	Roger Federer	David Nalbandian	44	4359
	2008	Roger Federer	Jo Wilfried Tsonga	56	2651
Roma	2005	Rafael Nadal	Rafael Nadal	76	4056
	2006	Rafael Nadal	Rafael Nadal	25	4138
	2007	Rafael Nadal	Rafael Nadal	58	9888
	2008	Rafael Nadal	Novak Djokovic	55	1594
TOTAL				*2154*	*155521*

TABLE 2: Returns of the strategy after deducting transaction costs.

This table gives the strategy returns by tournament type after deducting 2% per bet from the winnings to account for transactions costs. $R_{k,N}^{TC} = \frac{\prod_{n=1}^{N} Q(W_n/\Omega_{t_n}) - Q(W_T/\Omega_{t_0})}{Q(W_T/\Omega_{t_0})} - VTranCost$ is the return to strategy k, which is assumed to involve N simple events and to allow full implementation of the strategy. $Q(W_T/\Omega_{t_0})$ is the odds on the simple event of the favourite winning the tournament, $Q(W_n/\Omega_{t_n})$ denotes the odds on each of the N simple events of the favourite winning his n^{th} qualifying match, and VTranCost denotes total transaction costs incurred in each case. Returns were calculated for three different odds: the volume-weighted average odds, the odds with the highest betting volume, and the closing odds. The table shows the p-value from the binomial test for each type of tournament. (1) indicates the p-value from the t-test of difference of means. (2) indicates the p-value from the Mann-Whitney test. * and # denote significance at the 5% and 10% levels, respectively.

	Full Sample					
	Volume-weighted average		**Highest volume**		**Closing odds**	
Total return	0.0358		0.0680	*	0.1027	#
t-statistic	1.32		2.36		2.19	
Positive	19		24		21	
Negative	17		12		10	
Total number	36		36		31	
p-value	0.3089		0.0144	*	0.0147	*
	Grand Slam					
Total return	0.0532	*	0.0962	*	0.1764	#
t-statistic	2.5		3.92		2.14	
Positive	13		16		13	
Negative	5		2		2	
Total number	18		18		15	
p-value	0.0154	*	0.0001	*	0.0005	*
	Masters Series					
Total return	0.0184		0.0398		0.0398	
t-statistic	0.36		0.77		0.78	
Positive	6		8		8	
Negative	12		10		8	
Total number	18		18		16	
p-value	0.8811		0.5927		0.4018	
	Grand Slam - Masters Series					
Total return	0.0347		0.0564		0.1366	
p-value(1)	0.5275		0.3330		0.1283	
p-value(2)	0.114		0.071		0.075	

TABLE 3: Tournament characteristics (Grand Slam vs. Masters Series)

This table compares the betting data for the composite event "Favourite Winning the Tournament" in Grand Slam versus Masters Series tournaments. The characteristics included are number of different odds, volume-weighted average odds, the standard deviation of the odds, the skewness coefficient of the odds, total betting volume in pounds sterling, number of bets, average bet calculated as total betting volume over number of bets, and duration of the betting period, that is, the number of days on which betting on a given outcome is allowed. * and # denote significance at the 5% and 10% levels, respectively

	Average sample	**Grand Slam**	**Masters Series**	**GS - MS**	***p-value***	
Number of odds	27.72	34.89	20.56	14.33	0.007	*
Volume-weighted average odds	2.02	1.92	2.13	-0.22	0.346	
Standard Deviation	0.123	0.128	0.117	0.010	0.811	
Skewness coefficient	3.654	7.485	-0.178	7.664	0.323	
Total betting volume	925174.98	1761551.08	88798.88	1672752.19	0.000	*
Number of bets	3598.14	6433.50	762.78	5670.72	0.000	*
Average bet size	177.65	248.55	106.76	141.78	0.000	*
Duration (in days)	51.41	97.33	5.49	91.84	0.000	*

TABLE 4: Returns of the "Main" GS vs. "Alternative" GS strategy

This table compares returns based on the main betting market odds on the favourite with those based on the odds in the alternative betting market. $R_{k,N} = \frac{\prod_{n=1}^{N} Q(W_n/\Omega_{t_n}) - Q(W_T/\Omega_{t_0})}{Q(W_T/\Omega_{t_0})}$ is the return to strategy k, which is assumed to involve N simple events, and to allow full implementation of the strategy. $Q(W_T/\Omega_{t_0})$ is the odds on the simple event of the favourite winning the tournament, and $Q(W_n/\Omega_{t_n})$ denotes the odds for each of the N simple events of the favourite winning his n^{th} qualifying match. Returns were calculated for three different odds: the volume-weighted average odds, the odds with the highest betting volume, and the closing odds. The table shows the p-value from the binomial test for each type of tournament. (1) indicates the p-value from the t-test of difference of means. (2) indicates the p-value from the Mann-Whitney test. * and # denote significance at the 5% and 10% levels, respectively

	Odds		
	Volume-weighted average	**Highest volume**	**Closing odds**
	Grand Slam		
Total return	0.0759 *	0.1198 *	0.2014 *
t-statistic	3.38	4.55	2.36
Positive	14	16	13
Negative	4	2	2
Tournaments	18	18	15
p-value	0.003 *	0.000 *	0.000 *
	Alternative Grand Slam		
Total return	0.0854 *	0.1194 *	0.1706 *
t-statistic	3.31	4.92	2.75
Positive	14	16	15
Negative	4	2	3
Tournaments	18	18	18
p-value	0.003 *	0.000 *	0.000 *
	Grand Slam - Alternative Grand Slam		
Total return	-0.0095	0.0004	0.0307
p-value(1)	0.784	0.991	0.778
p-value(2)	0.728	0.899	0.745

TABLE 5: Multivariate analysis

This table shows the results of the regression $R_k = \alpha + \beta_1.GS + \beta_2 Fav + \beta_3 \ln(AvgBet) + \beta_4 \ln(Vol) + \varepsilon_k$, where Rk is the return to strategy k, GS is a dummy variable that takes a value of 1 for GS tournaments, and 0 otherwise, Fav is another dummy variable that takes a value of 1 when the volume-weighted average odds are less than 2 (that is, the market estimates the favourite to have a greater than 50% chance of winning the tournament) and 0 otherwise, ln(AvgBet) is the log of the average bet, and ln(VOL) is the log of the total betting volume. The p- values are shown in parentheses, and * and # denote coefficients that are significant at the 5% and 10% levels, respectively.

	Raw Returns			Returns after deducting transaction costs		
	Volume-weighted average odds	**Highest volume odds**	**Closing odds**	**Volume-weighted average odds**	**Highest volume odds**	**Closing odds**
α	-0.0309	0.0517	-0.0552	-0.0631	0.0188	-0.0840
p-value	(0.895)	(0.850)	(0.843)	(0.782)	(0.943)	(0.756)
GS	-0.0319	-0.0067	0.0683	-0.0291	-0.0044	0.0669
p-value	(0.669)	(0.931)	(0.305)	(0.683)	(0.952)	(0.300)
Fav	0.1607	0.1721	0.0526	0.1646	0.1766 #	0.0633
p-value	(0.148)	(0.132)	(0.599)	(0.122)	(0.098)	(0.516)
ln(AvgBet)	-0.0173	-0.2140	-0.2137 #	-0.1596	-0.1995	-0.2037 #
p-value	(0,226)	(0.149)	(0.081)	(0.241)	(0.160)	(0.083)
ln(Vol)	0.0695 #	0.0804 *	0.0971 *	0.0645 #	0.0749 #	0.0928 *
p-value	(0.089)	(0.049)	(0.027)	(0.097)	(0.057)	(0.028)
R^2	0.192	0.212	0.201	0.201	0.216	0.198

TABLE 6: Bootstrap results

This table shows the results of the bootstrap procedure, in which bootstrap samples of the same size as the original sample of tournaments were taken with replacement. The average return for each bootstrap sample was then calculated. This process was repeated 1000 times before calculating the average of the distribution of bootstrap means, and the level of significance of the null hypothesis of 0 average returns (simulated p-value). Returns were calculated for three different odds: the volume-weighted average odds, the odds with the highest betting volume, and the closing odds. Returns with and without transaction costs were also calculated.

	Raw returns			**Returns after deducting transaction costs**		
	Volume-weighted average odds	**Odds with the highest volume**	**Closing odds**	**Volume-weighted average odds**	**Odds with the highest volume**	**Closing odds**
	Grand Slam					
Return	0.0775	0.2050	0.1212	0.0532	0.1742	0.0959
Simulated p-value	0.000	0.000	0.000	0.002	0.000	0.000
	Masters Series					
Return	0.0402	0.0612	0.0619	0.0188	0.0336	0.0406
Simulated p-value	0.422	0.142	0.214	0.756	0.434	0.434