Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Facultad de Ciencias Económicas y Empresariales

TRABAJO FIN DE GRADO EN
(Doble Grado Internacional en Administración de Empresas y en Economía)

BEHAVIOUR OF THE OVER-ROUND IN TENNIS BETTING MARKETS

Pamplona-Iruña, 18 de mayo de 2019

Autor: Joxe Mari Barrutiabengoa Ortubai

Tutor: Luis Fernando Muga Caperos

Módulo: Finanzas

ABSTRACT

By understanding how the over-round is constructed, this work draws a parallel between financial and sport betting markets, where the over-rounds work analogously to bid-ask spreads in stock exchanges. As in financial markets, when the probability of facing better informed investors increases, bookmakers increase the over-rounds or market spreads in sport betting markets. In this sense, assuming that in the WTA circuit the adverse selection costs are higher than in the ATP circuit, the paper examines if over-rounds in WTA games are higher than in ATP games. The logic works as follows; if less information on WTA players is publicly available, is not unreasonable to think that the chances of confronting a better-informed bettor are higher for bookmakers in women's tennis games (information asymmetries), leading us to believe that WTA games should exhibit higher spreads than ATP games. In short, the study confirms this hypothesis and it displays significant evidence suggesting that in professional women's tennis games, bookmakers impose higher over-rounds than in professional men's tennis games. Furthermore, based on the study's findings, the paper presents a concluding discussion on the wage gap between men and women professional tennis players.

Keywords: Information asymmetries, Betting markets, Regression analysis, Bid-ask spreads, Adverse-selection costs

TABLE OF CONTENTS

1. INTRODUCTION

The sport betting industry has suffered an unprecedented expansion during the last decade. The boom in online betting, together with scarce regulation in many countries and a huge investment in publicity, have exponentially increased its public, becoming addiction problems more and more common through the population. For sports' fans, is not unusual to see how Cristiano Ronaldo, Rafael Nadal, or Neymar Jr. collaborate in numerous betting ads, encouraging viewers to bet in their next match.

In the next lines, an attempt to combine the worlds of sport and finance has been made, making a brief overview of the industry and understanding how the "Over-round" is constructed in betting markets. In price-driven markets, when the probability of facing better informed investors increases, market-makers increase the bid-ask spread (the difference between the buying and selling price of a security or Over-round), with the intention of protecting themselves against huge loses.

Analyzing the over-round differences between the games played by ATP and WTA players, and assuming that, among others, the WTA circuit creates less media attention, the level of compliance of financial theory in betting markets has been tested. It should be pointed out that WTA stands for Women's Tennis Association, while ATP implies Association of Tennis Players (male players). So, if a significant over-round difference exists between men and women matches (higher over-round for women's games), as long as we have considered that the probability to find investors better informed than bookmakers is higher in the WTA circuit, evidence will indicate that financial theory also applies in betting markets. Take into account that betting markets are comparable to price-driven markets, as market-makers have the duty to ensure market liquidity, determining quotes.

In the past, other authors as Shin (1991) argued that bookmaker's over-rounds can be interpreted as an analogy of the bid-ask spread in financial markets. The logic works as follows: Betting markets correspond to a market for contingent claims with n states. In this market, the value of the securities is determined by the betting odds (implied probabilities), and the sum of all the implied probabilities (all possible outcomes) should be equal to 1. So, if the sum of the implied probabilities is higher than one, an over-round or market spread exists (Coleman, 2007; Law and Peel, 2002), providing a clear, unambiguous, and an accessible measure of the size of the market spread.

So, in the study, after a preliminary analysis of the data, three different econometric models have been constructed, in order to see if significant over-round differences exist among the econometric model for men, the econometric model for women, and the final model that encompasses both. Results will be discussed in more detail in section [7], but in general, the models prove that ceteris paribus, significant over-round differences exist between men and women professional tennis games, being the spread higher for women matches. This result can be in part explained by the fact that less public information is available on women tennis players, especially for the ones that are outside the top 50. This increases the possibility of investors to have private information that escapes the knowledge of market-makers, increasing bookmakers' expected mark-to-market loses, and in consequence, as it happens in financial markets, increasing the spread. Furthermore, as unsophisticated or passion investors are less frequent in WTA games, again, the chances of facing institutional and better-informed counterparties increase in women tennis. As an example, analyzing Betfair's betting exchange data[1], it can be seen that in Wimbledon 2016, although the total bet volume was considerably higher in men's games (98,023,248.20 € against 53,041,054.94 €) the average bet size was higher in women's matches (459.91€ against 415.89€). Hence, data suggests that the proportion of institutional investors in betting exchanges may be higher in women's tennis matches. In this regard, Lin et al. (1995) re-examine the relationship between the trade size and the components of the bid ask spread, finding a positive relationship between the trade size and the adverse selection component.

In short, it seems that in the WTA circuit, the adverse selection costs are higher than in the ATP circuit, being information asymmetries the most plausible cause to explain the difference. Remember that in financial theory, bid-ask spreads are affected by 3 main factors; inventory costs, operating expenses, and dealers' risk of transacting with better-informed clients (Levin and Wright, 2004). In this sense, given that no stock must be kept in inventory when trading on probability, it could be assumed that for betting firms, inventory costs are equal to 0.

This research contributes to the existing literature in two ways. First, it proves that an over-round difference exists between ATP and WTA games, and it cites higher information asymmetries as the most plausible explanation. Second, previous studies have documented major similarities between betting and financial markets, and this paper also contributes in this regard: As in financial markets (an increase in the spread could be observed), bookmakers increase the over-round when the chances of confronting better-informed counterparties increase.

The paper is organized as follows. First, in section [2], an overview of the sport betting industry is made, analyzing, at the same time, which are the principal characteristics of price

---

[1] Data reflects betting volume from the quarterfinals to the final. See Appendix 1 for further results.

and order-driven markets and how odds and prices are stablished in bookmaker's markets. Second, in sections [3] and [4], an explanation is given to clarify which are the variables that influence price formation in betting markets, and continuedly, the main hypothesis of the research is presented. In the following sections (sections [5] and [6]), an overview of the dataset and the methodology is made, explaining the steps taken to contrast the ideas presented in previous sections. Third, in section [7], in order to obtain further evidence in favor of the hypothesis, a bunch of econometric models are estimated, together with the respective analysis of the results. To conclude with the study, the main findings, the final remarks, and the future research pathways are highlighted in section [8], finalizing with a closing discussion on the wage gap between men and women professional tennis players in section [9].

2. BETTING MARKETS

2.1 Sport betting industry

The sport betting industry, or the activity of trying to predict sporting outcomes and placing a positive economic quantity in the guess, has always been considered a social instrument. On the one hand, for high standing people, bets have been an excuse to foment social gatherings and interactions, and on the other hand, for low class people, gambling has been the easiest way to experience emotions and to target a new life. For example, betting was common in Ancient Rome, where people were used to place bets on chariot races or in combats between the fiercest gladiators. Nevertheless, the betting industry has always been a double-edged sword. Since the very beginning, social conflicts were common in casino's areas, and people suffered from economic instability due to the high amounts of money that they spent in the casinos.

In this situation, along with the creation of the first national lotteries, the most advanced nations of the world started to regulate the betting industry at the end of 18th century, in an attempt to obtain new revenue flows, to fairly distribute the prizes, and to take care of the public order and health (Schwartz, 2013). For example, in some countries like China and U.S.A., betting was restricted to certain areas (Macao and Las Vegas), but the legality of the activity was never put into question. In this sense, national regulatory differences have limited the ability of countries to result in a market segmentation, or at least, to define similar regulatory barriers, and in spite of agreements, legal niches have allowed sport betting firms to internationalized distribution (Gomber et al., 2008).

However, the real revolution in the sport betting industry took place recently, more concretely at the beginning of the 21st century, with the arrival of the internet. In fact, it was not until 2008 when sport betting firms obtained the first licenses to operate in the communities of Madrid and the Basque Country (Spain). Brindley (1999), already predicted that synergies between the gambling industry and the internet would change supply and consumption, due to the fact that in 1999, the sport betting markets were dominated by odds that were almost impossible to compare. In other words, due to physical distance, bettors only had access to one or two locations where bets could be placed. So, physical and time barriers restrained the capacity of the investment public to compare bookmarkers' quotes, and in consequence, competition between betting firms was almost inexistent.

In an online world, the betting industry has acquired the capacity to break unimaginable barriers and to connect with new publics, especially youth from 20 to 29 years who were inaccessible some years ago. Solely in Spain, the sport betting industry has more than multiplied its revenues in the last 6 years, and from 2012 onwards, it has experimented annual gain increases of 20% (Expansión, 2018). To put it into perspective, the volume of the sport betting industry in Spain in the year 2010 was 742 million euros, slightly lower than the annual revenue of Real Madrid, that consist on 750 million euros.

Thanks to their aggressive marketing strategies, sport betting firms are able to increase its revenues year after year, and the prospects are even more promising. Betting firms are attracting more and more young public, betting on a game is not socially rejected, and thanks to their "live" distinctive characteristic (people have the possibility to bet on a game while they are seeing it), satisfaction is more intense. However, the increase in sport betting volumes is raising some new concerns in the old continent, with Italy and Spain, for example, imposing more restricted regulatory norms regarding online betting.

So, based on the relevance that the sport betting industry has right now, the aim of this work is to understand how the over-round evolves in betting markets, analyzing if significant differences exist among women and men tennis games. In a world where the sport betting industry is gradually gaining importance, research and empirical work are necessary for the correct understanding of the industry and to facilitate future regulation. With all the required humbleness, this work tries to row in that direction.

2.2 Price and order-driven markets

Price-driven markets, also known as quote-driven markets, are electronic stock exchange systems in which prices are determined from bid and ask quotations made by market-makers, dealers, or specialists (Investopedia, 2018). Put another way, market-makers create the market by quoting, buying, and selling orders, and establishing the maximum quantity that they are willing to offer at the quoted price. In a quote-driven market, dealers fill orders from their own inventory or by matching them with other orders.

On the other hand, order-driven markets are opposite to price-driven markets, as all buyers and sellers publicly display the prices at which they want to buy or sell a security and the desired quantity. In short, the main advantage that price-driven markets have over order-driven markets is that order execution is guaranteed, as market-makers have the obligation to meet the quoted bid and ask prices (Demsetz, 1968). However, although less liquid, order-driven markets are more transparent. As it has been mentioned, orders of both, buyers and sellers are publicly shown, and in consequence, all the market orders and the prices at which investors are hoping to buy or sell the security are displayed.

Regarding betting markets, since the beginning of the 2000s, betting markets have been characterized by the coexistence of order-driven and price-driven markets (Flepp et al., 2017). As in financial markets, in price-driven betting markets, market-makers operate on their account by betting odds at which bettor can place their bets (Croxson and Reade, 2013), whereas in order-driven betting markets, betting exchanges act as a market place in which buy or sell orders are matched in a continuous double auction. In other words, by guaranteeing liquidity at the odds placed, price-driven betting markets increase rapidity and reduce the gap that arises from the different arrival rate of buyers and sellers, while in order-driven betting markets, liquidity is provided by the flow of orders from market participants (De Jong and Rindi, 2009).

So, like market-makers in financial markets, in the betting industry, bookmakers (e.g. Bwin, Ladbrokers or William Hill) serve as intermediaries between the investors that want to place a bet on particular outcome (buyers), and the rest of the people that want to invest on the opposite result (sellers). Gomber et al. (2008) defined the bookmaker market model as a bilateral dealer market, where the bookmaker offers the traditional way to place a bet. In this extent, by determining the price at which they are willing to accept bets (on the outcome of certain sport event), bookmakers unilaterally determine the odds for a given outcome, and they earn a commission, which is known as the spread or the "Over-round" (Harris, 2003). The over-round, which is already priced into the odds, compensate them for providing liquidity and assuming the risk of an unfavorable outcome. So, in short, bookmaker's markets are comparable to price-driven markets, as prices are established from quotes (bid-ask) made by market-makers. On the other hand, one significant difference between price-driven financial and betting markets is that more information is available in sports betting markets. Rothschild described it in the following way: "While publicly available sports statistics are very deep, in financial markets there is more hidden, idiosyncratic information that investors have to gather" (Institutional investor, 2017). Furthermore, sport betting is purely organized "Over the Counter" (Gomber et al., 2008). Put another way, investor's protection may be lower than in financial markets, as a centralized regulated market does not exist, and the trades are completed via a dealer's network.

Previous studies have obtained evidence suggesting that betting exchanges (order-driven markets) suffer less operational risk (Koning and Van Velzen, 2009), have a higher prediction accuracy in their odds (Franck et al., 2010; Smith et al., 2009), and bear lower information costs (Davies et al., 2005). However, the most usual form of betting, especially for small investors, is still through bookmakers, that year after year, continue being successful.

For the porpoises of this work, as long as our aim is to confirm the hypothesis that due to higher adverse selection costs, women's tennis games should exhibit higher "over-rounds" or spreads, betting exchanges will be intentionally excluded from the analysis.

2.3. Odds and prices in bookmaker's markets

In betting markets, the traded instruments are bets. As it happens with derivatives in financial markets, bets represent a contingent contractual claim on a future cash flow (Flepp et al., 2017). Put another way, bets are rights that could possibly arise (at the moment that the outcome is known) to demand future cash flows. With contingent, the intention is to express that is subject to change.

At the same time, the cash flow is influenced by two different parameters: the outcome of the underlying asset, and the price of the contract or the odds (Sauer, 1998). In this regard, if the bookmakers set a price p for a concrete event, the probabilities that the bookmaker assigns to each event are also displayed (Probability of an event = $1/p_{\text{event}}$). Remember that the bookmaker determines the probability of an event and it takes the opposite position in every transaction (Franck et al., 2010). In this extent, Forrest and Simmons (2008), and Levitt (2004) used bookmaker's quotes to calculate the probabilities of victory, and Hvattum (2013) and Sauer (1998) also made use of them to understand how the markets work.

The probabilities are public before the event takes place, "ex ante", and they evolve through time. Furthermore, the investor has the possibility to accept the quote or to refrain at every moment before the event is finished. So, in bookmaker's markets, the investors know that the bookmaker is always on the other side of the bet, ensuring payments (Elizalde, 2015), and as in financial markets, bookmakers change the price of an event when new information is available to the public (a goal, a major injury, or an event that may have an influence on the final result of the game).

Regarding the gains of the bookmaker, the way of calculating them will vary depending on the strategy that the firm decides to take: If the probabilities are solely based on the real probabilities of the event (imagine that p represent the fundamental value of an event), the bookmaker will win, on average, a benefit equal to the over-round that is imposed to market participants (in the analogous way in which market-makers obtain their benefit from bid-ask spreads). However, if the bookmaker is able to guarantee the same volume of bets against and in favor of an event, the bookmaker will always be able of paying the successful investors with the unsuccessful ones, and in consequence, win the over-round in every occasion, without any regard to the outcome. Nevertheless, if we tend to infinity, the bookmaker's profit will be equal in both ways.

In tennis games, if the spread is not taken into consideration, the quoted probability (the inverse of the odds) that player x or player y have to win the game is 100%, and therefore, this should be reflected in the prices. In other words, if the bookmaker pays 2,5€ for every euro invested in the victory of player x (1/2,5=40%), it needs to pay (1/0,6) 1.67€ for every euro invested in the victory of y (only two possible outcomes, the victory of player y or x). Being more precise, as long as the assigned probability of winning the game is higher for player y, player y will be the favorite. On the other hand, player x will be the long shot. So, as it can be observed in the previous example, betting firms establish higher prices for bets in favor of the favorite. However, as their goal is to obtain a profit, the outcome of the event is uncertain, and information asymmetries could be present in the market, bookmakers always include a spread (over-round) in their quotes.

Extending the previous example, imagine that the bookmaker takes the decision to include a 5% over-round into the quotes (distributed equally). In this situation, the quotes will be the following:

2.5/1.05 = 2.38 for player x

1.67/1.05 = 1.59 for player y

As it can be seen, if the implied probability is calculated, (1/2.38095) %42 + (1/1.59) %62.9, the sum is approximately 105%. This 5% is known as the over-round.

So, the over-round can be defined as a price indicator that reflects the probability percentage that is above 1. Put another way, the over-round is the inverse sum of the prices of the different outcomes that exist in an event. The over-round is strictly positive, and it only takes negative values (sure bets) when an error is in between. In that way, the bigger the over-round is, with all the rest equal, the more profit for betting firms, and the less for bettors.

In this extent, Hvattum (2013) suggested that the competitive positioning of bookmakers, along with the information that they have, are essential to understand how over-rounds are constructed. In mathematical terms, as it has been highlighted, the over-round is the difference between the sum of the inverse of the odds and one:

$$[1] \ \ \lambda_{mi} = \Sigma_j \ (\frac{1}{p_{mij}} - 1)$$

m stands for match m, i stands for bookmaker i, and j refers to the odd status (player *x* wins or player *y* wins). So, without further prove, it can be said that the shorter the odds, the higher the over-round, and consequently, the higher the margin of the bookmaker (Deschamps and Gergaud, 2007).

In this context, a widely studied phenomena in betting markets is the Favorite-Longshot Bias (FLB hereafter), or the tendency to overvalue "longshots" and undervalue favorites (see Cain et al., 2010; Williams and Paton, 1997). Put another way, the FLB is the longstanding empirical regularity that betting odds provide biased estimates of the probabilities of the sports' outcomes; longshots are over bet, while favorites are under bet (Snowberg and Wolfers, 2010). In consequence, empirical evidence indicates that betting on the favorite is a much better idea than betting on the long shot. For example, in the long run, losing 5% by betting on the favorite, but losing 40% on longshots is not uncommon (Sobel and Raines, 2003). However, people remain ignoring the evidence and willingly betting on the long shot. Risk-loving behavior, misperceptions of probabilities, and irrational behavior are some of the possible answers to this phenomenon that can be found in the literature. In order to better understand the FLB, see the following example:

As we have seen in the previous illustration, quotes change from 2.5 (for player *x*) and 1.67 (for player *y*) to 2.38 and 1.59 when a 5% over-round is included (assuming that the margin is equally spread across each player). However, especially when a clear favorite exists, this is not what happens. In betting markets, in terms of margin percentage, the over-round is distributed unequally, and therefore, quotes like the following are easier to find:

1.67/1.01 = 1.653
2.5/1.1126 = 2.247

The spread is still 5%, (1/1.653) 60.5% + (1/2.247) 44.5%, but the implicit probability of victory of the long shot is now overvalued (42% vs 44,5%). The FLB is more pronounce in markets with higher trading volumes, heavier attention on the favorite, and less sophisticate and informed investors (Abinzano et al., 2017). Finally, it is important to mention that the FLB is inconsistent with the widely known decision-making model that Kahneman and Tversky presented in 1979, when they affirmed that a natural tendency to avoid a loss rather than make a gain exists among humans (also known as loss aversion).

3. VARIABLES THAT AFFECT PRICE FORMATION

Price formation in betting markets have two distinctive characteristics. First, unlike in financial markets, sports bets are completely idiosyncratic, meaning that they have no relation to any risk premia or aggregate risk (Moskowitz, 2015). Remember that in the Capital Asset Pricing Model developed by Sharpe (1964), the choice between the potential risk and return of a portfolio is explained by the risk premium that investors bear. Furthermore, as Williams (1999) mentioned, in contrast to stocks, the contract of a sport bet is related to a single event, and in consequence, is not subject to other considerations as the future performance or the future cash flows. Second, sports contracts are very short in time, and in consequence, as uncertainty disappears very quickly, mispricing can be easily detected (Moskowitz, 2015). When considering market efficiency of sport betting markets, it must be taken into account that betting is a zero-sum game, where bookmakers try to earn a profit in the long run. So, if bettors are able to make a profit, the prices cannot be considered as efficient, as long as it can be inferred that all bettors or bookmakers are not well informed (Sauer, 1998).

In this sense, sports betting contracts should be, in theory, subject to the same behavioural tendencies or biases that influence market anomalies in financial markets. Barberis and Thaler (2003) defined behavioural finance as the study of how irrational behaviour influences market prices (deviation of rational thinking and exposure to unnecessary risks), moving them from their intrinsic values. Furthermore, the rational expectation utility model is constructed around generic risky gambles, so it should apply for both, sports bets and capital market securities. On this subject, Franck et al. (2010) suggested that the presence of irrational bettors can lead bookmakers to bias their betting odds. However, the authors recognized the existence of large gaps in the analysis of the impact of trading on price setting, as obtaining information on trading activity is still complicated.

Regarding price formation in betting markets, Elizalde (2015) mentioned that among others, two main factors stand out in the process: Information and liquidity.

On the one hand, the possibility of facing better informed market participants influences prices. Thanks to information asymmetries, bettors may have advance information on the fundamental value of the asset, and in consequence, they may be able to obtain an easy profit. Hence, bookmakers must consider the probability to encounter with a better-informed bettor as a counterparty, and therefore, of incurring in loses. Remember that the bookmaker obtains a benefit when is trading against worse informed investors and from the bid-ask spread, so, if more better-informed investors are active in the market, the bid-ask spread also increases, worsening price discovery.

Furthermore, following the same logic, bet prices also suffer alterations when new information arrives to betting markets. In this way, if a sufficient number of bets of the same outcome are accepted, the bookmaker protects itself modifying its quotes (reducing the price and therefore increasing the implicit probability of the outcome). The same happens if enough bets (with contrarian outcome) are accepted. Take into account that in the analysis presented in the next lines, only two possible outcomes exist (either player x or y wins), so modifications on the price of x (the amount of euros that will be obtained for every euro invested in the victory of x), implicitly represent a change in the probability that the bookmaker assigns to the victory of player y. So, as bookmakers change their expectations on the value of the asset with the arrival of new information, quotes on the victory of player x and y will be changing, depending on the volume of bets that each player receives.

Moreover, as long as the outcome is more uncertain, the over-round is higher for games that are played on first rounds and in low level tournaments. This happens because the bookmaker's possibilities of suffering an undesirable outcome increases in these situations, as other investors may have more concrete or better information about the health of the players, how motivated they are, or if they have had any family issue recently (further explanations about the phenomena are given in following sections [4] and [5]). In other words, the over-round, that can be interpreted as an analogous of the bid-ask spread, will be higher when private information is not reflected in the prices. Remember that the bid-ask spread is an accepted measure of liquidity costs, and that as part of its framework, Fama

(1970) highlighted the notion of an efficient market, underwriting that an efficient market is a market in which securities' prices reflect all available information. In this extent, if the bookmakers increase the over-round in order to mitigate loses from information asymmetries, it can be said that the efficient-market hypothesis is rejected for bookmaker markets.

On the other hand, liquidity, or to allow assets to be bought and sold at stable prices, also affects price formation in betting markets. This is better understood with an example:

In betting markets, liquidity varies depending on the type of game. Put another way, the maximum number of euros that can be bet on the outcome of an event changes with the market of the respective game and its depth. As an example, is not the same to place a bet in the Roland Garros final, where the maximum permitted bet will be above 10000€, or to place a bet in the first round of Pune (ATP 250), where the maximum bet will be around 200€. So, limitations to execute the desired bet also vary depending on the market of the game. This is again connected with the fact that bookmarkers have the need to protect themselves against frauds created by information asymmetries. Regarding prices, if the market depth is thin, a small order may be sufficient to alter the price of an outcome (e.g. player *x* or *y* wins).

All in all, Moskowitz (2015) suggested that betting firms set an initial price or line on each contract. These initial prices are established in order to maximize their benefits, considering risks, and equalizing the dollar bets on each side of the contract. With this method, betting firms receive the spread with no risk exposure. However, if bookmarkers are on average better than gamblers at predicting results or betting volumes, they also have the possibility of increasing their risk exposure, and therefore, of increasing their profits. Once the initial price is established, as betting volume flows, the initial price is subject to variations when bookmakers try to balance their risk exposure. On the other hand, bettors also have the possibility to bet before and during the game, and as a result, until the final closing price is established, they constantly try to exploit mispriced quotes (old quotes that do not reflect an injury, the mood of the players, or other circumstances that can influence the result of the match).

4. HYPOTHESIS

As it has been mentioned in previous lines of the work, the objective of this paper is to determine if the over-round is higher in WTA games than in ATP games, understanding at the same time, which are the possible triggers behind this difference.

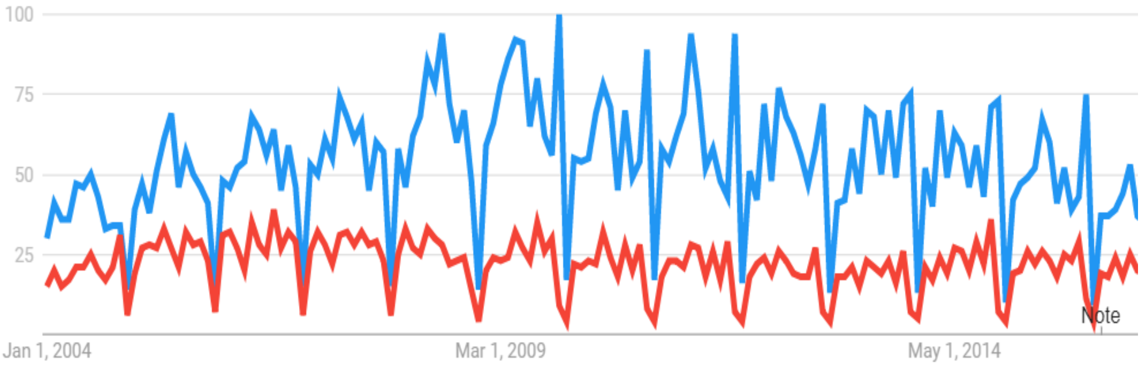
In this extent, the null hypothesis (H_0) is that bookmakers impose the same over-round in WTA and ATP games. However, as our main hypothesis is that due to higher information asymmetries, women's tennis games should exhibit higher over-rounds than men's tennis games, we expect to reject the null hypothesis.

The selection of tennis as the test group is based on two main reasons: First, there are many players on the ATP and WTA tour, a large number of games are played during a season, there are no draws, and as Forrest and McHale (2007) mentioned, the structure of the tournaments facilitates matches between players with very different rankings, thus leading to a wide range of available odds. Second, abundant public information about professional tennis players and tournaments is publicly available, easing the extraction of conclusions.

So, the logic behind the main hypothesis works in the following way: Public information about female tennis players is less accessible for bettors, as the WTA world tour receives less media attention, attendance to WTA tennis games is lower, and in consequence, less resources are devoted to trace the circuit. This creates an environment where private information is easier to obtain for investors (in comparison to the ATP circuit), and in consequence, bookmakers face a higher risk of suffering adverse selection costs. In other words, it is more plausible for an investor to obtain an information that the bookmaker does not receive.

The following graph illustrates the worldwide web search interest relative to the highest point on the chart for a given time. The graph has been obtained from Google Trends and a value of 100 is the peak popularity for the term. As it can be seen in the graph, the term ATP (in blue) arouses more interest in terms of searches than the WTA (in red) for the whole period. This is consistent with the believe that more information is available for male tennis, and in consequence, market participants are in general better informed, leaving less room for profitable private information.

Figure 1. Public interest on the ATP circuit vs the WTA circuit. Source: Google Trends.



The same pattern repeats if we look to television audience measurements or to the number of bets placed in each tournament. As an example, in the year 2015, the ATP circuit had 973 million viewers in comparison to the 395 million viewers that the WTA circuit had. Furthermore, in the year 2016, 351,025 bets were placed at the betting exchange Betfair in Wimbledon, of which 23,5696 were placed in games played among men players, and 11,5329 in games played between female players (see appendix 1). A considerable difference.

In this extent, if less information is publicly available for WTA players, is not unreasonable to think that the chances of facing better-informed bettors than themselves increase for bookmakers in the WTA circuit, and in consequence, the spreads will be higher than in the ATP circuit. In other words, if the bookmaker does not impose a higher over-round in women's games, their per game gains will be lower than in men's matches, as the possibilities of incorporating private information in the bets, capitalizing the superior information, will be higher for bettors. Financial literature has proven that superior information can be used to create profitable strategies in price-driven markets (see Grinblatt and Titman, 1994; Daniel et al., 1997).

However, to ensure the meaningfulness of the theory, the difference in the spread (due to a higher probability of facing information asymmetries) must occur with the rest of the variables equal. So, complying with previous literature, at the time of constructing the regression models, other variables that create variations in the over-round have to be included; the round number (Lahvička, 2014; Abinzano et al., 2017), the type of tournament (Lahvička, 2014; Abinzano et al., 2017), the yearly variations due to technology improvements (Hidalgo et al., 2016), the ranking and point differences between the players (Lyócsa and Výrost, 2017; Moskowitz, 2015), and if at least one of the players is among the best 50 of the world (Lahvička, 2014; Abinzano et al., 2017). If controlling for these variables, a significant over-round difference is still appreciable between men's and women's tennis matches, it can be concluded that the analysis gives favourable evidence in favour of the hypothesis that higher information asymmetries create an over-round difference between ATP and WTA games, and in consequence, financial and betting markets present major similarities at the time of fixing the spreads.

5. DATABASE

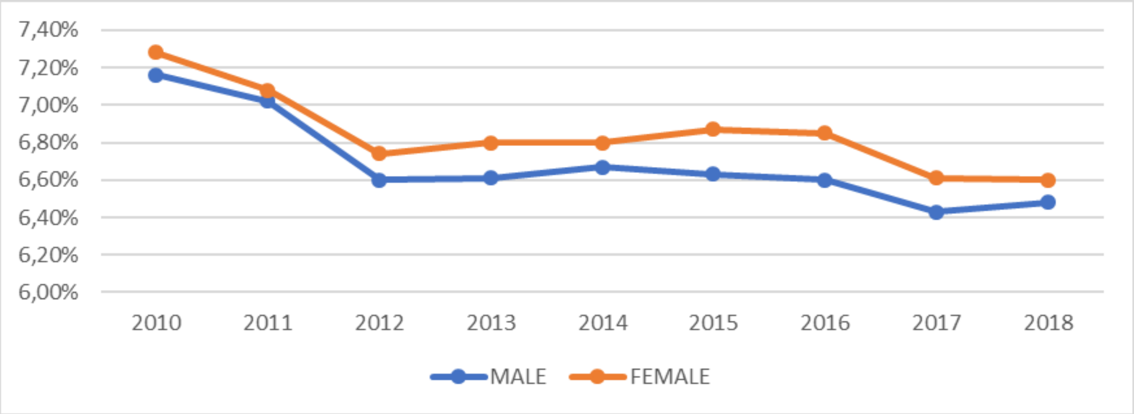
The sample consist on 45,661 tennis matches played on the ATP and WTA world tours from 2010 to 2018. 23,568 games were played among men, while the other 22,093 games were played between women. The data were retrieved from www.tennisdata.co.uk, a source used in a number of studies covering the tennis betting market, e.g. Forrest and McHale (2007), Scheibehenne and Bröder (2007), McHale and Morton (2011), Baker and McHale (2013), or Brown and Minor (2014).

The dataset is composed with the basic match information that is meaningful for the purpose of this study, including the ranking of the players, the ranking points that each player has, the match results measured as the number of games won, the type of tournament, the date of the match, the round number, and the closing odds quoted from Bet365 for a given match. All the games that had some blank data have been automatically discarded. Also, by applying Equation [1], the over-round values for each match have been obtained.

Regarding the ranking points, it is convenient to remember that they are awarded according to the type of tournament and the stage of tournament reached. Furthermore, the rankings are updated weekly, so players go up and down in the ranking depending on their performance. Being more precise, after every tournament, the rankings drop all the points earned in the previous year at the respective tournament and replace them with the points won in the just ended contest.

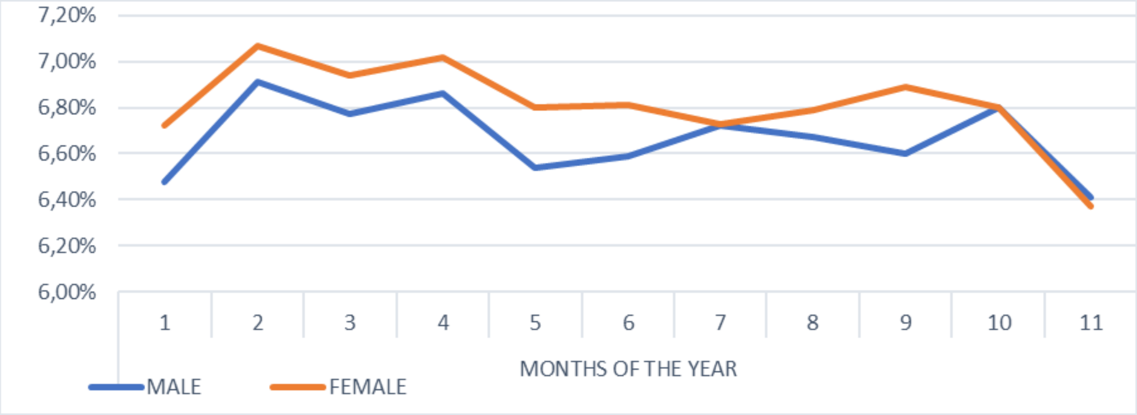
Figure 2 shows the evolution of the over-round for women's and men's world tour matches from 2010 to 2018. As it can be seen in the data, the over-round has experienced a sharp decrease in the last years, both for men and women. This is consistent with the findings of Hidalgo et al. (2016) who argued that on average, football bets have experienced an average over-round reduction of 40 percent in the last decade. The gradual expansion of online betting and the appearance of new competitors, have significantly increased competition among sport betting firms, and in consequence, over-rounds have decreased. This is also consistent with financial theory. Among others, Huang (2002), Mayhew (2002), Brogaard et al. (2014) and Battalio et al. (1997), obtained favourable evidence suggesting that competition in providing liquidity tights the spreads in financial markets.

Figure 2. Evolution of the average over-round from 2010 to 2018, both for men and women.



As shown in Figure 3, the average over-round also varies depending on the month of the year that the game takes place, both for men and women. For the purposes of this study, and in line with standard financial market practices, December matches have been dropped from the sample, as long as the matches played in this month are residual. In this extent, the higher over-rounds can be found in the period from February to April, and although a variety of variables could create these variations, the scarce number of prestigious tournaments that take place in this period in comparison to the rest of the year, emerges as the most plausible trigger of the difference. Furthermore, as the period coincides with the months that are between the Australian Open and Roland Garros Grand Slams, it could be think that the uncertainty regarding the performance and the interest of the players in the tournaments is higher than usual. Abinzano et al., (2017) informed about the relevance of market uncertainty in the price-setting process.

Figure 3. Average over-round in the different months of the year (2010-2018).



Furthermore, Table 1 shows that the average over-round (in percentage), is lower for men's games than for women's games. Further evidence will be needed to prove that this difference can be in part explained by the higher information asymmetries that bookmakers may encounter in the WTA circuit, but at first sight, it is evident that an over-round difference exist between the WTA and ATP world tours (6.85% vs 6.69%). Table 1 also shows that the over-round decreases with the importance of the tournament, with a considerable difference between grand slam tournaments (GS) and the rest. In order to make the results of women and men comparable, an equivalence between tournaments have been made, as long as WTA and ATP circuits do not share the same tournament hierarchy. For that, monetary rewards, stadium attendances, and television audiences have been examined, leading to the following equivalences:

- ATP 250 and 500 – WTA International
- ATP Master 1000 – WTA Premier
- ATP Master Cup – WTA Finals

On this subject, Grand Slam tournaments are almost identical for both sexes, with only one significant difference: Women's matches are played to the best of three sets, while men's play best of five sets. This could create problems at the time of analysing over-round differences between female and male GS matches, as the over-round variations may be built on the tournament design difference. Further studies may be necessary in the future to understand the influence that the tournament design may have on the observed over-round.

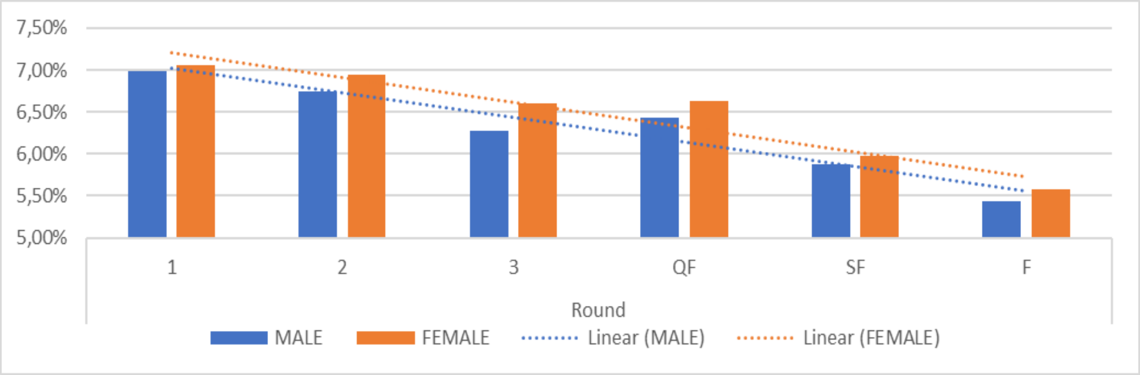
A possible explanation to the over-round difference between GS and non-GS tournaments is that players tend to free ride less in GS tournaments, and therefore, suspicious behaviours are less frequent. For example, the data shows that male favourites (higher assigned probability of victory) won the 77.03% of the GS matches, while they only succeeded in the 67.95% of the non-GS games. A striking difference that is replicated in women's games, where the favourites won the 71.88% of the games in GS matches, whereas they only defeated the long shot at the 65.49% of the times in non-GS matches. Again, the difference in the proportion of success between women and men favourites at grand slams, may be partially built on the reality that women's games are played to the best of 3 sets, leaving room for more surprising results. However, is also important to mention that from 2010 to 2018, with Rafael Nadal, Roger Federer, Novak Djokovic, and Andy Murray, the GS tournaments have witnessed an almost tyrannical superiority of these players, while the WTA circuit, by not having such superior players except from Serena Williams, has given room for a higher number of unexpected GS champions and results.

The over-round is even smaller in the Master Cup (males) and WTA Finals tournaments (5.57% and 5.98%), but due to the particularities that both tournaments entail (only the players that are ranked among the top 8 participate), both have been excluded from the analysis. Moreover, Figure 4 shows that as players advance in the tournaments, the over-round decreases gradually, finding the most important differences in the first and the second rounds. This could be explained by the fact that bookmakers receive key information in the first rounds of the tournament: If the player is taking the tournament seriously, and which is her/his physical and emotional condition. Put another way, the bookmaker may suffer more exposure to information asymmetries at the early stages of the tournament, and in consequence, it increases the over-round. At the same time, more low rank players participate in first rounds, and in consequence, as information is scarcer for these players, bookmakers protect themselves against loses increasing the over-round. Usually, when the last games of the tournaments are played, bookmakers have more information to impound on prices, and therefore, the market becomes more efficient.

Table 1. Average over-round by tournament type (2010-2018).

		Tournament type				
	2010-2018	**250**	**500**	**1000**	**GS**	**MC**
Male	6.69%	6.97%	6.79%	6.75%	5.95%	5.57%
Female	6.85%	7.03%		6.99%	6.26%	5.98%

Figure 4. Average over-round by round (2010-2018).



Finally, Table 2 contains the average over-round differences for matches that were played between two top 25 players, two players that were ranked among the 25 and the 50, and two players that were between the 100 and 500 positions in the ranking. As it can be seen in the table, over-rounds are lower when two high ranking players face each other, both for women and men tennis players.

This could be explained by the fact that almost no private information exists on high ranking players, and in consequence, bookmakers can accurately predict the fundamental values of the bets. Furthermore, a higher volume of bets is received when two top players face each other, improving the liquidity of the markets. The logic works as follows: Bettors, as a group, are more willing to bet a higher amount of money on both outcomes of the game, and as a consequence, bookmakers are more able to balance their risk exposure. This is consistent with financial theory, which affirms that stocks and indexes with higher trading volumes have narrower bid-ask spreads than those that are infrequently traded, since a broker requires more compensation for handling the transaction.

Table 2. Average over-round by the players ranking position (2010-2018).

	Ranking: 1-25	**Ranking: 26-50**	**Ranking: 100-500**
Male	6.05%	7.26%	7.32%
Female	6.46%	7.28%	7.40%

In short, it seems that the preliminary analysis of the data gives evidence in favour of the main hypothesis that the over-round is higher in the WTA circuit, as the over-rounds have been higher in every analysed situation. Furthermore, data suggests that bookmaker's exposure to information asymmetries could be key to understand over-round differences.

6. METHODOLOGY

6.1 Univariate analysis

As it has been shown in the previous section, differences exist between the sample means that have been analysed. However, in order to determine if the disparities between the sample means are significant enough to consider that a difference also exist in the population means, t test analysis have been conducted. The results of the t-test are also interesting to conclude which variables should be included in the model. Put another way, the objective is to determine if the average over-rounds of the populations (see male vs female matches, GS vs non-GS matches, 1st round vs non 1st round matches…) are different, not just if the sample means are dissimilar. Small differences between the sample means may be caused by sampling variability.

In concordance with the preliminary analysis and previous studies (see Lahvicka, 2014; Abinzano et al., 2017; Moskowitz, 2015), the tests have been performed for the following variables:

1. Grand Slam vs Non-Grand Slam games (being 1 if the game was a GS game and 0 if not)
2. First round vs Non-First round games (being 1 if the game was a 1st round game and 0 if not)
3. At least one player in the top 50 vs No player in the top 50 (being 1 if a top50 player was in the game and 0 if not)
4. ATP games vs WTA games (being 1 if the game was an ATP game and 0 if not)

The null hypothesis (H_0) has always been that both population means are equal, while the alternative hypothesis (H_1) have sustained that a significant difference exists between the average over-rounds of the two populations.

See the following example:

1. H_0: $\mu_{GS} - \mu_{NGS} = 0$ against H_1: $\mu_{GS} - \mu_{NGS} \neq 0$

In this respect, in order to know if the variances of the populations can be assumed to be equal or not, Levene tests have been performed (being H_0: the variances of both populations are equal). So, as it happens in the cases that we have studied, if the probability associated with the Levene statistic (p value) is lower than 0.01, it can be assumed with a 1% significance level that the population variances are not the same[2]. Hence, for the t-test statistics, it has been assumed that the population variances are not equal. However, the results are almost identical in both cases.

So, at a 5% significance level, if the two-sided p value or sig. is lower than 0.025, it can be said that strong evidence is obtained suggesting that the null hypothesis does not hold, and therefore, population means are not equal ($\mu_x - \mu_y \neq 0$). The same logic applies if 0 is not included in the limits of the confident interval, or if the calculated t values are lower or higher than the critical t values at a 5% significance level and *x* degrees of freedom.

6.2 Regression analysis

The following lines describe the methodology used to analyse the over-round variations in the betting firm Bet365. As our aim is to compare the over-round evolution and see if, ceteris paribus, the over-round is higher for female professional tennis games than for male ones, basic linear regression models have been used. First, an OLS have been constructed for ATP games, then, the same procedure has been replicated for WTA matches, and finally, a general model has been developed, where matches from both categories have been included. In this last model, in order to see if more probable information asymmetries create a higher over-round in women's games, a dummy variable, $DATP_i$, has been included (1 if the game is an ATP game, 0 if not). If our main hypothesis holds (and therefore the null hypothesis is rejected), $DATP_i$ has to present a negative coefficient, and of course, it needs to be significant.

Furthermore, in line with previous findings and standard procedure in the literature (see Hidalgo et al., 2016; Lahvicka, 2014; Abinzano et al., 2016), the study variable (the over-round) has been included as the dependent variable, while a mix of dummy and continuous variables have been included as independent variables, in order to control for other forces (apart from a higher probability of facing information asymmetries in the WTA circuit) that could affect the over-round. The regression models will therefore be estimated including the following variables:

Regarding continuous variables, two have been included: The points difference between the players in absolute terms (Points dif.), and the "World Tour" ranking difference, also in

---

[2] The Levene tests results can be viewed in Table 3 and in Appendixes 2 and 3.

absolute terms (Rank dif.). Both are proxies that reflect the difference in the skills or level of the players, but as long as no multicollinearity problems have emerged, it has been decided to include both in the model. Consider that in the presence of high multicollinearity, confidence intervals tend to become very wide, and in consequence, null hypothesis are more difficult to reject. These problems have not turned out in our analysis.

On the other hand, regarding dummy variables, 12 have been included:

- DGS$_i$, which takes the value of one when the game is a Grand Slam game, and 0 if not. The logic says that the variable should have a negative coefficient, as long as more information is available on GS tournaments, the players tend to free ride less, and all the top players participate in the event.
- DT50$_i$, which takes the value of one when a at least one of the game participants is ranked in the best 50 tennis players, and 0 if not. Again, common sense predicts that the coefficient of the variable should be negative, as more information is public for top players, and bookmakers have less doubts about their performance.
- D1R$_i$, which takes the value of one if the match is a first-round game and 0 else ways. First round matches should exhibit higher over-rounds (as explained in section [5]), so the direction of the variable should be positive. Put another way, the bookmakers protect themselves in a greater extent in the first round, as uncertainty is higher, and less information is available.
- $\sum_{i=2011}^{2018} D_i$, which take the value of one if the game was played in the respective year (0 if not), and control for progressive decreases in the over-round due to technology improvements and an increase in competition (see Hidalgo et al., 2016). In order to prevent perfect multicollinearity, we have omitted the year 2010, which has been taken as the reference category. Moreover, as long as online platforms are constantly improving and expanding, the variables should exhibit negative coefficients, that should be even more negative as years go by (the over-round in 2011 was smaller than in 2010, the over-round in 2012 was smaller than 2011 etc.).

So, the models are:

[2] $$\underline{\text{ATP Over-round}_i} = \alpha_1 + \alpha_2 \text{DGS}_i + \alpha_3 \text{D1R}_i + \alpha_4 \text{DT50}_i + \beta_5 \text{Points dif.}_i + \beta_6 \text{Rank dif.}_i + \sum_{i=2011}^{2018} D_i \alpha_i + \varepsilon_i$$

[3] $$\underline{\text{WTA Over-round}_i} = \alpha_1 + \alpha_2 \text{DGS}_i + \alpha_3 \text{D1R}_i + \alpha_4 \text{DT50}_i + \beta_5 \text{Points dif.}_i + \beta_6 \text{Rank dif.}_i + \sum_{i=2011}^{2018} D_i \alpha_i + \varepsilon_i$$

$$[4]\quad \underline{\text{General Over-round}_i} = \alpha_1 + \alpha_2 DGS_i + \alpha_3 D1R_i + \alpha_4 DT50_i + \boldsymbol{\alpha_5 ATP_i} + \beta_6 \text{Points dif.}_i + \beta_7 \text{Rank dif.}_i + \sum_{i=2011}^{2018} D_i\, \alpha_i + \varepsilon_i$$

7. RESULTS

7.1 Univariate analysis

In order to determine if significance differences exist between the sample means, a univariate analysis of the data has been conducted, as described in the previous section. In this part of the study we have not deal with causes or relationship, and the sole intention of the analysis has been to determine if the disparities between the sample means are significant enough to assume that a difference also exist in the population means.

The results that are presented in Table 3 include data from both, women's and men's tennis matches, but the tests have also been replicated analysing both samples independently (women's and men's matches), with identical results (see Appendix 2 and 3). As it can be seen in Table 3, the null hypothesis (the over-round mean of the populations is equal) is rejected in each of the cases, leading us to the following conclusions.

Table 3. Levene tests for equality of variances and T-tests for equality of over-round means.

		Levene test for equality of variances	*T-test for equality of over-round means*	
		F	**T- test**	**Mean dif.**
Male/Female	$\sigma^2_M = \sigma^2_F$	55.463 *	-13.550 *	- 0.001566296
	$\sigma^2_M \neq \sigma^2_F$		-13.563 *	- 0.001566296
GS/ NGS	$\sigma^2_{GS} = \sigma^2_{NGS}$	541.459 *	-59.215 *	-0.008269882
	$\sigma^2_{GS} \neq \sigma^2_{NGS}$		-53.749 *	-0.008269882
1R / N1R	$\sigma^2_{1R} = \sigma^2_{N1R}$	336.037 *	42.704 *	0.0048633818
	$\sigma^2_{1R} \neq \sigma^2_{N1R}$		42.887 *	0.0048633818
T50 / NT50	$\sigma^2_{T50} = \sigma^2_{NT50}$	593.141 *	-54.846 *	-0.006985752
	$\sigma^2_{T50} \neq \sigma^2_{NT50}$		-58.702 *	-0.006985752

* represents that the tests are significant at a 5% significance level

Regarding the mean over-round for Grand Slam and non-Grand Slam tournaments, the null hypothesis can be rejected at a 1% significance level, meaning that there are strong evidence suggesting that an over-round difference exists between the two populations. The mean difference is negative, meaning that GS have lower over-rounds than the rest of the tournaments. Therefore, if the game takes place on a GS or not is something to be considered in our model.

The same logic applies when we analyse the over-round for matches that had at least one player in the top 50. The mean difference is again negative and the null hypothesis can be rejected at a 1% significance level, suggesting that the over-round is higher when a top 50 is not in the game. The reasoning behind these differences is the one that have been explained in section [5], and again, if a top50 player is or not in the game, has to be considered at the time of constructing the econometric model.

On the other hand, the mean difference is positive for matches that have been played on the first round. What is more, a significant over-round difference exists among the first-round games and the rest, rejecting the null hypothesis at a 1% significance level. The t-test result makes perfect sense, as long as bookmakers receive important information in the first round of the tournaments, as it has been mentioned earlier.

Finally, the null hypothesis that the over-round is equal in women's and men's tennis matches is also rejected at a 1% significance level. Put another way, there are strong evidence suggesting that an over-round difference exist between women's and men's tennis games. The mean difference is negative, so bookmarkers impose lower over-rounds in games played by male tennis players. This means that there is room for analysis in this field, and in consequence, our main hypothesis may be correct. However, the fact that a significance over-round difference exists in female and male tennis matches does not prove anything per se. In consequence, in the model that is presented in the next lines, an attempt is made to control for other variables that may have an influence in the over-round, analysing if controlling for these variables, the disparities persist. However, for the future, in order to give more strength to the hypothesis that more frequent information asymmetries are the trigger of these differences, it may be interesting to find an information asymmetry proxy, and see if indeed, the logic applies.

7.2 Regression analysis

After conducting the estimation using the OLS method, three linear regression models have been obtained, which have confirmed the intuitions discussed in the paper. In the following lines, a deep overview of the results is carried out.

All three models exhibit similar characteristics (see the results on Table 4). As expected, the over-round is lower when the game takes place in a GS, one of the players is ranked among the best 50 tennis players of the world, the match is not a first-round game, and the ranking and points differences are higher. Furthermore, the control variables that have been introduced to limit the effects that technology improvements and the increase in competition have had in the over-round, also show negative coefficients, result that enforces the idea that the over-round has followed a decreasing pattern in the last decade, driven by the evolution of the above-mentioned factors.

Table 4. Results of the models using the ols method.

VARIABLE	ATP MODEL	WTA MODEL	GENERAL MODEL
Constant	0.078 *	0.078 *	0.079 *
DATP			- **0.01** *
DGS	- 0.006 *	- 0.007 *	- 0.007 *
D1R	0.002 *	0.002 *	0.002 *
DT50	- 0.002 *	- 0.002 *	- 0.002 *
Rank dif.	- 0.00001863 *	- 0.00001636 *	- 0.00001719 *
Points dif.	- 0.000002296 *	- 0.000002524 *	- 0.000002453 *
D2011	- 0.002 *	- 0.001 *	- 0.001 *
D2012	- 0.005 *	- 0.005 *	- 0.005 *
D2013	- 0.005 *	- 0.004 *	- 0.005 *
D2014	- 0.005 *	- 0.005 *	- 0.005 *
D2015	- 0.005 *	- 0.004 *	- 0.005 *
D2016	- 0.005 *	- 0.005 *	- 0.005 *
D2017	- 0.007 *	- 0.007 *	- 0.007 *
D2018	- 0.007 *	- 0.007 *	- 0.007 *
ADJ R2	0.395	0.260	0.333

*represents that the variable is significant at a 5% significance level

The only remarkable difference between the models that try to explain the over-round variation in WTA and ATP games is that, although the constant term has the same value, the over-round decreases in a greater extent in the WTA circuit when the match is played on a GS tournament and the points difference is higher. However, on the contrary, the ranking difference has a more negative coefficient in the ATP circuit. A possible explanation for this apparently contradicting result may be the following: In the male circuit, although two players may be very close in terms of ATP points, a significant difference exists between the level that both players have, meaning that the player with the highest ranking is usually the one that obtains the victory. This generates less uncertainty for the bookmaker at the time of quoting the odds, especially in face to faces matches with minor rank differences. On the other hand, in the WTA World Tour, the level of the players is more even, and in consequence, the raking of each player has less importance. Put another way, in women's tennis, if both players have similar world tour points, the result is more uncertain than in men's tennis, and in consequence, the points difference is more important than the ranking itself. Being more precise, data shows that the player with the best ranking wins the game the 66.37% of the times in the ATP circuit, while in the WTA circuit, the percentage decreases to 64.71%. Because of that, for forthcoming studies, it may be interesting to include interactions terms to the model, as adding interactions to a regression can greatly increase the understanding among the variables in the model, allowing to test hypothesis with greater robustness. Put another way, the possibility to analyse if higher adverse selection costs have a more or less pronounce effect on the over-round when the game takes place on a GS or not, when the match is a first-round game or not, or when a top 50 player is in the game or not, among others, would be real. That is, the presence of a significant interaction would indicate that the effect of the explanatory variables in the over-round is different when the dummy ATPi takes the value of 1. The new regression equation would be the following:

$$[5] \quad \underline{\text{Int. General Over-round}_i} = \alpha_1 + \alpha_2 \text{DGS}_i + \alpha_3 \text{D1R}_i + \alpha_4 \text{DT50}_i + \boldsymbol{\alpha_5} \mathbf{ATP_i} + \beta_6 \text{Points dif.}_i + \beta_7 \text{Rank dif.}_i + \sum_{i=2011}^{2018} D_i \alpha_i + \alpha_{14}(\text{DGSi}\mathbf{ATP_i}) + \alpha_{15}(\text{D1Ri}\mathbf{ATP_i}) + \alpha_{16}(\text{DT50i}\mathbf{ATP_i}) + \beta_7(\text{Points dif.}_i\mathbf{ATP_i}) + \beta_8(\text{Rank dif.}_i\mathbf{ATP_i}) + \varepsilon_i$$

Going back to the original model, regarding the dummy variable DGS_i, the following may occur: As we have been predicting since the beginning of the paper, public information on female tennis players is scarcer and more difficult to find. Because of that, when female players participate in a Grand Slam, and media coverage increases exponentially, the extra information that bookmakers obtain in comparison to other less renown tournaments is higher in women's tennis than in men's tennis. In consequence, bookmakers are able to impound more information into prices, decreasing the over-round and making markets more efficient. Regarding the R^2 coefficients, it has to be mentioned that the same model is able to explain the 39.5% of the over-round variability in the ATP circuit, while it only explains the 26% of the over-round variability in the WTA. This supports our ideia that differences exist in the over-round formation between ATP and WTA games.

Furthermore, in the estimation of model three, where data from both women's and men's tennis matches is used, the coefficient of the dummy variable, $DATP_i$, is negative. This means that with all the rest equal, male tennis matches exhibit lower over-rounds in betting platforms (a more in deep analysis of the DATP variable will be provided in section [8]).

In order to give an illustration of how the estimated models should be interpreted, see the interpretation of model [4] (models [2] and [3] can be read in the same way, with the exception of the dummy variable, DATP, that does not exist in these models):

If the match did not take place on a Grand Slam, if it was not a first-round match, if it was played by two female players that were below the 50th position in the ranking, if it took place in 2010, and no points and ranking difference existed between the players, the model predicts that the over-round of that match had to be 0.079. However, as no match with these characteristics took place (two players cannot have the same ranking position), the value of the model resides in the interpretation of its parameters. Take into account that when two players have the same amount of world tour points, the tie is break considering the following factors in order of importance; the quantity of points that have been obtained in GS or Master 1000 tournaments, and the number of tournaments needed to obtain that quantity of points.

In this way, the model forecasts that with all the rest equal, the over-round decreases 0.007 points when the game is played on a Grand Slam. Furthermore, as long as the p value is 0.000 (see Appendix 6) we can reject the null hypothesis that $\alpha_2 = 0$ at 1% significance level. So, there is evidence that the over-round is lower in GS games, and as a result, DGS_i should be added to the model. Put another way, DGS_i is statistically significantly different from zero. The reasoning behind the direction of this variable is the same explained in previous sections.

Regarding $D1R_i$, the direction of the variable is positive. In other words, ceteris paribus, the over-round is 0.002 higher when the match takes place on the first round. Moreover, at a 1% significance level, we reject the null hypothesis that α_3 is equal to 0 (p value is 0.000) and in consequence, as long as it has a significant influence at the time of explaining the variability of the over-round, $D1R_i$ is also included in the model,.

$DT50_i$ is also significant with a 99% confidence level, and it predicts that, other things being equal, the over-round is 0.002 lower when a at least one of the best 50 tennis players of the world is in the game. Apart from the reasons already mentioned, the players that are ranked among the best 50 tennis players of the word have higher incentives to maintain their raking, and in consequence, less reasons to free ride. It is convenient to remember that in terms of earnings from advertising agreements, tournament prizes, or international prestige, an important difference exists between the best players and the rest.

In relation to the points and ranking differences, the model specifies that the over-round decreases in 0.00001719 points when the ranking difference increases in one unit, while the negative effect is 0.000002453 for each point difference between the players. Both variables are significant at a 99% confidence level, and the logic behind the direction of the two variables is quite intuitive: The higher the ranking and point differences are, the more noticeable will be the level or skill differences between the players, imposing less uncertainty and a lower probability of transacting with better informed counterparties to the bookmaker when quoting the odds. The transcendence of the variables may seem insignificant at first sight, as the coefficient of the variables are small, but nothing could be further from the truth. As an example, the point difference between Stephanos Tsitsipas (tenth position in the ranking) and Lucas Pouille (thirtieth) was of 1985 points (3160-1265) at the 18th of march, 2019. That means that ceteris paribus, the points difference decreases the over-round in 1985*0.000002453 = 0.004869205, for the reasons explained above. A considerable variation. So, it can be concluded that Points dif.$_i$ and Rank dif.$_i$ are explicative and significant, and in consequence, they should be included in the model. The problems that may arise from the non-linearity of ATP points will be examined in more advance sections of the paper.

Finally, regarding the dummy variables $\sum_{i=2011}^{2018} D_i$, the null hypothesis ($\alpha_i = 0$) is rejected in every case, at a 1% significance level. The reason to include these control variables is to exclude alternative explanations while testing the hypothesis that higher information asymmetries motivate a higher over-round in the WTA circuit. Put another way, the inclusion and interpretation of the control variables is therefore theoretically motivated, rather than statistically.

7.3 Discussion

In order to avoid any hasty conclusion, it may be convenient to analyse if the model is correctly specified in advance. For that, in the next lines, an overview of the model is done, detecting possible problems that the model could have.

First, the models do not present any inconsistency with economic theory or common sense. All the variables exhibit logical directions, and their interpretation does not contradict previous studies in this field (see Abinzano et al., 2017; Flepp et al., 2015; Gomber et al., 2008; Elizalde, 2015). In other words, the models are data admissible and all the predictions are logically possible.

Furthermore, the coefficients of the variables show constancy when they are individually added to the model, and no one is excluded from it when the computer performs a stepwise regression or "data mining". In other words, if we estimate 13 different models, adding an extra variable for each one of them, the coefficients do not experiment any significant changes or direction variations. The overall significance of the model has been also studied, imposing the null hypothesis that $\sum_{i=1}^{15} \alpha_i$ is equal to 0. So, by assessing multiple coefficients simultaneously, it can be affirmed that the fit of the intercept-only model is lower than the general model (F value 1627.207 and p value 0.000).

The R^2 coefficient of the general model is 0.333. Put another way, the model explains the 33.3% of the over-round variability in Bet365 bookmaker's market. In this sense, concerns regarding the goodness of the fit may be logical at first instance, but as our intention is focused on determining if a significant over-round difference exist between WTA and ATP games, the R^2 coefficient is not a crucial indicator. Previous studies that were focused on studying the significance of a dummy variable also showed low R^2 coefficients (see Bladh and Sandberg, 2013; Casado et al., 2011). Take into account that if the model explains the majority of the over-round's variability, the possibility of making accurate predictions in betting markets would be real, and in consequence, easy profits could be obtained. Thus, a low R square coefficient is consistent with the efficient market hypothesis. In this extent, the same problem arouses with the Durbin-Watson coefficient, as the general model exhibits a Durbin-Watson coefficient of 1,6.

So, in order to see if the introduction of a new variable changes the main conclusions of this work, a new linear regression model has been estimated, including the implicit probability of the favourite ($1/p_{favourite}$) as a new variable. The implicit probability of the favourite has been already used in several betting market studies, including Abinzano et al. (2017) and Deutscher et al. (2017). Table 5 shows the results of the estimation.

Table 5. Results of the new model (including Fav. prob.) using the ols method.

	Const.	DGS	DT50	D1R	Rank dif.	Point dif.	DATP	Fav. prob.	$\sum_{i=2011}^{2018} D_i$
β	0.113	-0.005	0.000	0.003	-0.000002106	-0.0000009464	-0.01	-0.052	
t	339.432	-46.708	-34.266	-2.363	-4.926	-35.421	-8.889	-29.561	Yes
Sig.	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.000	
DW=1.444		**R_2=0.486**		**Adjusted R^2=0.485**					

With the introduction of the new variable, Fav. $prob._i$, which measures the level of uncertainty of the game (less uncertainty for higher implicit probability $1/p_{\text{favourite}}$), the general model suffers some considerable changes. First, the estimation shows that a 1% increase in the implicit probability of the favourite decreases the over-round by 0.052*(1/100)=0,00052.

Second, the dummy variable $DT5O_i$ becomes insignificant at a 1% significance level. Third, other variables see their coefficients slightly altered: DGS_i, Rank dif.$_{i}$, and Point dif.$_{i}$ exhibit less negative coefficients due to the introduction of the new regressor, while the variables $D1R_i$, $D2011_i$, $D2016_i$, $D2017_i$ and $D2018_i$ suffer the reverse effect.

However, while the majority of the model's variables suffer some sort of change, the significance and the coefficient of the dummy variable, $DATP_i$, remain invariant. Put another way, the introduction of the new variable creates alterations in an otherwise stable model, but the direction, significance, and interpretation of the study variable does not change. This enforces the view that, although the possibility of omitting a relevant variable in the general model cannot be completely discarded, the hypothesis holds even when new regressors are considered. In short, the new model still presents evidence suggesting that due to a higher probability of facing information asymmetries, bookmakers impose a higher over-round in the WTA world tour than in the ATP circuit, replicating what happens in financial markets with the bid-ask spread.

Nevertheless, although the inclusion of the variable Fav. Prob.$_{i}$ could be interesting to provide a higher robustness to the results, it also presents serious problems, as endogeneity and multicollinearity problems are created in the model. This happens due to the fact that the over-round and the implicit probability of the favorite ($1/p$) are codetermined, with each affecting the other. Hence, by estimating either equation by itself, endogeneity is created, and as the rest of the explanatory variables linearly predict Fav.Prob.$_{i}$ with a substantial degree of accuracy, multicollinearity problems are also revealed. Remember that the over-round is calculated by summing the implicit probabilities of the favorite and the long shot and subtracting 1 (see Equation [1]). In that way, as long as the downside of the new variable (serious simultaneity problems are created) is more evident than the upside, is has been decided not to include Fav.Prob.$_{i}$ in the general model.

On the other hand, the variable Points Dif.$_{i}$ may provoke misspecifications problems, as the point distribution in the ATP and WTA world tours is not lineal. Figure 5 illustrates the non-linearity of the points distributions in tennis rankings, reproducing the ATP raking at the 18^{h} of march, 2019. As it can be seen in the graph, the number of points that each player has increases dramatically with the first three players of the ranking (Novak Djokovic, Rafael Nadal and Alexander Zverev respectively), and in consequence, the variable Points dif. loses explanatory power. The exact same thing happens in the WTA circuit. So, with the intention of increasing the model's accuracy, we have transformed the variable Points dif.$_{i}$ into ln(Points dif.$_{i}$). In that way, the highest values of the dataset are compressed, while the lowest ones are expanded, becoming the distribution symmetric. The results of the new model are displayed on Table 6.

Figure 5. ATP ranking at 18/03/2019.

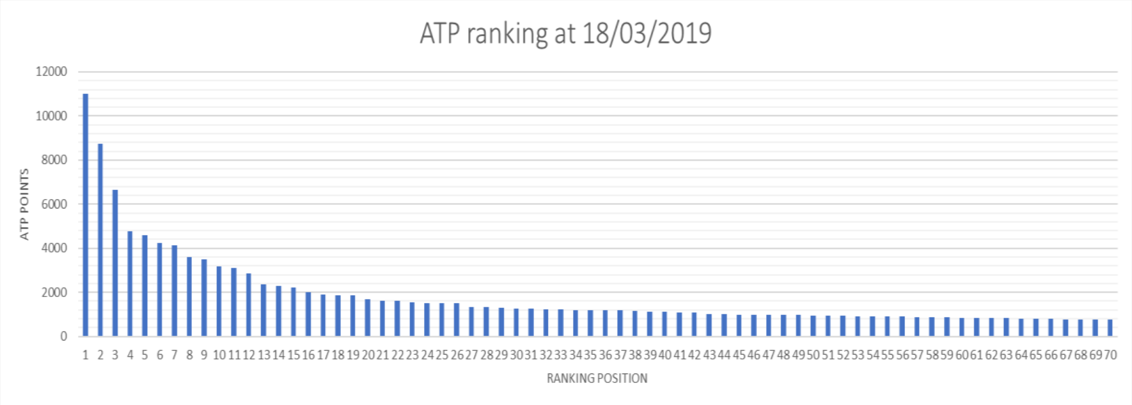


Table 6. Results of the model (modifying the specification of Points dif.) using the ols model.

	Const.	DGS	DT50	D1R	Rank dif.	Ln(Point dif.)	DATP	$\sum_{i=2011}^{2018} D_i$
β	0.095	-0.007	0.003	0.000	-0.00001064	-0.003	-0.002	Yes
t	300.699	-56.309	23.819	1.678	-21.046	-67.929	-8.684	
Sig.	0.000	0.000	0.093	0.000	0.000	0.000	0.000	
DW=1.642		**R_2=.0282**		**Adjusted**	**R^2=0.281**			

In the transformed model, the interpretation of the variable ln(points $dif._i$), which is significant at a 1% significance level, is the following: Ceteris paribus, a 1% change in the points difference between the players, is associated with a decrease in the over-round of 0.003*0,01 = 0.00003. Put another way, the estimated effect of Points $dif._i$ is no longer linear, even though the effect of log(Points dif.i) is linear.

Again, the only mentionable difference between the original and the new model is that now, the variable $DT50_i$ is not significance at a 5% significance level. Furthermore, the variable $DATP_i$ continues being significance with a 99% confidence level, so the specification change does not have any effect on the accuracy of our initial prediction. What is more, after the adjustment, the estimated coefficient of $DATP_i$ is even more negative, being the estimated over-round 0.002 lower when the match is an ATP game (with all the rest equal). The rest of the variables remain almost invariant, with minor changes. However, as the R^2 coefficient is even lower in this new model, it can be mentioned that the specification change has not improved the goodness of the fit.

8. DTAI CONCLUDING REMARKS

After the analysis conducted, there are enough evidence to suggest that an over-round difference exists between ATP and WTA matches, and consequently, the null hypothesis that ATP and WTA games exhibit the same over-rounds can be rejected. The variable $DATP_i$ is significant in each one of the general models, and even the models [3] and [4] (constructed with ATP and WTA data respectively), present evidence suggesting that an over-round difference exists between ATP and WTA games. As it has been mentioned in previous lines of the work, our main hypothesis is that this dissimilarity is provoked by the higher possibility that bookmakers have to encounter better informed counterparties in the women circuit. Being more precise, the models estimated in this paper predict that, ceteris paribus, the over-round is between 0.001 and 0.002 lower in ATP games than in WTA matches, a percentage difference of about 2% of the over-round.

As information for women's games is less extensive, driven by less media coverage and lower stadium attendances among others, private economically meaningful information is easier to find, and in consequence, bookmakers must defend themselves from huge loses. Put another way, knowledge about women players is lower among market participants, and in consequence, as less information is impounded into prices (the markets are less efficient), there is more room for the existence of unknown information that escapes the awareness of bookmakers. Hidalgo et al. (2016) obtained evidence that supports a parallel conclusion for football betting markets, as they found that bookmakers impose higher over-rounds in minor and foreign soccer's leagues.

These findings are consistent with financial theory, as theoretical models of the bid-ask spread hold that a proportion of the bid-ask is based on asymmetric information. Among others, Glosten and Harris (1988) found that by discomposing the bid-ask spread into four components: Information asymmetries, inventory costs, clearing cost, and monopoly power, the hypothesis of information asymmetries as the cause of bid-ask spreads cannot be rejected. For the future, it may be interesting to test if bookmaker's over-rounds can be employed to test for an increase in information asymmetries, prior to an anticipated information event, as press appearances of the players or open doors trainings. Similar studies in financial markets have analysed bid-ask spread variations prior to dividend or earning announcements (see Venkatesh and Chiang, 1986). Remember that bookmakers obtain a benefit from trading against worse informed investors and from the bid-ask spread, so, if a greater proportion of better-informed investors are active in the market (with advance information on the fundamental value of the asset) the bid-ask spread will be higher, worsening price discovery.

So, by obtaining evidence that suggest that due to information asymmetries, and with all the rest equal, the over-round is higher in professional women's tennis games, this article approaches the issue of the similarity between financial and betting markets, and it gives evidence implying that major similarities exist at the time of quoting the odds, and in consequence, of deciding the spreads among financial and betting markets.

9. CONCLUSIONS

The study presents significance evidence supporting that an over-round difference exists between professional women's and men's tennis matches. In other words, the findings of this paper indicate that financial and betting markets react against adverse selection costs in an analogous way, increasing the market spread. However, far from market microstructure contributions, the conclusions of this analysis also have a social implication.

The disparities in the monetary prizes awarded for participating and winning ATP or WTA tournaments are notable, except in the case of Grand Slams, where the economic incentives are exactly the same. In this sense, tournaments' organizers defend themselves arguing that these differences have their justification in the higher monetary returns that men players generate (in the form of greater audiences, greater stadium attendances, or greater television coverage). However, one may argue that people's interest in both circuits is very similar, and although there may be audience differences from time to time, these are cyclical and changing. Put another way, they depend on the charisma and the tennis level of the players of each era, and in consequence, both, women and men tennis players, should have the same prizes.

So, after analysing which are the main factors that drive the over-round differences between WTA and ATP games, we can also predict, with an acceptable degree of accuracy, which are some of the causes behind the wage gap. In this extent, other scholars as Smith et al. (2006) or Hetherington (2006) have affirmed that betting markets create an adequate environment for hypothesis testing, even outside the financial world. As an example, Vaughan et al. (2016), presented evidence suggesting that betting markets have a more accurate capacity to predict electoral processes than polls.

So, based on our data, and leaving moral implications aside, it can be said that the prize differences between men and women tennis players could have a theoretical backing. If men players create more revenues in terms of merchandising, stadium attendances, or television audiences, more public information about them would be publicly available and this would be reflected in lower over-rounds at betting markets. In some sense, this chain effect would have its roots in the higher interest that ATP games theoretically awake.

In some sense, our study is also built on this assumption, and the empirical research has given evidence in favour of this hypothesis, confirming that ceteris paribus, over-rounds are lower in the ATP circuit, and pointing adverse selection costs as the most plausible trigger of the difference. So, controlling for other variables, it seems that the ATP circuit generates a wider media coverage and higher stadium attendances than the WTA circuit, factors that drive the difference in both, the over-round and the wage of the players.

However, there is something of a chicken and egg situation here, since it cannot be affirmed if a wider media coverage of ATP games has its root in a higher public interest, or the other way around. That is to say, as the accessibility to male tennis has been greater for years, in terms of media coverage or advertising for example, fans may have learned to prefer ATP games over WTA matches. Further studies in this field will be necessary in the future to understand if a cause and effect relationship exists.

All in all, although at first sight it seems very distant, a deeper understanding of the over-round could be helpful to figure it out the reasons behind the salary gap, and to stablish corrective or preventive measures that could assist in the long road to gender equality in sports, and more concretely, in tennis. So, as a final remark, it may be concluded that if more resources are devoted to track the professional female circuit, more information on WTA players will be public, increasing market efficiency, and decreasing the over-rounds. At the same time, as bettors will bear lower transaction costs, betting volumes will also experience an upward trend, increasing the revenues generated by the WTA circuit, and thereby, reducing the wage gap in professional tennis. As an example, in Grand Slam tournaments, where media coverage is considerably higher than in other tournaments for female players, prizes for both genders are identical.

REFERENCES

Abinzano, I., Muga, L., and Santamaria, R. (2017). Behavioral biases never walk alone: An empirical analysis of the effect of overconfidence on probabilities. Journal of Sports Economics, 18, 99–125.

Abinzano, I., Muga, L., and Santamaria, R. (2017). Hidden Power of Trading Activity: The FLB in Tennis Betting Exchanges. Journal Of Sports Economics, 20(2), 261-285. doi: 10.1177/1527002517731875

Baker, R., and McHale, I. (2013). Optimal Betting Under Parameter Uncertainty: Improving the Kelly Criterion. Decision Analysis, 10(3), 189-199. doi: 10.1287/deca.2013.0271

Barberis, N., and Thaler, R. (2003). A survey of behavioral finance- Handbook of the Economics of Finance. 1(18), p. 1053-1128.

Battalio, R., Greene, J., and Jennings, R. (1997). Do Competing Specialists and Preferencing Dealers Affect Market Quality?. Review Of Financial Studies, 10(4), 969-993. doi: 10.1093/rfs/10.4.969

Bladh, J., and Sandberg, C. (2013). The US Holiday Effect - Evidence from Nordic markets on the impact of US investors. Stockholm School of Economics.

Brindley, C. (1999). The marketing of gambling on the Internet. Internet Research, 9(4), 281-286. doi: 10.1108/10662249910286798

Brogaard, J., Garriott, C., and Pomeranets, A. (2014). High-Frequency Trading Competition. Bank Of Canada Working Paper 2014-19. doi: 10.2139/ssrn.2435999

Brown, J., and Mino D. (2014). Selecting the Best? Spillover and Shadows in Elimination Tournaments. Management Science. 60 (12), p. 3087–3102.

Cain, M., Law, D., and Peel, D. (2000). The Favourite-Longshot Bias and Market Efficiency in UK Football betting. Scottish Journal Of Political Economy, 47(1), 25-36. doi: 10.1111/1467-9485.00151

Casado, J., Muga, L., and Santamaria, R. (2013). The effect of US holidays on the European markets: When the cat's away…. Accounting and Finance, 53, 111– 136.

Chiang, R., and Venkatesh, P. (1986). Information Asymmetry and the Dealer's Bid-Ask Spread: A Case Study of Earnings and Dividend Announcements. Journal of Finance. 41. 1089-1102. 10.1111/j.1540-6261.1986.tb02532.x.

Coleman, L. (2007). Just how serious is insider trading? An evaluation using thoroughbred wagering markets. The Journal of Gambling Business and Economics (2007) 1, 31–55

Croxson, K., and James Reade, J. (2013). Information and Efficiency: Goal Arrival in Soccer Betting. The Economic Journal, 124(575), 62-91. doi: 10.1111/ecoj.12033

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997). Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. The Journal Of Finance, 52(3), 1036-1058.

Davies, M., Pitt, L., Shapiro, D., and Watson, R. (2005). Betfair.com: Five technology forces revolutionize worldwide wagering. European Management Journal, 23,533–541.

De Jong, F., and Rindi, B. (2009). The microstructure of financial markets. Cambridge: Cambridge University Press.

Demsetz, H. (1968). The costs of transacting. Quarterly Journal of Economics, 82, 33-53. http://dx.doi.org/10.2307/1882244

Deschamps, B and Gergaud, Olivier. (2007). Efficiency in Betting Markets: Evidence From English Football. The Journal of Prediction Markets. 1. 61-73.

Deutscher, C., Frick, B., and Ötting, M. (2018). Betting market inefficiencies are short-lived in German professional football. Applied Economics, 50(30), 3240-3246. doi: 10.1080/00036846.2017.1418082

Elizalde, A. (2015). Microestructura y formación de precios en mercados financieros y de apuestas deportivas (Trabajo fin de grado). Universidad Pública de Navarra.

Expansión. (2018). Las apuestas deportivas se disparan hasta 742 millones. Retrieved from http://www.expansion.com/directivos/deporte-negocio/2018/11/01/5bda1ac9e5fdea133c8b45a4.html

Fama, Eugene.(1970). Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance, 25, 2, p. 383-417.

Flepp, R., Nüesch, S., and Franck, E. (2017). The liquidity advantage of the quote-driven market: Evidence from the betting industry. The Quarterly Review Of Economics And Finance, 64, 306-317. doi: 10.1016/j.qref.2016.07.016

Forrest, D., and Mchale, I. (2007). Anyone for Tennis (Betting)?. The European Journal Of Finance, 13(8), 751-768. doi: 10.1080/13518470701705736

Forrest, D., and Simmons, R. (2008). Sentiment in the betting market on Spanish football. Applied Economics, 40, 119–126.

Franck, E., Verbeek, E., and Nüesch, S. (2010). Prediction accuracy of different market structures — bookmakers versus a betting exchange. International Journal Of Forecasting, 26(3), 448-459. doi: 10.1016/j.ijforecast.2010.01.004

Glosten, L., and Harris, L. (1988). Estimating the components of the bid/ask spread. Journal Of Financial Economics, 21(1), 123-142. doi: 10.1016/0304-405x(88)90034-7

Gomber, P., Rohr, P., and Schweickert, U. (2008). Sports betting as a new asset class—current market organization and options for development. Financial Markets And Portfolio Management, 22(2), 169-192. doi: 10.1007/s11408-008-0077-7

Grinblatt, M., and Titman, S. (1994). A Study of Monthly Mutual Fund Returns and Performance Evaluation Techniques. The Journal Of Financial And Quantitative Analysis, 29(3), 419. doi: 10.2307/2331338

Harris, L. (2003). Trading and exchanges: Market microstructure for practitioners. Oxford University Press, Inc

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

Hidalgo, T., Del Corral, J., and Gómez-González, C. (2016). Variabilidad en el mercado de apuestas deportivas. Universidad de Castilla La-Mancha.

Huang, Roger D.(2002). The Quality of ECN and Nasdaq Market Maker Quotes. Journal of Finance, 57, issue 3, p. 1285-1319.

Hvattum, L.M. (2013). Analyzing information efficiency in the betting market for association football league winners. The Journal of Prediction Markets, 7:55-70.

Institutional investor. (2017). Playing the Odds in Sports Parallels Financial Markets. (2017). Retrieved from https://www.institutionalinvestor.com/article/b1505pg5zw6mq9/playing-the-odds-in-sports-parallels-financial-markets

Investopedia. (2018).Quote-Driven Market. (2018). Retrieved from https://www.investopedia.com/terms/q/quotedriven.asp

Jiang, W. (2003). A nonparametric test of market timing. Journal of Empirical Finance 10,399–425

Koning, R., and Van Velzen, B. (2009). Betting Exchanges: The Future of Sports Betting?. International Journal of Sport Finance. 4. 42-62.

Lahvička. (2014). What causes the favourite-longshot bias? Further evidence from tennis. Applied Economics Letters. 21. 10.1080/13504851.2013.842628.

Law, D., and Peel, D. (2002). Insider Trading, Herding Behaviour and Market Plungers in the British Horse-race Betting Market. Economica, 69(274), 327-338. doi: 10.1111/1468-0335.00285

Levin, E., and Wright, R. (2004). Estimating the profit markup component of the bid-ask spread: evidence from the London Stock Exchange. The Quarterly Review Of Economics And Finance, 44(1), 1-19. doi: 10.1016/s1062-9769(03)00005-x

Levitt, S. (2004). Why are gambling markets organised so differently from financial markets? Economic Journal, 114, 223–246.

Lin, J., G. Sanger, and G. Booth. (1995). Trade size and components of the bid-ask spread. Review of Financial Studies, vol. 8, nr.4, p.1153-1183

Lyócsa Š., and Výrost T. (2017): To bet or not to bet: a reality check for tennis betting market efficiency. Applied Economics, DOI: 10.1080/00036846.2017.1394973

Mayhew, S. (2002). Competition, Market Structure, and Bid-Ask Spreads in Stock Option Markets. The Journal Of Finance, 57(2), 931-958. doi: 10.1111/1540-6261.00447

Mchale, I.G. and Morton, Alex. (2011). A Bradley-Terry type model for forecasting tennis match results. International Journal of Forecasting. 27. 619-630. 10.1016/j.ijforecast.2010.04.004.

Moskowitz, T. (2015). Asset Pricing and Sports Betting. SSRN Electronic Journal. doi: 10.2139/ssrn.2635517

Sauer, R. (1998). The economics of wagering markets. Journal of Economic Literature, 36, 2021–2064.

Scheibehenne, B., and Bröder, A. (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition. International Journal Of Forecasting, 23(3), 415-426. doi:10.1016/j.ijforecast.2007.05.006

Schwartz, D. (2013). Roll The Bones: The History of Gambling. Winchester Books. ISBN 978-0-615-84778-8.

Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. The Journal Of Finance, 19(3), 425. doi: 10.2307/2977928

Shin, H. (1991). Optimal Betting Odds Against Insider Traders. The Economic Journal, 101(408), 1179. doi: 10.2307/2234434

Smith, M., Paton, D., and Williams, L. (2006). Market Efficiency in Person-to-Person Betting. Economica, 73(292), 673-689. doi: 10.1111/j.1468-0335.2006.00518.x

Smith, M., Paton, D., and Williams, L. (2009). Do bookmakers possess superior skills to bettors in predicting outcomes?. Journal Of Economic Behavior and Organization, 71(2), 539-549. doi: 10.1016/j.jebo.2009.03.016

Snowberg, E., and Wolfers, J. (2010). Explaining the Favorite–Long Shot Bias: Is it Risk-Love or Misperceptions?. Journal Of Political Economy, 118(4), 723-746. doi: 10.1086/655844

Sobel, R., and Travis Raines, S. (2003). An examination of the empirical derivatives of the favourite-longshot bias in racetrack betting. Applied Economics, 35(4), 371-385. doi: 10.1080/00036840110111176

Williams, L. (1999). Information Efficiency in Betting Markets: a Survey. Bulletin Of Economic Research, 51(1), 1-39. doi: 10.1111/1467-8586.00069

Williams, L., and Paton, D. (1997). Why is There a Favourite-Longshot Bias in British Racetrack Betting Markets. The Economic Journal, 107(440), 150-158. doi: 10.1111/1468-0297.00147

Williams, L., and Reade, J. (2015). Forecasting Elections. Journal Of Forecasting, 35(4), 308-328. doi: 10.1002/for.2377

APPENDIX

Appendix 1. Betfair's Betting Exchange, Wimbledon 2016 data.

		Number of bets	Volume	Volume per game	Average bet
Male	*QF*	118,176	47,059,922.72 €	11,764,980.68 €	398.22 €
	SF	86,658	36,033,617.93 €	18,016,808.97 €	415.81 €
	F	30,862	14,929,707.55 €	14,929,707.55 €	483.76 €
	Total	235,696	98,023,248.20 €	14,003,321.17 €	415.89 €
		Number of bets	**Volume**	**Volume per game**	**Average bet**
Female	*QF*	64,108	26,354,398.93 €	6,588,599.73 €	411.09 €
	SF	23,623	13,742,385.29 €	6,871,192.65 €	581.74 €
	F	27,598	12,944,270.72 €	12,944,270.72 €	469.03 €
	Total	115,329	53,041,054.94 €	7,577,293.56 €	459.91 €

Appendix 2. Levene tests for equality of variances and T-tests for equality of over-round means (ATP data).

		Levene test for equality of variances	*T-test for equality of over-round means*	
		F	**T- test**	**Mean dif.**
GS/ NGS	$\sigma^2_{\mathrm{GS}} = \sigma^2_{\mathrm{NGS}}$	368,603*	-46,155*	-0.009136136
	$\sigma^2_{\mathrm{GS}} \neq \sigma^2_{\mathrm{NGS}}$		-40,584*	-0.009136136
1R / N1R	$\sigma^2_{\mathrm{1R}} = \sigma^2_{\mathrm{N1R}}$	230,501*	34,442*	0,0055086273
	$\sigma^2_{\mathrm{1R}} \neq \sigma^2_{\mathrm{N1R}}$		34,789*	0,0055086273
T50 / NT50	$\sigma^2_{\mathrm{T50}} = \sigma^2_{\mathrm{NT50}}$	533,788*	-43,099*	-0.007505703
	$\sigma^2_{\mathrm{T50}} \neq \sigma^2_{\mathrm{NT50}}$		-48,412*	-0.007505703

Appendix 3. Levene tests for equality of variances and T-tests for equality of over-round means (WTA data).

		Levene test for equality of variances	*T-test for equality of over-round means*	
		F	**T- test**	**Mean dif.**
GS/ NGS	$\sigma^2_{\mathrm{GS}} = \sigma^2_{\mathrm{NGS}}$	167,055*	-38,103*	-0,007463030
	$\sigma^2_{\mathrm{GS}} \neq \sigma^2_{\mathrm{NGS}}$		-35,976*	-0,007463030
1R / N1R	$\sigma^2_{\mathrm{1R}} = \sigma^2_{\mathrm{N1R}}$	90,245*	25,442*	0,0041091435
	$\sigma^2_{\mathrm{1R}} \neq \sigma^2_{\mathrm{N1R}}$		25,440*	0,0041091435
T50 / NT50	$\sigma^2_{\mathrm{T50}} = \sigma^2_{\mathrm{NT50}}$	114,735*	-34,061*	-0.005889865
	$\sigma^2_{\mathrm{T50}} \neq \sigma^2_{\mathrm{NT50}}$		-34,878*	-0.005889865

Appendix 4. Results of the model using WTA data (ols method):

	β	t	Sig.	95% confidence interval (inferior)	95% confidence interval (superior)
Constant	.078	323.782	.000	.078	.079
Points dif.	-2.524E-6	-80.216	.000	.000	.000
DGS	-.007	-44.283	.000	-.008	-.007
Rank dif.	-1.636E-5	-27.818	.000	.000	.000
D1R	.002	15.212	.000	.002	.002
DT50	-.002	-14.506	.000	-.003	-.002
D2011	-.001	-5.081	.000	-.002	-.001
D2012	-.005	-19.884	.000	-.006	-.005
D2013	-.005	-19.537	.000	-.006	-.005
D2014	-.005	-17.554	.000	-.005	-.004
D2015	-.005	-17.658	.000	-.005	-.004
D2016	-.005	-19.716	.000	-.006	-.005
D2017	-.007	-27.710	.000	-.008	-.007
D2018	-.007	-27.746	.000	-.008	-.007

Appendix 5. Results of the model using ATP data (ols method):

	β	t	Sig.	95% confidence interval (inferior)	95% confidence interval (superior)
Constant	.078	296.699	.000	.078	.079
Points dif.	-2.296E-6	-47.956	.000	.000	.000
DGS	-.006	-33.914	.000	-.006	-.006
Rank dif.	-1.863E-5	-25.088	.000	.000	.000
D1R	.002	15.165	.000	.002	.003
DT50	-.002	-9.109	.000	-.002	-.001
D2011	-.002	-5.416	.000	-.002	-.001
D2012	-.005	-16.359	.000	-.006	-.004
D2013	-.004	-13.426	.000	-.005	-.003
D2014	-.005	-15.009	.000	-.005	-.004
D2015	-.004	-14.429	.000	-.005	-.004
D2016	-.005	-15.578	.000	-.005	-.004
D2017	-.007	-21.753	.000	-.007	-.006
D2018	-.007	-22.335	.000	-.007	-.006

Appendix 6. Results of the model using ATP+WTA data (ols method):

	β	t	Sig.	95% confidence interval (inferior)	95% confidence interval (superior)
Constant	.079	427,015	.000	.079	.079
Points dif.	-2.453E-6	-92.143	.000	.000	.000
DGS	-.007	-54.924	.000	-.007	-.006
Rank dif.	-1.719E-5	-37.035	.000	.000	.000
D1R	.002	21.630	.000	.002	.002
DATP	-.001	-14.378	.000	-.002	-.001
DT50	-.002	-16.055	.000	-.002	-.002
D2011	-.001	-7.409	.000	-.002	-.001
D2012	-.005	-25.521	.000	-.006	-.005
D2013	-.005	-23.124	.000	-.005	-.004
D2014	-.005	-22.975	.000	-.005	-.004
D2015	-.005	-22.738	.000	-.005	-.004
D2016	-.005	-25.022	.000	-.005	-.005
D2017	-.007	-34.951	.000	-.007	-.007
D2018	-.007	-35.278	.000	-.007	-.007