UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M.FANNO"

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

CORSO DI LAUREA MAGISTRALE / SPECIALISTICA IN ECONOMICS AND FINANCE

TESI DI LAUREA

"SPORTS BETTING: A NEW ASSET CLASS TO BET ON"

RELATORE:

CH.MO PROF. CLAUDIO FONTANA

LAUREANDO/A: DAVIDE TURCHETTO

MATRICOLA N. 1234400

ANNO ACCADEMICO 2021 – 2022

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

The candidate declares that the present work is original and has not already been submitted, totally or in part, for the purposes of attaining an academic degree in other Italian or foreign universities. The candidate also declares that all the materials used during the preparation of the thesis have been explicitly indicated in the text and in the section "Bibliographical references" and that any textual citations can be identified through an explicit reference to the original publication.

Firma dello studente

....
(firma)

TABLE OF CONTENTS

Introduction…………………………………………………………………5

1 The betting evolution……………………………………………………7

1.1 Bookmakers business model……………………………………………**8**

1.2 Exchange betting business model………………………………………**12**

1.2.1 Exchange vs Bookmakers………………………………………**15**

1.3 Sports betting vs financial markets………………………………………**16**

2 Sport trading……………………………………………………………18

2.1 Forecasting price movements in betting exchanges……………………**19**

2.2 Investing strategies………………………………………………………**21**

2.3 The role of psychology…………………………………………………**23**

2.4 Sport and stock market…………………………………………………**26**

3 Is the bookmaker market efficient?……………………………………31

3.1 The efficiency of the fixed-odds betting market………………………**34**

3.1.1 Price setting mechanism……………………………………………**35**

3.1.2 Biases in the football betting market………………………………**39**

3.1.3 Efficiency tests………………………………………………………**43**

3.2 The efficiency in the exchange betting market…………………………**52**

4 Arbitrage opportunities……………………………………………..………………**55**

4.1 matched betting……………………………………………………………………..**56**

4.2 sure betting………………………………………………………………………….**58**

4.3 value betting………………………………………………………………...….**61**

4.3.1 Value found through the exchange market…………………………...….**62**

4.3.2 Value found through bookmakers' disagreement……………………………….**64**

4.3.3 Value found through analysis of historical data…………...…………………….**68**

4.3.4Value found through machine learning approach…………...………………......**75**

4.3.2 Money management……………………..…………………………………………**83**

5 Personal value betting experiment……………………….…………………………**87**

5.1 The strategy…………………………………………………………………………**87**

5.2 Results……………………………………………………………………………….**91**

5.3 Final considerations……………………………………………………………...**96**

Conclusion……………………………….…………..………………………….………**98**

Bibliography and Websites……………………..………………………….………**101**

Introduction

This dissertation has the aim to present a complete overview of the current features and activities related to the sports betting industry and to explain the reasons why it can be considered a new asset class to invest on.

Sports betting can be seen as a branch of the wide gambling industry. At the beginning, it was not legal, thus not regulated. The rules applied today are relatively young. For instance, in Italy the fixed-odds bet on a football match is legal only since 1998. Nowadays, it is allowed and regulated in almost every continent, as governments have found in it a good source of revenues through licences and taxes, giving back to costumers protection against fraud, manipulation, nondisclosure of substantial information, and insider trading. On the other hand, it can be the cause of psychological diseases. This is the main reason why it is not legal everywhere. For example, in the United States it is permitted only in some states, but, with the rising of online betting, it is becoming more difficult to prohibit it, and several states are starting to concede to betting companies to operate.

Globally, the amount wagered every year is estimated in 1,500 billion of Euro. The numbers of this market are growing rapidly since the last two decades. This because the internet revolution has deeply changed this business, developing the offer available for the clients. First of all, it gave birth to the online betting market, lowering the effort needed to access to this service and fixed costs, and opening new market opportunities for existing and new bookmakers. The other big innovation is the appearance of the exchange market, a new business model based on the peer-to-peer betting system, where the players bet against each other. Its functioning is similar to a stock exchange market. Betfair is the principal platform operating now. It processes around seven million trades a day, more than the number of daily trades on all the European Stock Exchanges combined.

In this dissertation, the first chapter wants to explain the mechanism which runs the bookmakers and the exchange business model, their differences and points in common. The last subchapter compares the betting exchange market with the stock exchange market. Their similarities are the basement for the second chapter, that analyse the trading activity applied to the sports betting world. The rising of the exchange platforms has given another point of view for this market: now odds can be seen as assets to trade with the other market participants. The target of this chapter is to explain the fundamental aspects and characteristics to consider in order to successfully operate in this particular market. A possible football match outcome is not a regular stock, and the characters that belong to these platforms are, for the majority, not investors. Thus, strategies, profit opportunities and possible dangers must be highlighted. Moreover, there is a direct connection between the financial market and the sports betting world that it is worth to consider.

Despite the long history of betting as a business, scientific literature about this topic has been developed only recently. This is partially attributed to the fact that on-line betting and betting exchange were established only in this century. By now, the betting market has attracted the attention of several researchers, since its new structure has given rise to new investment opportunities and new aspects of research. The betting market has become a sort of a testing ground for various ideas, practices, and technologies of profit generation. Moreover, the exchange market has been frequently used as a proxy of the financial one, since their characteristics are similar. With this purpose, the sports betting market, intended as the union of fixed-odds and exchange platforms, has been deeply analysed concerning its statistical and economical efficiency. Due to its simpler features, it fits better for tests related to this topic. Therefore, there are lots of studies that examine in depth whether the market is efficient or not, the origins of the inefficiencies and the biases they reflect in the market.

The thesis proceeds exposing the principal arbitrage betting strategies, based on the biases and imperfections detected by the efficiency tests. The matched-betting exploit bonuses and promotions, while the sure bet strategy takes advantage of the bookmaker's discordance. The value betting strategy aims to find mispriced odds and to wager on them when the probability of an event to occur is higher than the one implied in the odds, hence when there is a positive expected profit. The scientific literature has focused its efforts on how to correctly estimate the real probability of a sport event. This work presents the tools used for making this strategy profitable, from the simplest, that consists of just looking at the exchange probabilities, to the most complex, the implementation of machine learning and artificial intelligence techniques, highlighting their degree of reliability and potentialities, without hiding flaws and feasibility.

The last chapter exposes a personal research project that tries to put in practice the value betting strategy, where the value is found through the bookmaker's disagreement. This experiment gives a contribution to the market efficiency and statistical arbitrage research, showing that, even if retroactively, a profitable betting strategy is attainable. Its main point of innovation is that it covers a time range until the end of 2021, thus it is possible to evaluate the effects of the pandemic on the strategy's profitability.

The last section concludes the thesis, giving some incipits and ideas that could definitely legitimize sports betting as a new asset class.

1 The betting evolution

The origin of gambling dates back to the Paleolithic age[1]. During history, human beings have wagered money on every kind of event or game, with the desire to have fun and to earn some money. Betting is a type of gambling activity that consists in challenging someone, putting money "on the table", that a specific event will occur in the future. In other words, it is a contract between two parts that establish future cash flows associated with different outcomes of an event. Their size is proportional to their quantified likelihood, their probability, that is reflected by the price, the odd in this case. This business has been moving huge amounts of money for ages, but only in the last decades it is becoming recognized and regulated by public governments. Its economic and social impact is not negligible. It represents a source of revenue for governments (licenses and taxation) and employment for people, but, on the other hand, it generates regulatory costs and social issues, such as game addiction and the appeal for criminal organizations.

The biggest field on which people love to bet is sport. The first recognized bookmaker was opened in 1790 by Harry Ogden in the United Kingdom. Nowadays, it is possible to bet on every sport and in every aspect of a match or a race. Since its legalization in Italy in 1998, this business has grown rapidly. In its "Report 2021", the Italian football federation ("FIGC") has drawn the situation in Italy, up to the end of 2020. The gambling industry as a whole has moved 110.5 billion of Euro. Sports betting has generated a business turnover of 13.1 billion of Euro, where, not surprisingly, football alone covers almost 10 billion, then tennis and basketball are the most followed sports. In Italy, Bet365.com is the most used website, but it operates only online. The other popular companies, such as Snai, Goldbet, Lottomatica, Eurobet and Sisal, combine physical shops with an online offer.

The market revenues are measured by the Gross Gaming Revenue (GGR), also called "game yield", which is a key metric used by gambling and betting companies. The Corporate Finance Institute defines it as the difference between the amount of money players wager and the amount that they win, where the first is intended as the amount of money collected from gambling and/or betting transactions, and the second refers to the amount of money that has been paid out to customers for winning. It is constantly growing, and now it is 1.58 billion of Euro.

The taxation rules are different within each country. In Italy, the tax rate is established as 20% and 24% of the GGR, for physical and online bets respectively. The 2020 revenue for the state was 253 million of Euro, excluding money from licences. The economic newspaper "Il sole 24

[1] www.britannica.com/topic/gambling/History

ore" has registered a revenue reduction for the treasury from the entire gambling sector from more than 11 billion to 7 billion of Euro. This is mainly due to the Covid pandemic. The FIGC has noticed a drop on the volume betted in physical shops of 45.3%, while the online betting has grown by 12.6%. This data demonstrate that the pandemic has accelerated the innovation process on this market.

As in many other activities, the recent technological revolution has deeply transformed the settings of this world. Firstly, bookmakers have moved their shops online and they have changed their business model. Secondly, internet has given the opportunity to build new types of markets and new ways of operating. As in the landing field, the peer-to-peer (P2P) disintermediation process has been developed, giving birth to a new way of placing bets: the betting exchange market. It consists in wagering against another player, and not against a bookie (see chapter 1.2). Its structure presents several similar traits with the stock market, and with that also investment opportunities. For this reason, participants are now not only gamblers, but also traders, speculators and investors. Therefore, sports betting has become an increasingly frequent subject of financial research.

To understand regular bookmakers and exchange platforms functioning and differences, it is necessary to analyse them singularly.

1.1 Bookmakers business model

The two main actors in this market are: the betting company and the player. The first one provides a service that gives the possibility to its clients to wager, the latter "signs" the contract by accepting the price set unilaterally by the counterpart. Hence, the price is nothing else than the odd. This concept is of crucial importance in the functioning of this business.

To begin, odds are expressed in different ways across the world. First and foremost, they are displayed in decimal, for instance "1.58" or "2.5" and they establish the total return the customer will receive if he places a bet at that price. Thus, with a €10 bet, the cash out would be €15,80 and €25 respectively, if the player wins. In the UK, which has the most liquid market in the world, prices are shown in fractions, for example "4/1" or "7/4", where the numerator expresses the amount of money you would earn if you bet the amount expressed by the denominator. The last way is the Moneyline or American one. In this case, the quote can be expressed by an either positive or negative figure[2] (+150 or -120). If the figure is positive, e.g.

[2] If the implied probability is higher than 50% the figure is positive, otherwise is negative.

+150, it shows the amount of money you would win on a 100$ bet. That is, if we bet 100$ we would gain a profit of 150$ in case of win, otherwise we would lose our 100$. Whereas, if we have a negative odd, e.g. -120, it expresses the amount of money you need to wager in order to win 100$, so the pay-out would be 100$ if we win or we would lose 120$ otherwise. Comparing them to the decimal odds, the American quote will be positive if the decimal odd is higher than 2 and it will represent the net profit if we bet 100$, or negative if decimal odds is lower than 2, indicating, in this case, the amount to wage in order to have the possibility to earn 100$. In this dissertation, we refer to odds in decimal, since it is the simplest and most used representation in Europe. To clarify this, we will make an example considering the principal market of a football match: the "1, X, 2 bet", where 1 stands for home win, X for draw and 2 for away victory. If we assume that the decimal quotes offered in the market are "1,95" for home win (1), "3,3" for a draw (X) odd and "3,6" for away victory (2), and we bet €5 on away win (2), our cash flow would be €0 if the match ends both 1 or X (thus we would lose our €5), alternatively we would cash in €18.00 (5x3,6) if the away team wins. This amount includes the one we gambled ("stake") and our net profit, which in this case would be €13.

The most important thing in a business model is how to create value, in other words how to make profits. Usually, in the other gambling products, such as roulettes or slot-machines, probabilities of winning are known with certainty, but they are not equally reflected by the odds. The organizers gain by the law of large numbers, since odds are stacked in their favour (typically they keep an edge around 14%). For instance, in a common Russian Roulette, gambling on colours (red or black) gives a pay-out of twice your bet if you win, but, since a colour covers less than half of the numbers, because the number zero (that is green) makes you lose, the pay-out should be higher. In fact, if we bet €10 on red, the table pays as it was risking in 50% cases (the odds is 2), but, in reality, the probability to lose is only 0.4865% (there are 18 red numbers,18 black numbers and zero; the probability is 18/37). Consequently, the organizers' concern is to attract as many players as possible. Similarly, bookies must attract players too, but probabilities are "home-made", subjective, uncertain, and therefore they must be careful: if bettors are able to recognize and exploit mispricing, bookies can sustain large losses. The article by J. Smith (2001) reports some examples. In 1946 over half of all British bookmakers were bankrupted in the first running after World War II ended, because of a horse wrongly given at odds of 50 to 1. Moreover, Coral Eurobet reported losses of £12 million on internet betting in the quarter-final round of Euro 2000 soccer championship. The firm offered the most competitive odds on the favourite teams and all four of them won easily.

Given these examples, it is clear that for bookmakers making a good business means to be able to offer good odds, both profitable and attractive. But how to be profitable? In order to answer this, we must be conscious that the probability of a specific outcome is given by the reciprocal of the quote. For example, if we flip a coin, the probability of obtaining head or cross is 50% in both cases, thus the fair odds should be 2 (since 0.5 = 1/odds). Hence, odds are inversely related to probabilities associated to that event, i.e. they are as higher as less probable an outcome is.

Bookmakers want to be rewarded for the service they provide and for the counterparty risk the support. They achieve this goal by imposing a margin to their quotes, that is to short the odds. In this way the quote doesn't reflect their estimated true probability anymore, but an "artificially" increased one. Mathematically, this is what happens:

$$\theta_i = \frac{1}{\delta_i}; \qquad M = \sum_{j=1}^{n} \theta_i - 1$$

Where θ_i is the probability associated to the odd δ_i, while M is the margin. Adding all the probabilities together they will not sum to 1, but more. Keeping flipping the coin as example, if we would bet on this game a common bookie would offer us odds around 1.85. The margin would then be 1/1.85 + 1/1.85 = 1.081 - 1 = 8.1%. If we assume that bets are equally distributed on both sides, the table would earn 8.1% of the total amount bet on this game. To come back to the "true" probability ρ_i, we can simply use this formula:

$$\rho_i = \frac{\theta_i}{1 + M}$$

Running their business, bookies face two types of risk: the first is to present misleading odds, so to wrongly compute the real probabilities; the second is to be too exposed on one side of the market, facing the risk to bear some losses. This latter case happens when bets are not equally distributed across all possible outcomes, then the final profit for the bookmaker depends on the final result. If the book is particularly unbalanced, the pay-out could be higher than the amount stacked on the other outcomes, resulting in losses for the company. For this reason, the price is not the market clearing price and the zero-sum condition does not hold in this market. Hence, betting on each outcome would lead the player to lose money.

The price setting mechanism is as crucial as peculiar. The bookmaker odds are influenced by both the true outcome probabilities and the bettors' demand. Gaming companies are not gamblers, they don't want their profit to be related to the match or race result, their aim is to minimise the risk. To do that, they have to compute consistent "true" probabilities, and forecast bettors' preferences. If they are extremely good at determining in advance the price which equals the quantity of money wagered on each side of the bet, they will make money regardless of who will win the game. Even though this is the most efficient strategy, there is anecdotal evidence that they adopt riskier strategies depending on the selected level of risk tolerance and the anticipated profitability of alternative strategies. Franck, Verbeek, and Nuesch (2010) have demonstrated, for example, that bookmakers offer more (less) favourable terms for bets on teams with a comparatively large (small) fan base in order to attract a disproportionately large betting volume. At the end, supposing now that the bets are not equally distributed across all possible outcomes, the expected margin, that is the profit for the bookie, will be:

$$E(M) = 1 - \sum_{i=1}^{n} P_i * w_i * \delta_i$$

Where P_i is the real probability associated with each outcome, w_i is the percentage of bets on each outcome and δ_i is the quoted odd. The margin is not the same in every match. It reflects the riskiness of the match, that is related to the quantity of information available for computing the probabilities and the number of possible outcomes. Big matches with famous teams usually do not present huge margins, because there is a lot of liquidity invested, the market is more competitive and lots of information and statistics are available to prepare the quotes. For example, bet365[3] takes a margin usually around 5/6% in a Champions League match, while in secondary leagues, like Africans or Asian, it is around 11/12%[4]. The same phenomenon happens when the possible outcomes are not just two or three, but way more, as in horse racing for example. In this case the risk of being too exposed on a horse is high, and bookmakers protect themselves by increasing a lot the margin. The same reasoning can be applied depending on which teams are playing. Mangold and Stübinger (2020) highlight that the margin percentage changes according to the difficulty of forecasting the result of a specific team. Figure 1[5] displays the relation between average points and average risk premium for the teams of the Spanish Primera Division from 2006 to 2018. Top teams like Real Madrid and Barcelona achieve around 2.3 points per match, bad teams like Granada and Gijon around 1 point per match. They observe an average risk premium of approximately 6.2 percent for the top teams as well as the bad teams. Teams of average quality in terms of long-run performance like Zaragoza and Mallorca

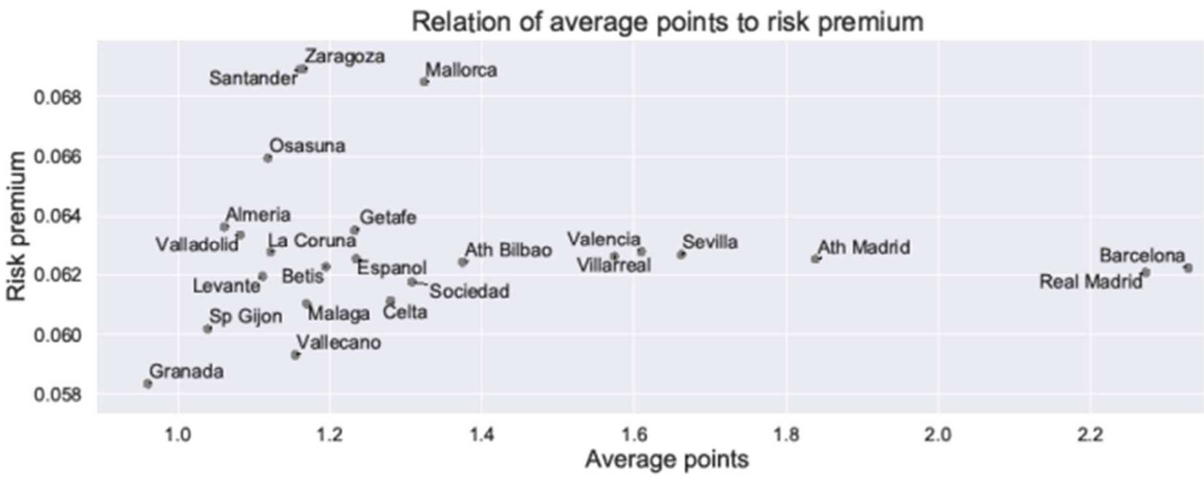
[3] The most used betting website in Europe (10,9% of the entire market) according to *"Online gaming"* by Deloitte.

[4] Data are used just as an example. Other bookmakers can have different values, but the principle still apply.

[5] All the figures presented in this dissertation are taken from the specific paper under consideration.

possess a higher risk premium. Thus, it is possible to deduct that it is much more difficult to predict the outcome of matches involving teams of average quality.

Figure 1 Relation of average points and average risk premium for the teams of the Primera Division from 2006 to 2018.



1.2 Exchange betting business model

As preannounced, the disintermediation revolution has also involved the gambling world. The first peer-to-peer betting platform was announced in 2000 in the UK. The website name was www.flutter.com. In this market customers could publish their bets, and the counterpart could choose to bet against one or more among them. Betfair.com, the biggest platform still operating, was launched only a few months later[6], but its way of operating was more efficient, hence Flutter decided to merge with Betfair in December 2001. It is structured as a common stock exchange, where people can purchase or sell assets, that in this case are bets on possible outcomes, at agreed prices, represented by the quotes. In other words, they contract their opposing opinions with each other. After the bets have been matched, both individuals hold a contract on a future cash flow. The size of the cash flow is determined by the price of the contract, while the direction of the cash flow is tied to the outcome of the underlying event.

The name Betfair is emblematic: the purpose of this innovative market is to overcome bookmaker biases on odds and to offer fair ones. In this platform odds must reflect the true probability of an event to happen. This aim is reached through letting players offer and buy odds, without including any margin on them. It is an order-driven market. Companies profit by imposing a commission only on winning bets, since they do not bear any kind of risk. There is no interest in the outcome, the only concern is to match as many bets as possible. The percentage charged is set around 5% of the net winnings, depending on the country regulation. Although, according to how much a client wagers on the site, it is possible to reduce the base rate until

[6] In Italy the launch dates back to April 2014.

2%. Moving huge quantities of money, as sport traders do, makes this market even more convenient. As a result, online betting exchanges have experienced a fast expansion.

To explain in detail how it works, it is useful to show the way it looks like:

Figure 2: Betfair 1X2 market

Juventus	**3.75** €1762	**3.8** €1136	**3.85** €414	**3.9** €32	**3.95** €3752	**4** €2105
Chelsea	**2.24** €5399	**2.26** €2188	**2.28** €1248	**2.3** €476	**2.32** €3731	**2.34** €1908
Pareggio	**3.1** €3671	**3.15** €5707	**3.2** €1753	**3.25** €789	**3.3** €1185	**3.35** €3485

Keeping as an example a 1 (home win), X (draw), 2 (away win) market of a football[7] match, this is how the double auction is presented. Numbers in bold are the odds, while the amount below stays for the related liquidity in Euro available at that moment. In other words, it shows the maximum amount of money you can instantaneously match on that quote at that moment. If it is not enough, your money will stay in the market until someone will accept that quote. If the offer has not found a counterpart yet at the beginning of the match, it will be deleted as soon as the match starts, unless the punter orders to keep it. The table must be seen in lines, where each one shows the odds available for the outcome related on the left. Basically, for each outcome it is possible to bet on (i.e. back) or against it (i.e. lay). If we back a team, we are stating that we are expecting that this team will win. The blue box shows the best available odds on which it is possible to back a team. For example, choosing the blue box on the first line, we are betting on a Juventus win against Chelsea at the odds of 3.85. The coloured box is the best option currently available, and the odds on the left are the second and third options respectively. The same reasoning, but in the right way, applies if you lay a team. When we lay an outcome, we are saying that that result is not going to occur. Thus, by laying Juventus means that the player is betting on a Chelsea win or a draw. At the current situation, the best quote for betting against the Juventus win is the one in the pink box: 3.9.

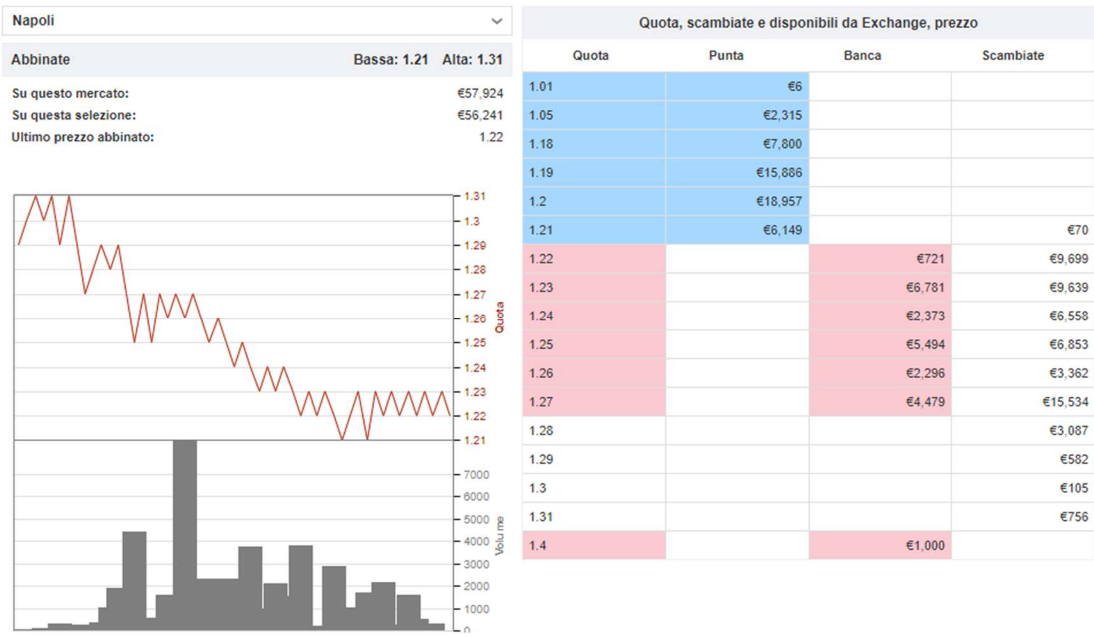
The back bet works as a usual fixed-odds one. We put the amount we want to wager, and the system gives us our potential profit. Thus, betting €10 on Juventus on the best odd available, the net profit would be (3.85*€10) - 5% = €36,57 if our team wins, otherwise we would lose 10€. The lay bet in the exchange market works in a mirror way: we must digit how much money we wish to win, and the website shows us the amount we should risk to earn that money, that in jargon is called "liability". The latter is calculated in this way: gross outcome * (odds-1). For

[7] Football is the most popular sport, thus the most liquid.

example, supposing to always have €10 at our disposal, at the odds of 3.9 we should lay €3.45 on Juventus (10/(3.9-1)). Consequently, if Juventus wins, we would lose €10, if Chelsea wins or ties the net pay-out would be: €3.45 - 5% = €3,28. Laying a team is more costly as the probability is lower and the odd is higher, because we are betting against an outcome that is unlikely to occur.

The way the book is run is very similar to a stock exchange market. Orders are sorted following a time precedence principle and everyone can propose his buy or sell orders at the price they want. If there is a counterpart available, the system will immediately close the contract, otherwise the offer will be put on queue. Players can start trading a few days before the event. Betfair provides useful statistics and charts to help players understanding how the market is moving. This is strongly important for sports traders. In Figure 3, we have a complete overview of this team's odd characteristics. Betfair gives at its clients disposal the complete history of the odds movements for each match. On the left are displayed the highest and the lowest odds agreed, the total amount matched until now on the entire market and on this specific outcome (Napoli's win in this example), and a chart showing quote's movements and volume traded time by time. On the right side, instead, it is summarised the current situation. The first column shows the odds, where again the blue ones are of back type and the pink are lay type, the second and the third column show the liquidity available on the back and lay side respectively, and the fourth column shows the amount already matched for each specific odds.

Figure 3 Betfair market statistics



Quota	Punta	Banca	Scambiate
1.01	€6		
1.05	€2,315		
1.18	€7,800		
1.19	€15,886		
1.2	€18,957		
1.21	€6,149		€70
1.22		€721	€9,699
1.23		€6,781	€9,639
1.24		€2,373	€6,558
1.25		€5,494	€6,853
1.26		€2,296	€3,362
1.27		€4,479	€15,534
1.28			€3,087
1.29			€582
1.3			€105
1.31			€756
1.4		€1,000	

The most revolutionary innovation launched by Betfair is the Cash Out. It is a tool that permits to players, after they close a bet, to sell their position in the market. In other words, bettors can quit the market before the end of the match, or even before it begins. Moment by moment, the website shows us our position (the money we would earn or lose by cashing-out at that moment), depending on how odds have moved and how the match is going. Cash Out is available only on top leagues because the market must be liquid, otherwise it's impossible to sell ongoing bets. The site displays in every moment the updated payoffs if the player would cash out at that moment, and he can choose to do it fully or partially. At the same time, the player can choose how to exit the market, that means with the same pay-out for every outcome ("all green") or weighting each position as he prefers.

1.2.1 Exchange vs Bookmakers

From the previous paragraphs, it is easy to see the huge difference between the two types of businesses, although they long to serve and attract the same market. The bookies' fixed odds system is simpler and easy to learn, there are dozens of companies, also with physical shops, thus it suits better for occasional players. Moreover, it is possible to create multiple bets, including different matches in the same contract. The pay-out will be the amount invested multiplied by every quote. It is not possible to lay on an event, but just bet on the logical opposite outcome. It is less transparent, because it is possible to see only the quotes offered in that moment, while Betfair makes available the whole limit order book as well as the matched price history.

The exchange platform is a more complicated system that involves three characters: two players and the platform's manager, which put together supply and demand. Here, above all in secondary matches, odds can change frequently, even if only for a decimal ("tick" in jargon). When liquidity is low, it is not difficult to absorb it, causing relevant fluctuation in the market. Hence, the market is more volatile, which means profitable and dangerous at the same time. For this reason, it suits better for professional traders than for sporadic gamblers. Liquidity is essential for exchanges, both for the owners and players, while the latter don't have to worry about it in a regular fixed odds bet, because bookies provide it. On the other hand, if in this case the only constraint is liquidity, regular bookmakers can choose to change the quote right before closing the contract, or to refuse your order because you are wagering too much money, and this would expose them to a risky situation. More than that, they can unilaterally suspend your account. In this way, they marginalise players that win with some regularity. In the exchange platform, as soon as you find a counterpart, you can keep matching bets. Furthermore, winners are more than welcome since it charges commissions only to pay-outs. Consequently, odds are higher because they don’t include any implicit margin. In both cases, the maximum amount you can win with a single bet is €10.000 and live betting is available.

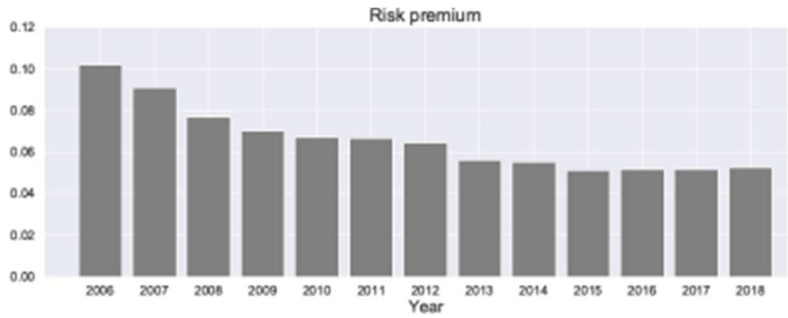
To conclude, Cash Out is available only on exchanges.

1.3 Sports betting vs financial markets

Sports betting and financial markets share several features, but also important differences. Within both markets, participants (traders, investors, speculators) hold different expectations, estimates, beliefs and information about the future, and seek to realise profit through exchanges as the uncertainty about the future is resolved gradually over time. They both have a winner and a loser. Large amounts of money and trades occur every day, and all of them incorporate a risk, since you will get a reward only if your choice turns out to be correct or lose money otherwise. Fixed-odds sports wagering markets have, contrary to financial markets, detailed price and pay-off information bound to the specific outcome of a sporting event. Furthermore, since the event will take place with certainty, there is an exactly specified time horizon between the bet and the realisation of any profit or loss. Consequently, the maturity of their products is different: while a financial asset could not have a specified outcome and maturity, a bet has a specific cashflow related to each outcome, and the maturity is always short, since it is just related to the duration of a match or a race. A parallel that works is to compare bets with binary options, since by buying this option, you are wagering that in a specific moment in the future the value of the underlying will be equal or higher than a predetermined stake price. If you are right, you will get the agreed amount of money. They work in the exact same way.

Both in the financial and in the sports betting world, there is a rapid dissemination of information that is of crucial importance. Bookmakers generate odds processing information about the likelihood of sport events in the same way that financial markets aggregate information about uncertain future asset payoffs. Another point in common is the possibility and profitability of insider information parallels closely that found in the stock market. Moreover, while there are different theories regarding the efficiency of the financial market, there is no doubt that the betting market is getting more efficient thanks to the increasing use of big data and quantitative analysis. Therefore, more and more betting providers are entering the market and the information available is increasing. Figure 4 proves this fact: bookmakers are reducing their margin due to the increasing number of companies operating in this market. In this sense, competitiveness pushes the market through a higher level of efficiency.

Figure 4 Average risk premium of the bookmaker Bet365 from 2006 to 2018.



But, above all in the gambling sector, the majority of players are not rational; they are, instead, victims of psychological biases: they end up merely betting on their favorite team, for example, or they play in an impulsive way, often with the desire to recover from huge losses.

The entrance on the gambling market of order driven platforms has moved closer the two worlds. Nevertheless, while the financial market can rely on big regulated markets that guarantee protection for investors, gaming companies operate in a decentralised and heterogeneous market, divided into numerous entities and brands. The main markets can be categorised as follows: bookmakers, exchanges and betting brokers. The fact that these entities are decentralised can be a source of inefficiency in the market because the same event can have different prices across different bookies, opening the possibility of profitable opportunities.

2 Sport trading

Since the introduction of betting exchanges, there has been an increasing interest about how to monetise on this new technology. The exchange characteristics are tempting for speculators since odds are more volatile than fixed-odds markets and the players are often irrational and affected by psychological biases. Consequently, a new professional figure was born: the sport trader, a full-time high-volume trader who buys and sells odds just like financial traders buy and sell stocks. The aim is trying to profit from the quote's swings. Basically, as compared to regular gamblers, traders follow a specific strategy and a strict plan, that allows them to be rational, to avoid sentimental mistakes and to support the stress. Another fundamental aspect is to understand how the market usually behaves, which are the moments on which it is more dynamic and how it reacts to new information. For example, after the publication of the line-ups or a few minutes before the beginning of the match, the market is particularly animated. Once he gets more confident with these mechanisms, the trader can implement diverse strategies, according to his willingness to risk and time he can invest. In other words, the trader must play against the crowd, and try to bring its moves forward. When a bet is matched, it is impossible to know who the counterpart is and what are its reasons to accept that odds. Thus, you must be better informed than him for gaining a profit. Their activity is mostly based on in-play bets, that are bets placed while the match is ongoing. They are used to enter and quit the market through specific software during the match, using the cash out. They plan their trades using statistical analysis and they try to exploit "surprise" biases on the market.

Nowadays, there are just a few professionals in sports betting that trade mostly on their own, but the number is increasing and there are a lot of courses and opportunities to approach this business. In the future it could be a valuable investment tool, but, for now, the actual regulation does not permit to run this business on a large scale.

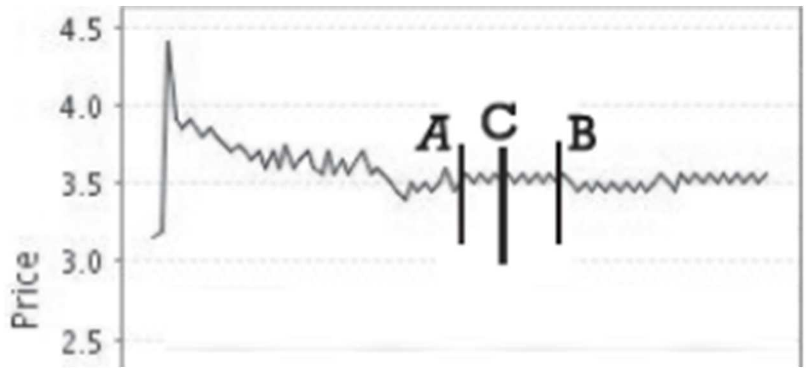
This chapter will present the main strategies and tools implemented to invest in sports outcomes. An important remark: trading is not arbitrage. Even if following a plan can reduce the risk, a trader can incur in a period of losses if money management and risk control are incorrectly applied.

2.1 Forecasting price movements in betting exchanges

The goal of a sports betting trader is to bet on one outcome at a higher price and to bet on the opposite outcome at a lower price, with the aim of making a profit regardless of what the final outcome is. Since these markets usually have a short duration (a few hours or minutes), probabilities, and therefore odds, are continuously changing throughout the sporting event, according to every influential factor. Hence, forecasting price movements correctly is the key to success in this activity. The core issue is to detect whether the odds will move up or down, for how much, when, and how fast.

Ioulianou et al. (2011), have demonstrated, using an econometric analysis, that in an exchange market a price range of 1% to 2% is very possible and repetitive. Figure 5 shows the pattern of a "draw" selection in a football match. The horizontal axis is 'time', and the vertical axis refers to the odds. The time period before point A is about one or two days before the match. In practice, the authors check values 24 hours at maximum before a game, excluding the last minutes before the event start, since, statistically, in this time interval periodicity stops, maybe due to the last-minute excitement of unprofessional gamblers who place bets without following a particular plan.

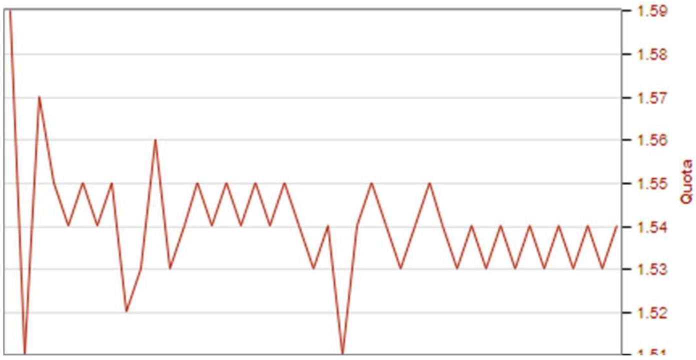
Figure 5 Price pattern example



Between A and B points there is a great opportunity for trading. In fact, the system recognizes the pattern that the price statistically varies from 3.50 to 3.55 periodically and the signal for profit opportunity is given in point C, making a buy-sell or sell-buy strategy. The price is going up and down by a small amount, but it is doing it continuously. The accuracy of the proposed system lies on the fact that it is more than 80% possible that this periodical trend of price range between two values will continue, if we observe at least 5 to 6 repetitions of the pattern. In the other cases, the unexpected trends were equally distributed between favourable and unfavourable cases, thus without significantly affecting the expected returns. In this situation, we can have an almost risk-free arbitrage opportunity with a low but not negligible profit (approximately 1% - 2% of the money invested). The strategy applied would be the buy-sell or the sell-buy one. The next subchapter will explain it in detail.

This kind of pattern is very common in this market. Figure 6 shows an example from www.betfair.com of a world championship qualification match.

Figure 6 Betfair example



An opportunity like this one can be detected manually, by just looking at these charts, or automatically, using artificial intelligence. In the paper by Ioulianou et al. (2011), the authors have created an informatic system that alerts them every time there is an opportunity. As shown in Table 1, in three months of trading, they could increase their profits remarkably, updating their system continuously.

Table 1 Experimental result

Details	First Month	Second Month	Third Month
Average Alerts Per Day	125	240	520
Profit per Month (initial capital 1000€)	18%	33%	54%
% of Advantageous Changes	9.6%	9.58%	11.53%
% of Disadvantageous Changes	16%	11.25%	10.38%

Another useful tool for predicting odds movements is the past. A recent study by Arbeloa et al. (2010) has carried out a case-based system which captures some features of a current event and finds similarities with other past events. Then, observing the price evolution in these historical events, the agent will be able to predict the more accurate future prices depending on what happens during the event. Thus, forecasts refer to an ongoing match. The authors have used a data set from soccer matches played in the 2008-2009 season in the Barclays Premier League, the first UK football league and the most popular league in the world. Consequently, it is also a high liquidity one, that is important for obtaining more reliable results. They focused their research on the price prediction for the under/over 2.5 goals market[8]. Although all the football events are completely unrelated (different players, different teams, weather conditions and so

[8] It refers to the sum of goals in a match: 0, 1 or 2 goals are "under", more than 2 goals are "over".

on), if we consider the prices of the outcomes of different markets, we can find some similarities between an unknown event and a past one. The first step is the data acquisition. Similarities are found with respect to these features:

- The exact minute of the game.
- The current score.
- The prices of the under/over 2.5 markets.
- Odds prices for the home wins, visitor wins and draw wins.

Every selection includes back and lay prices. After a data filter process (for excluding samples which may not reveal a real probability at a given moment), the CBR-agent creates the case base which will be matched to past data to forecast future odds at 1, 5, 10 and 15 minutes. The results are promising: future events follow price movement patterns similar to past events. As the agent increases the number of past observed cases, the accuracy of the future price prediction is also increased. With 250 cases in the case base, the precision accuracy is around 90% for an error rate of 0.05. Moreover, for predictions in the next minute this accuracy is almost 100% for the same error rate, while the next 5, 10 or 15 minutes have lower but similar precision. In conclusion, the authors have demonstrated that despite each event being different, under similar circumstances some price movement patterns are repeated.

This result is reinforced by another study again by Arbeloa et al. (2021), which demonstrates that the CBR agent is more accurate than a sample of 20 traders, which proves that the human mind can lead to some involuntary biases.

This information puts the basement for a successful sport trading activity and underlines how experience is fundamental in this business.

2.2 Investing strategies

Sports trading can be considered a low-risk speculation activity. Investors only care about how prices are moving, and their mission is to buy low and sell high. In this business, a profit is always generated if you are able to exchange odds related to the same outcome, where the back one is higher than the lay one. This strategy can be done both in a pre-match situation and in live when prices are very volatile. To prove it mathematically, suppose that a trader makes a bet on Betfair by risking μ units at a price of ρ_1 for a specific outcome ν. If the bet is finally won, the trader wins $\mu\times(\rho_1-1)$ units and loses the μ units if the other outcome is the final one.

Now, the player decides to bet the same amount μ on the lay side of the same outcome, then, if ν is indeed the final result, he will win $\mu\times(\rho_1-1)$ and lose $\mu\times(\rho_2-1)$. The final profit is then:

$$\mu \times (\rho_1 - 1) - \mu \times (\rho_2 - 1) = \mu \times (\rho_1 - \rho_2)$$

The final profit, whether ν is the final outcome, will then be positive if $\rho_1 > \rho_2$, that is if the back price is higher than the lay price. In the case that ν is not the final result, the trader will not lose any units, since he wagered the same amount on both sides of the market. The same reasoning can be applied if you first lay and then back. An important thing to highlight is that Betfair applies its commission only on the net cash flow, so if your bets balance each other, we would not see any cash flow, thus we would not be charged. The fee is applied exclusively if you get net profit on that outcome. A trade like the above one can also be set in "all-green" position, that in jargon means to split the potential profit on each possible outcome, in order to gain the same amount whatever will be the final result. For example, if we back €100 on a draw at 3.55 odds, we can lay at a lower tick (3.55/3.5) ×€100 ≈ €101.43. In this way, we would earn €1.43[9] whatever would be the final score. The general formula is: $\frac{\rho_1}{\rho_2} \times \mu = \theta$, where θ is the amount to lay. This is what the cash out does when the market has already moved. Combining this plan with an accurate statistical analysis, focused on recognising the best opportunity, can lead to a low-risk 1%-2% profit. In particularly liquid and stable matches, this margin can be reached just offering back and lay bets at the best respective available odds in a reasonable advance with respect to the match beginning. In this way there is a high probability that the market will absorb the liquidity, and both our bets will be matched. If we are able to match both our bets, a sure profit will be reached automatically.

The strategy that exploits odds fluctuation by buy and sell or sell and buy is called "scalping", similarly to the financial one. It consists in closing a lay position at a lower odds than the back one as quickly as possible. The second position can be closed both in a pre-match or live transaction using the cash-out. The ROI percentage is usually around 1% or 2%, but, depending on the willingness to risk, it could be higher during live match betting. Nevertheless, the intrinsic risk is quite low, since if the market does not go in the expected direction, it is always possible to close the open position using the cash-out, thus limiting losses. Furthermore, the maturity of the entire strategy is usually very short, could be only a few minutes, hence traders repeat it frequently, generating at the end a considerable profit.

[9] The net profit would be €1.36 with a 5% commission.

There exist several strategies to run this business, depending on the player's attitude to risk and the money he can afford to invest. Sports betting is not a science, but it can be profitable with a good combination of statistical analysis, market knowledge, predetermined strategy, and personal effort. A good investor should detect the match where it is worth to enter in the market by scanning several statistics of both teams and players involved. After that, he can decide which strategy fits better for that match. The most popular ones in football are the following:

- Lay the draw: enter in a pre-match or in-play bet when the result is in equilibrium but both teams are forcing, and then immediately cash-out when a goal occurs.
- Super back: live bet towards the favourite team in the first half, cash-out with a goal or at the end of the first half if no one scores (incurring in a loss of 15%).
- Super over: back bet on over 2,5 goals where statistical analysis suggests that someone will score in the first half. It is more convenient to enter in the first 5-10 minutes, because the odds usually increase about 10-15 ticks. Cash-out when a team scores or at half time if the result is still 0-0.
- Super under: pre-match bet on under 1,5 and cash-out after 10-15 minutes.

Professional traders perfect their work using informatic programs, like "Fairbot", that allows them to set the strategy on it. In other words, the trader can set the order to perform on it, and the software will execute them once the pre-determined condition set is satisfied. For example, in a super over strategy, it is possible to ask to "Fairbot" to immediately perform a cash-out if a goal is scored, or if the first half is ended 0-0. In order to protect the portfolio from huge losses, it is very important to follow the strategies strictly and blindly, because, once the match is over, there is no way to rescue this position. In the financial market, instead, if your investment is going badly, you can keep your assets and wait for their recovery. Besides that, sports trading shares common features with financial trading (riskiness, knowledge and experience requirement etc.). One of the biggest points of difference is the psychology that involves gambling in general, which can dangerously affect trader rational decisions.

2.3 The role of psychology

Human beings are not machines, and their rationality is undermined by feelings and emotional impulses. Especially in gambling games, this issue can bring them far from being winners, even if they are trying to be apathetic. Moreover, gambling can become a serious illness; thus, this aspect must not be undervalued.

The first question that is legitimate to ask is why people gamble or bet, despite the negative expected return characteristic of gambling products. The two main agreed explanations are: the wealth maximizing and the pleasure seeking. The former views gamblers as economic agents who invest money in gambling aiming at maximizing wealth. They are willing to bear the risk to have the possibility to increase their wealth. But betting is not purely wealth oriented. The negative expected value accepted is motivated by its pleasure of participation, in fact betting could be considered as a hobby for a representative consumer and it should influence the utility function directly.

Another interesting aspect linked to human psychology is the reaction to surprising information. Financial researchers have long been interested in whether market participants react to unanticipated events in an unbiased manner, or whether they exhibit behavioural biases such as overreaction or underreaction. D. Choi and S. Hui (2014) have conducted a study that aims to explain why investors overreact to certain events while underreacting to others. They hypothesize that overreaction and underreaction are driven by conservatism and “surprise”. Testing this hypothesis on financial market data is challenging because it is difficult to a priori quantify how surprising an event is and to unambiguously measure its impact on equity prices. For this reason, the sports betting market has attracted their attention. In particular, in-play market (live betting) is the ideal setting to test the hypothesis for several reasons: first, the arrival of a goal is apparent and its impact on odds can be objectively assessed with actual match outcomes; second, new information, like a goal or red card, is reported as soon as it occurs and it immediately becomes public knowledge among all market participants, while in financial markets there can be a long delay between media reporting and the occurrence of an unanticipated event, hence causing information asymmetry issues. Furthermore, we can clearly define how surprising a goal is by comparing the strengths of the two teams: a goal scored by the “underdog” is more surprising than a goal scored by the “favourite”.

The dataset is composed of second-by-second transaction records in 2017 soccer matches of the main European championship obtained from Betfair. The authors focus on the in-play win odds of the scoring team. They measure how surprising a goal is by the difference between the implied winning probability of the non-scoring team and that of the scoring team right before the goal. Overreaction and underreaction to the first goal of the match are studied using a sequence of logistic regressions. As a result, in general, market participants underreact to new events in general, underreaction decreases with surprise, and overreaction occurs when the event is extremely surprising. These biased reactions attenuate over time and disappear at around 5 min after a goal. In addition, they demonstrate the economic significance of underreaction and overreaction by developing a strategy that bets on the scoring team when underreaction is predicted, and against the scoring team if overreaction is predicted. Through a split-sample analysis, this strategy earned a profit of 2.46% after commissions if the bets are placed at 2 min after the goal. This research proves that psychology is influencing human behaviour also in live betting, and sports trader must be aware of it.

In order to test if sports traders are able to keep their feeling out from their business, J. Alberola and A. Garcia-Fornes (2012) has compared the behaviour, regarding trading decision under the same circumstances, of eleven human sports traders and their Case based reasoning (CBR) agent, previously used to forecast price movements on Betfair. They presented 20 scenarios related to specific states of matches, and they asked the participants about their trading decisions (back, lay or no bet) that they would take according to the state of the match. Then they must select when to place the closing bet. For the informatic trader, the trading decision is dependent on the odds prediction and the ω condition, where the ω is supposed to be a goal event, and its probability of being true in the next few minutes is taken as the percentage of matches from 2000 to 2010 in which ω was true. Then, according to the under 2,5 goals probability increase or decrease and the probability of ω being true or not in the next δ minutes, the CBR agent chooses the correct combination that maximizes the profits (back and lay or vice versa) and which δ is the most beneficial for placing the closing bet.

As we might expect according to the odds prediction accuracy of the previous experiments presented in chapter 2.1, the CBR agent makes better trading decisions than humans, and this is translated into higher profits. This means that the capacity for finding value is higher in the CBR agent. Humans take into account factors such as the teams of the match, the team that is winning, the specific league of the match, etc. This causes them to choose not to bet in more scenarios than the CBR agent, and to be too prudent in some cases, reducing the amount of profit they could have made by keeping their position, or too risky in some other, incurring on losses bigger than if they would have sold their position before. Furthermore, this experiment has found out that, for humans, the level of experience seems to influence the accuracy for prediction as well as the final profit.

All these research projects have confirmed that the human instinct and psychology are relevant factors to consider when we start approaching sports trading. Humans' predictions are guided by their feelings, their emotions, the teams that are playing, or the players of the match. Consequently, this activity must be applied in a as much rational way as possible, using accurate software or following statistical analysis and reasonable strategies. In conclusion, the scientific literature has corroborated the idea that a good experience level and a self-mind control are the two main ingredients to be a successful sports trader.

2.4 Sport and stock markets

Football is the most popular sport in the world, hence it moves a huge quantity of money, not only in the professional field. Especially in Europe, it can be considered a real industry and its economic and social impact must be taken into consideration. It gives a job to thousands of people, and it can directly affect the economy of a country. For instance, the Italian victory in the 2006 World Cup had a positive influence on the Italian economy, valued in an increment of 4,1% of the GDP. The recent win in the 2020 Euro Cup has positively impacted the GDP by 0,7% and has pushed the exports by 10%[10]. This is mainly due to the international prestige that a national team can reach in such important and followed tournaments. It becomes like an advertisement for its own country.

Particularly in behavioural finance, several contributions have investigated the effect of mood and investor sentiment on asset prices. Motivated by the abundance of psychological evidence showing that sports outcomes have a strong effect on mood and that investors' mood has an influence in stock prices, some analysts have tested whether sport results drive investors' mood substantially so that the game outcomes will be powerful enough to impact asset prices. The findings are in contrast. Edmans et al. (2007) used a regression model with panel-corrected standard errors and model stock return volatility by means of a GARCH (1,1) process, using data from the World Cups and the main continental cups between 1973 and 2004, in comparison with total return indexes from Datastream. The biases are associated with domestic investors, for which local returns are the relevant benchmark. The paper documents a strong negative stock market reaction to losses by national soccer teams. The size of the loss effect is economically significant in monthly terms, the excess returns associated with a soccer loss exceeds 7%. The effect is still significant but smaller in other sports. There is no evidence of a corresponding reaction to wins and the effect is more pronounced in countries where soccer is especially important and in small stocks predominantly held by local investors.

Q. Fan et al. (2017) have presented a similar research focused on the American market. The results are similar: the loss of MLB, NBA, NFL and NHL[11] teams reduce the return of the stock

[10] Data from *https://www.ilsole24ore.com*

[11] Respectively the first American baseball, basketball, American football and hockey leagues.

market. Moreover, the magnitude of loss effect is consistent with the popularity of the games and the effect is amplified as the local characteristics strengthen.

In contrast, C. Klein et al. (2009), while looking for a relation between World and Continental Cups match results and the specific national stock index returns during the period 1990–2006, could not find any significant statistical relation. To verify if surprising results might have a stronger influence on the mood and, therefore, on the stock prices, they modified the setup, quantifying the "surprise" effect by using betting odds. But neither in this framework they could find any hint of a link between soccer outcomes and stock prices. In line with this thesis is the result from a GARCH (1,1) model presented by two American professors, Beer and Lin (2019), who have conducted an analysis regarding the connection between cricket wins and losses and the Bombay Stock Exchange.

Major leagues clubs have become real industries. Professional sports, football above all, moves billions of Euros in the main European countries, and it is a real asset in the national economy. The main first leagues and international tournaments offer premiums in hundreds of millions of Euros, but to compete in this environment is extremely costly, because competition is very tough. Therefore, winning requires investments, and, as any other regular firms, some football clubs have decided to list their company in a stock exchange market, in order to get money to invest and to improve the club's international reputation. The first team who did it was the Tottenham Football Club in 1983. The peak was reached in 2002 when 36 teams in Europe were listed. In recent years, the cost has increased and the investors' pressure for good results in the short-term has led some teams to delist from the stock exchange. Today, 22 teams are still quoted, and 3 of them are Italian: Juventus F.C., A.S. Roma and S.S. Lazio.

Given this listing practice, in the last two decades financial analysts have done some studies that aim to quantify the relation between sport results and their stock returns. First, to isolate the matches outcomes effect is not an easy task. A club's stock price is influenced by several factors, such as the managerial decisions and sponsorship contracts. Second, as we saw previously, different researchers have used different tools and different teams to assess the matches' influence, thus the results are different according to the statistical instruments used, the country, the sport and the team or championship involved. A paper by Renneboog and Vanbrabant (2000), investigates whether or not the share prices of soccer clubs listed on the London Stock Exchange (LSE) and the Alternative Investment Market (AIM)[12] are influenced

[12] The AIM is part of the LSE and designed for small and growing companies. The listing requirements of the AIM are less strict than those of the official market.

by the soccer teams' weekly sport performances. Using data from the English and Scottish championships, National Cups and European competitions, the event studies, corrected for thin trading and with Baysian updating, reveal that at the first day of trading after a game, positive abnormal returns around 1% were realised following a soccer victory. In contrast, defeats or draws are penalised, respectively, by negative abnormal returns of 1.4% and 0.6% (2.5% and 1.7% cumulatively over the week). Interestingly, much larger abnormal returns are generated subsequent to promotion and relegation games as the Premier League and European games guarantee substantially higher (future) income in terms of television broadcasting rights and sponsoring income. Hence, listed football teams results seem to directly affect their stock price.

Listed football companies have attracted the attention of the financial researchers in order to evaluate the market's ability to collect every piece of publicly available information and incorporate them in the asset price. Georg Stadtmann (2006) has tested this news model analysing Borussia Dortmund results and price movements. The football industry proves a very appropriate candidate for applying this news model due to specific characteristics: signals are very frequent and easy to quantify, occur solely when the markets are closed, become publicly available to all agents at the very same time, and have observable expectations due to the existence of betting odds. According to the news model, the unexpected part of an information should influence stock prices, while the expected part should already be included in the current price. Betting odds information are used to control for the ex-ante expected match outcome. The results show that a win influences the stock price positively, and a loss has a negative effect. Furthermore, there is no evidence that the outcomes of European matches have a higher impact on the stock price than the outcomes of Bundesliga matches.

A similar study was conducted by Palomino et al. (2005), but, in this case, the authors considered also betting odds as a source of information, since they comprise a probability distribution over possible game outcomes. For this reason, they are correlated with the other type of news: the match results. The test involved 16 British soccer clubs listed on the London Stock Exchange. First, they studied the forecasting power of the quotes and they found out that they are very good predictors of the game outcomes. Then, they examined the price reaction to game outcomes and to betting odds. The results evidence that good and bad outcomes correspond, respectively, to positive and negative abnormal returns, and there is no evidence of a market reaction following the release of betting odds by the bookmakers. In other words, the results suggest that stock markets process information about realized performances efficiently but disregard some information about future performances. Given that and the fact that betting odds are assumed to be good result predictors, these odds contain unpriced information which can be used to predict short-run share price returns. Furthermore, this study confirms the hypothesis that the higher the level of salience, the faster the public information is processed by investors. Betting odds are publicly available but are only posted on bookmakers' websites and in their physical shops. In contrast, game results are extensively discussed in all daily newspapers, on the television news, and in a variety of sports shows on prime time. Therefore, odds info is slowly incorporated into share prices.

One more recent research has confirmed these main findings. Castellani at al. (2013) have analysed the relationship between soccer match results, betting odds, and stock returns of all listed European soccer teams. The data set includes all national and international match results and pre-match betting odds for 23 listed European soccer teams, in the period between 2007 and 2009. To measure the aggregate market reaction to match results they used an event-study approach. Their study confirms that wins are followed by positive abnormal returns, draws and losses are followed by negative abnormal returns. Furthermore, abnormal returns following losses are larger in magnitude than those following wins. Even with different intensities, abnormal returns follow not only unexpected results but also expected results. Executive managers and investors should exploit this evidence to their advantage. Pre-match betting odds offers unpriced information that must be used to anticipate market reaction. For example, the fact that abnormal returns following expected losses are larger in magnitude than those following unexpected wins indicates that the market is more sensitive to (expected) negative news than to (unexpected) good news. Thus, managers of listed teams could implement strategies to reduce their firm's downside exposure while providing the potential for positive abnormal returns in case of favourable events. Investors should choose to rebalance their personal portfolios by buying stocks of teams that are expected to win, and by selling those of teams that are expected to lose. A new contribution of this paper is the goal difference analysis: it has an impact on the way investors perceive expected and unexpected match results, and on their mood and emotional state. Hence, positive and negative abnormal returns are magnified by goal difference.

To conclude, we can state that there is a double direct connection between football and the stock market. The first, especially from national teams, refers to the results influence on human mood, which can affect financial performances. However, there is also literature in contrast with this hypothesis. In general, it depends on the statistical instruments used, the country, teams and sport analysed. The second relation is the one that connects match outcomes with the stock value of the listed teams. In this case, even if sports results are not the only factor affecting their share value, having a good knowledge of this world can help to invest successfully in this sector, and betting odds, with their predicting power, can be a helpful tool to anticipate the market.

3 Is the bookmaker market efficient?

A matter of considerable importance in economics and finance is how relevant information becomes impounded in market prices. The theory of efficient markets is one of the most important paradigms in economics and finance, referring to the process of price formation in financial markets. The most important contribution to the efficiency literature was given by E. Fama (1970), who formalized the Efficient Market Hypothesis (EMH). The primary implication of this hypothesis is that, given the intrinsic uncertainty, the price of a good serves as an unbiased predictor of the good's future value given all publicly available information. In other words, the efficient market hypothesis states that financial markets are informationally efficient if they are perfectly competitive and security prices reflect their true fundamental value, that is the price at which the counterparts would agree to trade if all the relevant information is known. Then, the price should be the discounted value of the security's future cash flows. Changes in market prices are random since they are influenced only by new random information, that is unpredictable. Hence, it is impossible to make an estimation of the new information. The efficient markets hypothesis predicts that asset prices will incorporate all the relevant information, and in the simplest interpretation, will do so immediately and completely. This implies that while facing new relevant information agents are rational and able to update their expectations correctly. Consequently, if the EMH holds, it should be impossible to consistently earn above average returns on a risk adjusted basis. If a discrepancy exists between the market price and the information based fundamental value, an arbitrage opportunity would arise, and only the informed investors would immediately exploit it and make it vanished, at the expense of the less well informed.

Depending on the meaning of "all available information", Fama has distinguished three types of efficiency:

- Weak: past prices alone cannot predict future prices.
- Semi-strong: all publicly available information doesn't predict prices.
- Strong: a special group is not able to achieve a higher than the average rate of return due to the group's monopoly over specific information.

If the weak efficiency form holds, prices are composed of only three components: the last period's price, the expected return of the asset, and a random error term with an expected value of zero. Hence, future price movements must follow a random walk and no one can obtain abnormal returns with a technical study by analysing prices from the past.

Semi-strong form efficiency implies that all publicly available information, such as facts about firms' products, operations, patents, and balance sheets, are reflected in prices of relevant financial assets fully, immediately, and in an unbiased manner. Thus, in addition to historical prices, it suggests that excess returns are impossible to generate by basing investment decisions on any publicly available information or any type of fundamental analysis.

The highest level of efficiency, the strong one, assumes that all available information, both public and private, is reflected in asset prices. In this case, excess returns cannot be achieved in the long run even if an investor holds insider information.

After the formulation of this theory, financial markets 'efficiency has been tested with different methods, raising both supporting and opponent results. Recently, behavioural finance studies have criticised this scheme, proposing an evolutionary alternative to market efficiency: the Adaptive Market Hypothesis (AMH). According to it, instead of being an all or none condition, market efficiency is a characteristic that varies continuously over time and across markets.

Scientific literature has spent a considerable effort trying to prove the EMH in the financial market. Regardless of the amount of evidence supporting the EMH, it can never be perfectly demonstrated. Two decades after the first publication of the efficiency theory, the same Fama has admitted that market efficiency itself is actually not testable because it must be tested jointly with some model of equilibrium, i.e. with an asset pricing model. Thus, even if we find anomalous evidence of the behaviour on some asset returns, it will remain unclear whether these anomalies are really due to market inefficiency or, alternatively, an incorrect model of market equilibrium. This challenge is called the joint hypothesis problem and because of it, despite the vast number of studies applying the EMH, it does not seem to provide infallible methodology for analysing efficiency of financial markets. The focal phrase of the EMH, that prices fully reflect all available information, is a statement about two distinct aspects of prices: the information content and the price formation mechanism. This drawback has been thereafter widely recognized in the literature, allowing many other authors to conclude that the EMH per se is a blurred theory, because, at the same time, it is not a well-defined hypothesis, but neither empirically refutable.

As in the case of the evaluation of the "surprise" effect, due to the difficulties in formulating direct tests of efficiency, the financial literature has looked beyond conventional financial markets to find a setting that would be better suited for tangible studies of efficiency. Again, the sports betting market has been considered as a good proxy of the stock market. This because of several sports betting characteristics: first, in sports betting markets, market participants receive an objective signal about the fundamental value of an asset quickly, often even within a couple of hours. A stock could be basically infinitely lived, its value today depends both on the present value of future cash flows and on the price that will be paid for the security tomorrow, while each asset in sports betting markets has a well-defined termination point at which its value becomes certain. Thus, efficiency tests do not have to focus on predictability of asset returns, which simplifies inferences about the learning behaviour of investors. Second, the range of possible asset payoffs is simple and often known with certainty in advance. Third, betting market participants are in general well-informed, motivated and experienced, and sports breaking news are usually reported promptly and are easily shared and processed by the agents. In other words, there is very little room for information leakage which, conversely, affects financial market efficiency. Fourth, tests of efficiency in sports betting markets reduce the scope of the pricing problem and almost completely remove the joint hypothesis problem, or at least mitigate it. In fact, this issue is impossible to delete completely, since no one can compute the exact real probability of an event, but, in case of a bet, the payoff structure is fixed, simple, with a short maturity and certain, both temporally and quantitatively. All these factors, together with the easily accessible information that makes the price setting mechanism more reliable, allow us to overcome almost fully the joint hypothesis problem.

Thaler and Ziemba (1988) are among the first to propose that sports betting markets have a better chance of being efficient because of their conditions, such as quick and repeated feedback, that tend to facilitate learning. After them, many other authors have later elaborated these views on the suitability of sports betting markets for tests of efficiency and choice under uncertainty. In the context of betting markets, efficiency has been viewed through two different lenses: statistical and economic. The former implies that betting odds, that are the probabilities determined by the information subset available, are unbiased predictors of actual results; the latter implies that it is not possible to systematically earn profits following any betting strategy. Thus, contrary to financial markets where market efficiency is commonly viewed only through the economic lens, betting markets provide an additional, statistical lens to shed light on efficiency. Anyway, statistical betting market inefficiency does not necessarily indicate economic inefficiency, mainly due to the transaction cost incorporated in the bookmaker margin.

While tests of statistical efficiency are primarily of academic interest, tests of economic efficiency represent the stricter and decisive test of efficiency, being full of practical interest. Statistical tools are used also to measure economic efficiency, in order to build arbitrage betting strategies. If there exist systematic biases in the market’s ability to incorporate relevant information into odds, one might be able to formulate betting strategies that would profitably exploit these biases. As in the financial market, once you have detected the inefficiency, it is important to measure the extent of inefficiency, because, to generate profits, the larger the biases the greater can be the transaction costs, and vice versa. In the current competitive online betting markets, characterized by low transaction costs, deviations from statistical efficiency do not have to be large to allow profitable betting.

The distinction between the three types of financial efficiency applies in a natural way to betting markets, where asset prices are replaced by betting odds. In his approach, weak form efficiency equates available information with historical odds and returns, semi-strong form efficiency adds public information to the set of available information, and strong form efficiency allows for the existence of particular market participants who may act as insider traders exploiting their monopolistic access to specific information. Hence, empirical examination of betting market efficiency has many shades, depending on the form and lens through which efficiency is considered. Each form of efficiency can be examined with either pure statistical tests or direct economic tests; the former tests look at statistical properties of betting markets, while the latter tests attempt to detect unexploited profit opportunities.

3.1 The efficiency of the fixed-odds betting market

Even before the rise of the exchange platforms, scientific literature has spent its effort trying to detect inefficiency in the fixed-odds betting market. This statistical inefficiency research has been followed by an economic one, that aimed at building profitable betting arbitrage strategies. The first step consists in testing the market against the weak efficiency hypothesis. Statistical betting market efficiency requires that market probability distributions (i.e. odds), conditional on the relevant information subset, are equal to objective probability distributions, conditional on the same information set. If the objective probabilities forecasting model produces information about the match outcome probabilities that is not already reflected in the odds quoted by the bookmaker, then the odds fail to satisfy standard weak-form efficiency criteria: all historical information relevant to the assessment of the match outcome probabilities should be reflected in the quoted odds, that must be the best forecasted probability of the outcome. Weak form efficiency implies that no abnormal returns can be achieved using only price information and technical analysis. Thus, to identify some biases is the basement on which to formulate arbitrage strategies.

3.1.1 Price setting mechanism

The first issue that needs a detailed study is how bookmaker companies set their prices. We know that they employ odds compilers who have special knowledge of specific sports and computational software in order to estimate the true probability of each possible outcome. So far, even some bookies 'mangers have admitted that it is not always in their interest to set odds equal to their objective probabilities. The real point of interest is to understand the criteria used by the bookies to set the prices. In particular, we want to know whether bookmakers want to neutralize their counterparty risk by balancing each side of the book, or what other strategies they persecute.

Lee and Smith (2002, p.5) write: "Bookies do not want their profits to depend on the outcome of the game. Their objective is to set the point spread to equalize the number of dollars wagered on each team. The margin compensates bookmakers for making a market and for the risk they bear that the odds or total line may be set incorrectly". This approach protects the odds compilers companies by the risk of incurring in huge losses, and at the same time it moves the core activity of this firms from focusing on predicting a probability as more precise as possible to focusing on the characteristics and behaviour of their clients. But, if they are capable of doing both, they can substantially increase their expected profits by setting odds according to their customers' preferences, and not reflecting their estimated probability. For example, if bookmakers know that local bettors prefer local teams, they can skew the quotes in favour of the visitors.

Levitt (2004) has shown that this theory is not far from being realistic. Analysing the behaviour of 285 bettors playing bets at an online bookie on the National Football League[13], the first American football league, for a total of almost 20,000 bets, he demonstrated that the bookmaker does not wish to set prices in order to balance their book. In almost one-half of the cases, at least two-thirds of the bets fall on one outcome. Furthermore, they often set a spread that is not able to embed all the relevant information (e.g. which team is the home team), that would push the market closer to an equalized condition. For example, when the home team is an underdog, on average two-thirds of the wagers are placed on the away club. An explanation for this setting price behaviour emerges in the paper's second finding (p.7): "bookmakers appear to be strategically setting prices in order to exploit bettors' biases". Bookmakers find more convenient to set prices in such a way that the favourites teams win in reality only less than

[13] In this sport, since draws are rare, odds are offered head-to-head, that means betting on a winner including the eventual overtime.

50% of the times. This is true because bettors seem to be affected by a favourite bias, as in any case more than half of the total bets fall in the favourite side of the book. In this way, their profits are estimated to be greater by 20–30% compared to a risk avoiding setting price technique. They do it as much as possible, but without allowing for a strategy that bets only on low probability events to be profitable. In short, being an efficient character by letting supply and demand regulate prices, is not a profit maximiser behaviour for fixed odds bookmakers.

Kuypers (2000) has shown mathematically that bookmakers can maximize their profit without imposing the odds to be equal to the calculated probability, and neither to be balanced on each side of the possible outcome. His model assumes that bookmakers have no private information, but they are able to evaluate publicly available information as well as any other individual or organization. We suppose that the punter is betting on a football match, hence the possible outcomes are home win, draw or away victory, that are represented respectively by 1, X or 2. Let H be the handle, that is the total amount invested on this match, and h_1, h_x and h_2 be the amount bet on each outcome. Thus, the share of H on each outcome is defined as:

$$s_1 = \frac{h_1}{H}\ ; \qquad s_x = \frac{h_x}{H}\ ; \qquad s_2 = \frac{h_2}{H}\ ;$$

Suppose that the bookmaker's subjective probability that the result of the event would be 1, X or 2 is represented by b_1, b_x and b_2. By definition, $b_1 + b_x + b_2 = 1$. The bookmaker's posted odds is denoted by o_1, o_x and o_2. As we are assuming that the margin of the book is 11%, then:

$$\frac{1}{o_1} + \frac{1}{o_x} + \frac{1}{o_2} = 1{,}11$$

Now, let d be the implied probabilities from the odds, so:

$$d_1 = \frac{1}{1.11 * o_1}\ ; \qquad d_x = \frac{1}{1.11 * o_x}\ ; \qquad d_2 = \frac{1}{1.11 * o_2}\ ;$$

By definition, $d_1 + d_x + d_2 = 1$.

Supposing that the players accept the bookmaker's price for offering his service (i.e. the margin of the book), the model is only concerned with how punters distribute their bets over the three outcomes. Thus, the punters' reaction functions are only used to determine the share of the handle (s_i) which is bet on each outcome. Throughout the model it is assumed that bookmakers know the appropriate punters' reaction function. Bookmakers are supposed to be risk neutral and as such are expected profit maximisers. The single bookmaker's expected profit is:

$$E(\pi) = H - b_1(h_1 o_1) - b_x(h_x o_x) - b_2(h_2 o_2)$$

That is nothing else than the handle less his subjective probability of each outcome multiplied by the pay-out for each outcome. Winning punters receive the decimal odds multiplied by their stake ($h_i o_i$). Substituting $h_i = Hs_i$ and $o_i = \frac{1}{1.11 * d_i}$ into the bookie's expected profit equation:

$$E(\pi) = H - b_1\left(Hs_1\frac{1}{1.11 * d_1}\right) - b_x\left(Hs_x\frac{1}{1.11 * d_x}\right) - b_2\left(Hs_2\frac{1}{1.11 * d_2}\right)$$

Note that $d_1 + d_x + d_2 = 1$ and therefore d_2 = 1 - d_1 - d_x. Rearranging terms it becomes:

$$E(\pi) = H - \left(\frac{b_1 Hs_1}{1.11 * d_1}\right) - \left(\frac{b_x Hs_x}{1.11 * d_x}\right) - \left(\frac{b_2 Hs_2}{1.11 * (1 - d_1 - d_x)}\right)$$

The bookmaker wishes to maximize its expected profit by setting odds given the players' reaction function. Looking at the equation above, the bookmaker maximizes profit by setting the implied probability d, as each d implies a unique odd. It is assumed that the punter compares his subjective probability of an outcome with the odds. He will bet on outcomes that he believes are more likely than the odds suggest. Thus, the share bet is a function of the implied probability from the bookmaker's odds and the distribution of players 'subjective probabilities over the three outcomes called P. Hence, $s_i = f(d_i, P)$. The bookmaker's expected profit function becomes:

$$E(\pi) = H - \left(\frac{b_1 Hs_1(d_1, P)}{1.11 * d_1}\right) - \left(\frac{b_x Hs_x(d_x, P)}{1.11 * d_x}\right) - \left(\frac{(1 - b_1 - b_x)Hs_2(d_1, d_x, P)}{1.11 * (1 - d_1 - d_x)}\right)$$

Differentiating with respect to $d_1$and d_x the bookmaker's decision variables:

$$\frac{\delta E(\pi)}{\delta d_1} = -\frac{\delta s_1(d_1, P)}{\delta d_1} * \frac{b_1 H}{1.11 d_1} + \frac{b_1 Hs_1(d_1, P)}{1.11 d_1^2} - \frac{(1 - b_1 - b_x)Hs_2(d_1, d_x, P)}{1.11(1 - d_1 - d_x)^2}$$
$$- \frac{\delta s_2(d_1, d_x, P)}{\delta d_1} * \frac{b_2 H}{1.11(1 - d_1 - d_x)} = 0$$

$$\frac{\delta E(\pi)}{\delta d_x} = -\frac{\delta s_x(d_x, P)}{\delta d_x} * \frac{b_x H}{1.11 d_x} + \frac{b_x Hs_x(d_x, P)}{1.11 d_x^2} - \frac{(1 - b_1 - b_x)Hs_2(d_1, d_x, P)}{1.11(1 - d_1 - d_x)^2}$$
$$- \frac{\delta s_2(d_1, d_x, P)}{\delta d_x} * \frac{b_2 H}{1.11(1 - d_1 - d_x)} = 0$$

In this model, the market is efficient when $d_i = b_i$, therefore when the implied probability from the odds is equal to the bookmakers' subjective probability, which is assumed to be the best possible subjective probability. Imposing it in the first derivative equation and simplifying, it gives:

$$s_1(d_1, P) = b_1 \left[\frac{\delta s_1(d_1, P)}{\delta d_1} + \frac{s_2(d_1, d_x, P)}{(1 - b_1 - b_x)} + \frac{\delta s_2(d_1, d_x, P)}{\delta d_1}\right]$$

$$s_x(d_x, P) = b_x \left[\frac{\delta s_x(d_x, P)}{\delta d_x} + \frac{s_2(d_1, d_x, P)}{(1 - b_1 - b_x)} + \frac{\delta s_2(d_1, d_x, P)}{\delta d_x}\right]$$

Thus, in order for the market to be efficient, the function that determines the shares evaluated at $d_1 = b_1$ and $d_x = b_x$ must satisfy the two equations above. This is not necessarily the case, and a situation can be imagined in which the expected profit maximizing implied probabilities (in effect odds) are not equal to their subjective probability.

To clarify, a numerical example can be helpful. In a football match between Liverpool and Manchester United there are ten punters, each one betting 1 unit. We suppose that these players bet on the outcome which maximizes the difference between their subjective probability (p_i) and the probability implied by the odds (d_i). If for each outcome the subjective probability equals the probability implied by the odds, they bet on the most likely event, that is the one with the lowest odds. We assume that there are two types of players, six Manchester United fans and four neutrals. The firsts fans believe that Manchester have a better chance of winning the game than the bookmakers and have subjective probabilities $p_{1mu} = 0.4$, $p_{xmu} = 0.2$, $p_{2mu} = 0.4$. The neutrals are of the same opinion as the bookmaker, so $b_1 = p_{1n} = 0.5$, $b_{x=} p_{xn} = 0.2$, $b_2 = p_{2n} = 0.3$. Now assume that the bookmaker sets the market efficient level of odds $d_i = b_i$. In this situation, all the Manchester United fans would bet on the away win, since it maximizes their probability difference, and all the neutral fans on the home win, because it is the lowest quote. Hence, the shares of the bet are $s_1 = 0.4$, $s_x = 0$ and $s_2 = 0.6$. Then, the bookmaker's expected profit is:

$$E(\pi) = 10 - \frac{0.5 * 10 * 0.4}{1.1 * 0.5} - \frac{0.2 * 10 * 0}{1.1 * 0.2} - \frac{0.3 * 10 * 0.6}{1.1 * 0.3} = 0.99$$

However, the bookmaker could set odds to take account of the home team supporters' bias. For example, he could set odds such that $d_1 = 0.41$, $d_x = 0.20$ and $d_2 = 0.39$. Under the same conditions, the punters would bet in the exact same way, thus the new expected profit would be:

$$E(\pi) = 10 - \frac{0.5 * 10 * 0.4}{1.1 * 0.41} - \frac{0.2 * 10 * 0}{1.1 * 0.2} - \frac{0.3 * 10 * 0.6}{1.1 * 0.39} = 1.45$$

Thus, using the assumed punter reaction function, it is demonstrated that the bookmaker can increase his expected profit by setting market inefficient odds.

3.1.2 Biases in the football betting market

Bettors' biases are intended as the inefficient behaviours of the players, while they are choosing which odds to bet on. Bookmakers are particularly interested in detecting biases because they can maximize their profit if they systematically exploit them, instead of simply equalizing the total amount of stacked money. Forrest and Simmons (2008) have measured the sentiment bias in the Spanish online football betting market. They explain that markets seem to be affected by a sentiment bias, in the sense that bettors can have a nonfinancial behaviour, thus they wager on a particular team partly because of emotional attachment. Even if they maintain a certain degree of odds sensitivity, they will probably give up the bet rather than back their team's opponent if the price goes above their reservation price. In this case, bookmakers would find it worthwhile to shift odds in favour of supporters of the bigger team if they accounted for a sufficiently large part of the market for that match and if their demand is sufficiently price-elastic. At the end, competitive forces may then generate more, rather than less, favourable prices for those interested in betting in accordance with non-financial preference, in other words the sentimental bettors.

The most well-known bias reported in the betting market literature is unquestionably the favourite-longshot bias (FLB). Its first documentation is attributed to Griffith (1949), who observed that horses with short odds (favourites) yield on average higher returns than horses with long odds (longshots, underdogs). Consequently, market probabilities of longshots overpredict on average their empirical probabilities (computed from race outcomes) and the favourites' winning probabilities are, on average, relatively underestimated. In other words, the FLB implies that placing bets on favourites yields a higher return than placing bets on underdogs. As a consequence, the betting public has been observed to have a systematic tendency to overbet longshots and underbet favourites.

This phenomenon has been identified in a variety of sports and also in the financial market. Rubinstein (1985) detects a similar bias also in the equity options market: he finds that shorter maturity options (i.e. longshots) are overpriced. Moreover, Hodges et al**.** (2003) demonstrate that the FLB exists for the put options, for 3 months and 1 month horizons, on S&P 500 and FTSE 100 markets, where investors tend to overpay for all put options as an expected cost of insurance.

Two main theories have been proposed in the literature to explain the favorite-longshot bias: first the risk-loving betting behaviour, second the bettors' misperception of probabilities and the resulting price-adjustment by the bookmakers. The first states that if individual bettors love bet on high and risky odds, they are willing to accept a lower expected payoff when betting on longshots, which are also the riskiest investments. In other words, they are more attracted by the opportunity of earning a huge amount of money than to have good chances of winning a less significant amount. Alternatively, some agreed behavioural theories suggest that cognitive errors and misperceptions of probabilities play a role in market mispricing. These theories demonstrated, for example, that people find difficult to distinguish small and tiny probabilities and hence price both similarly. Conversely, the show an inclination to choose certain rather than extremely likely outcomes, leading highly probable bets on an event to be under-priced.

A deep analysis on the causes of this bias has been conducted by Snowberg and Wolfers (2010), who verified which class of models between risk-love or misperceptions has a stronger causing power, or if they perform similarly. Their results attribute the causes of the favorite-longshot bias to misperceptions rather than a risk-love behaviour. Moreover, this bias is likely to persist in equilibrium because misperceptions are not large enough to generate profit opportunities for unbiased bettors. That said, the consequences of this bias are also relevant, and debiasing an individual bettor could lead the gamble activity to a more efficient level.

Another branch of the literature about this bias has the aim to measure the degree of presence of this phenomenon. At the beginning, the research focused mainly on horse races, since the wider number of victory contenders fits better for this bias scheme. Nevertheless, evidence of favorite-longshot bias has been found also in the football betting market. Paton and Williams (1998) have employed a tobit regression with a small sample of 1X2 Premier League odds and they have detected a favorite-longshot bias, explaining the phenomenon with transaction costs. Also in English soccer, Cain et al. (2000) reported evidence of a FLB, similar to that found in horse racing, both for match results and scores when comparing the estimated fair odds with a bookmaker's actual odds. The data comprise the results of the 2855 Football League matches played in the UK during the 1991-92 season, together with the associated odds against each simple outcome (home win, away win or draw) and against each score quoted by the bookmaker William Hill. For the full sample of matches there is some evidence of the favourite-longshot bias: bets on longshots generated substantially lower returns than bets on favourites, but there is no evidence of potentially profitable betting strategies on favourites.

Also Deschamps and Gergaud (2007) tried to detect this bias in the English Premier League. The data set is composed of 8377 matches during seasons from 2002/2003 to 2005/2006. For each match, they considered the quoted outcome odds (home win, draw and away win) of six bookmakers. They detected biases analysing the relationship between odds and return, supposing that for a bet to be "fair" the expected value of a bet must be zero, thus there should not be evidence of abnormal return. They split the dataset creating a group for each possible outcome, thus three groups, and they analysed each group separately. For each type of outcome, they ranked the odds according to their implicit probability and they divided them into five categories, corresponding to a range of probability. The results showed evidence of a positive favourite-longshot bias for both home and away odds. The highest return is obtained with the shortest odds, while betting on improbable events would not have been a good idea.

Interestingly, an opposite phenomenon was registered for the draw: high odds have a relative high return (-5%), while probable draws have yielded a much lower return (-11%). Their data suggest a natural explanation for this reverse bias, which is that draw outcomes are extremely difficult to predict. They indeed find that there is virtually no relationship between the draw odds and the probability of a draw outcome. This is unique to draw odds, since home and away odds are strongly correlated with the probability of home and away win. Draw odds, on the opposite, are totally uninformative.

Vlastakis et al. (2009), while looking for the favorite-longshot bias on a wide range of European football matches, have discovered another type of bias. They found evidence that the home-field advantage is consistently overestimated. This was not evident at first glance since the rule that places bets on home teams have a significantly higher average return than the rule that places bets on away teams. However, the authors warned that if we wish to examine the effects of home-field estimation error we must first account for the favourite–longshot bias effects. Combining the favourite-longshot bias and the overestimation of the home field advantage leads to a new reverse 'home-underdog' bias in the European football odds, that the authors called the 'away-favourite' bias. In summary, it states that the expected return is higher for visiting teams that are favorites than it is for favorites at home. From an extensive database of odds quoted by five major online bookmakers and one physical fixed-odds bookmaker, the authors found evidence of both favorite-longshot bias and away-favourite bias.

In a recent research project, Daunhawer et al. (2017) studied the online football betting market to verify the persistence of these biases. Their data set is composed of opening and closing odds of European major leagues of 17 different bookmakers over a period of three months, ranging from February to May 2017. Their procedure is similar to previous research: the data are divided into ten bins of equal depth and the average return r of each bin $B_i$is computed as

$$r_i = \frac{1}{|B_i|} \sum_{t \epsilon B_i} (o(t)X(t) - 1)$$

where $o(t)$ stands for the odds of a data-tuple t and $X(t)$ denotes a dummy variable that is true if the tuple turns out to be a correct prediction of the outcome. Then, they detect the biases by looking at the average return for different ranges of odds and its variation across bookmakers. In sum, their result confirms the existence of the favorite-longshot bias, but they found little evidence for the away-favorite bias. An interesting finding that conflicts with previous research by Vlastakis et al. (2009) is a relatively high average return for home-longshots, which is even positive for individual bookmakers. However, these types of matches are relatively rare, which puts some doubts on the validity of the finding and also limits its application as a possible betting strategy.

Constantinou and Fenton (2016) made use of the Rank Probability Score (RPS) for the forecast assessment. In brief, it represents the difference between the observed and forecasted cumulative distributions in which a higher difference leads to a higher penalty, that is sensitive to small sample size. RPS formula for a single match instance is defined as:

$$RPS = \frac{1}{r-1}\sum_{i=1}^{r-1}(\sum_{j=1}^{i}(p_j - e_j))^2$$

where r is the number of potential outcomes (three in this case), and p_j and e_j denote the forecasts and observed outcomes at position j (so, for example, p_j is the forecast probability of and e_j is 1 if the outcome was a home win and 0 otherwise). A lower score indicates a more accurate forecast (less error). Basing the research on the odds provided by numerous bookmaking firms for 14 European football leagues and over a period of seven football seasons (from season 2005/06 to season 2011/12 included), they showed no indications of forecast improvement in bookmaking performance over this period, and this is true for almost all of the 14 distinct leagues. Overall, the results tend to suggest that the accuracy of the odds provided for top division leagues is higher than lower division leagues. Moreover, the presence of the favourite-longshot bias is still strong and, in many cases, also profitable. For instance, in the cases of English and Spanish seasons between 2005/06 and 2009/10. Betting on outcomes with >80% chance generated average profits of approximately 2%, 2%, 8% and 10% respectively. The profitability, and contextually the reduction of losses on underdogs bets, could be explained by the recently reduced bookmaker profit margins. In particular, it may be possible that the bookmakers reduce the bias in such a way that the continuous decrease in profit margins does not increase the chances of profit when betting on favourites. The results show also a clear home-away bias whereby the returns from bets on home win outcomes generate considerably higher returns, when compared to bets on away wins.

3.1.3 Efficiency tests

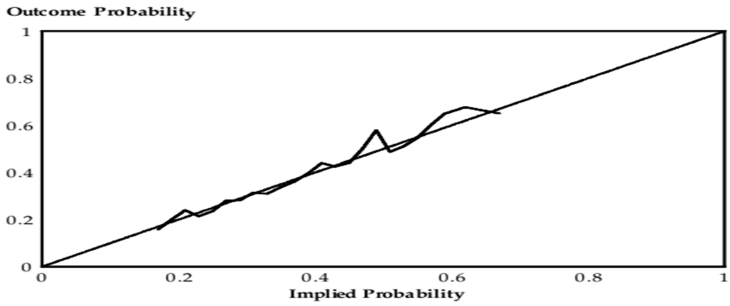
A significant number of papers have analysed football betting markets with respect to odds information efficiency. Thus, researchers have explored the value of odds in predicting football events, the existence of systematic odds-related biases, as well as the degree of variation of odds between different market operators. The studies presented in the previous paragraph are in themselves efficiency tests, since, whenever a bias is detected, the market consequently results inefficient. This section will present the most important efficiency tests published, where the efficiency is tested through statistical models, and not only comparing implied probability and returns as in the biases research.

Pankoff (1968) performs the first ever test of betting market efficiency by regressing match outcomes in the American National Football League (NFL), concluding that the market seems to be efficient on aggregate. Later on, the topic of betting market efficiency is rather developed in the literature, but the main findings are mixed empirical evidence on the degree to which betting markets are efficient. In particular, the efficiency of win-draw-lose match outcomes in football betting markets is still an open issue. Standard linear regression has been widely employed but it has been considered weak, revealing only whether a betting market is efficient in an aggregate sense.

Besides these standard models, the relationship between subjective and objective probability has been investigated using linear probability models (LPMs), first introduced by Pope and Peel (1989). They proposed an approach based on linear probability and logit models to test for betting market efficiency for the 1981/1982 football season in the UK. They concluded that there is no bias in the bookmaker's odds-setting processes for home and away wins and that no profitable betting strategy can be identified.

Kuypers (2000), further than the model of bookmakers' behaviour, ran a test about the efficient market hypothesis. He first adopted the classical test used for detecting biases, that is to divide odds in group level and then compare the profitability of each one, searching for differences between outcome probability and probability implied by the odds. In his result, there seems to be a good match between the two except around a probability of 0.5, as shown in Figure 7:

Figure 7 Implied versus outcome probability



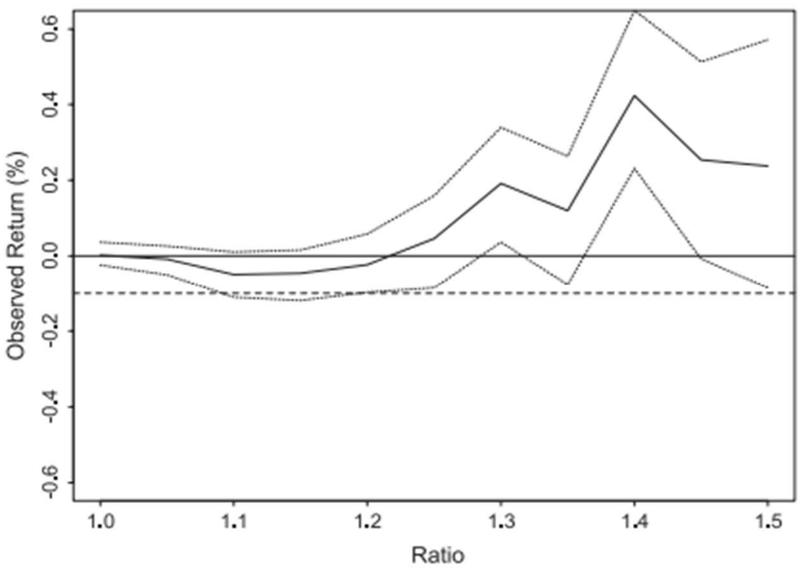
In order to validate the result, he carried out an OLS regression. The estimated equation was implied probability = b * outcome probability. The hypothesis that b = 1 (i.e. implied probability = actual probability) could not be rejected at the 95% confidence level. Thus, the regression analysis ended up with the same conclusion: there is rare occurrence of both inefficiency and profitable betting opportunities and there is no systematic bias.

Dixon and Pope (2004) have evaluated the economic significance of the statistical forecasts of UK Premier League match outcomes in relation to betting market prices. They extended the forecasting model built by Dixon and Cole (1997). This model accounts for the abilities of both teams in a match, which is proxied by summary measures of recent performances. In detail, two dimensions of ability are recognised: ability to attack (score goals) and ability to defend (not concede goals). The number of goals scored by the home and away teams are independent Poisson variables, with means determined by the respective attack and defence abilities of each side. Furthermore, the model is dynamic, in the sense that it updates the team's performance. Then, they presented a detailed comparison of odds set by three different bookmakers in relation to their forecast model predictions. The odds have very similar distributions among each bookmaker. Their posted draw odds are very concentrated, while home and away odds show a binomial distribution.

In contrast, the forecasting model probabilities displays far greater dispersion for the draw outcome. Unsurprisingly, for home and away victories, the distributions of model probabilities are relatively smooth and are not bimodal. This fact is not attributable to the restrictions in the parametric model. In the authors opinion, the model is flexible enough to display bimodality if it really existed in the data. There are many ways to explain this phenomenon. For example, if the league consisted of half very good and half very bad teams, then a histogram of estimates would tend to be very bimodal. The fact that it does not is evidence that the bimodality in the published odds is an artifact of the odds-setting process.

Moreover, they tested the existence of inefficiency in posted odds by considering the observed returns that would have been obtained with a particular trading rule. Inefficiency in the match outcome betting market will be indicated if expected returns are greater than the expected return of an uninformed, random betting strategy. To measure the profitability of this betting strategy, they calculated confidence intervals for the observed return using a bootstrapping technique. In the 1, X, 2 market, a bet is placed whenever the ratio of the model probability to the odds for a match outcome exceeds a critical value measured on the abscissa of Figure 8. The lines stand for a zero return (solid line) and the expected return under random betting (broken line). The dotted lines are 90% confidence intervals obtained by bootstrapping.

Figure 8 Observed returns



The average gross return and the 90% confidence interval obtained by bootstrapping suggest that the odds fail to capture all the information in the forecast model. The expected return conditional on the model forecast is systematically higher than the unconditional (negative) expected return. As we can see from Figure 8, above a ratio of 1.2 the expected return is positive, reaching a value well in excess of 20% for a ratio value of 1.4. The expected return is (marginally) significantly greater than zero for high ratios, even if the estimated standard error of returns increases with the ratio, because a higher "request" of quality implies less bets available. Therefore, the authors have concluded that the forecast model appears capable of generating economically and statistically significant abnormal returns. Thus, the model provides estimates which more accurately reflect the true probabilities than do the posted odds. Consequently, the market results inefficient, since a trading rule can exploit the probabilities model to achieve positive abnormal returns.

In the same year, Goddard and Asimakopoulos (2004) have published a study that tests for weak-form efficiency in the prices quoted by a prominent high street bookmaker for fixed-odds betting on English league football match results during the 1999 and 2000 seasons. If the forecasting model produces information about the match outcome probabilities that is not already reflected in the odds quoted by the bookmaker, then the odds fail to satisfy the standard weak-form efficiency criteria, that all historical information relevant to the assessment of the match outcome probabilities should be incorporated in the odds quoted. The main innovation of this research is that, for the first time, it forecasts future match results with an explicit match result forecasting model. More precisely, they built an ordered probit regression model estimated using 10 years' data, that are not only past match results data, but also several other explanatory variables. As well as past match results data, the significance of the match for end-of-season league outcomes (for example a promotion or relegation match), the involvement of the teams in cup competition and the geographical distance between the two teams 'hometowns, all contribute to the forecasting model's performance.

The scientific literature has gradually abandoned the linear probability models, in favour of probit and logit models, both binary and multinomial, basically to overcome two major issues: heteroscedasticity and allowance of unreal probabilities. By comparing the bookmaker's probabilities implicit in its odds, the model's estimated match result probabilities, and the match outcomes, the model outperforms the bookmaker in the 1999 season, while the reverse is true in 2000. In both seasons the difference in forecasting performance appears very small. Regression-based tests indicate that the forecasting model contains information about match outcomes which is not embedded into the bookmaker's odds, that can consequently be tagged as weak-form inefficient. Evidence of inefficiency is particularly strong for matches played during the final few weeks of the season in April and May. A strategy of betting on the match outcome for which the ex-ante expected return (calculated from the model's probabilities) is the highest would have generated a positive gross return of around +8% for matches played in April and May in both the 1999 and 2000 seasons. Thus, this is another evidence of economic inefficiency.

Also Angelini and De Angelis (2018) proposed a forecast-based approach to analyse the efficiency of online betting markets. They proposed an innovative approach where they tested the predictability of football match outcomes through the information contained in the odds posted in the market and modelling the bookmaker's forecast errors to formally test for market efficiency, after considering bookmaker commissions. In particular, they investigated the efficiency in the football online betting market among the 11 major national club competitions in Europe in order to find possible differences in the degree of market inefficiency. To achieve this purpose, they considered a data set composed of odds offered by 41 bookmakers over 11 years (from 2006 to 2017), for a total of 33,060 football matches.

The first variable included in the model is a dummy, y_i, which assumes value 1 if the match i is ended with the outcome under consideration, i.e. home team win, draw, or away team win. The variable y_i is then distributed as a Bernoulli with (true) probability π_i , i.e. $y_i|\Omega_i \sim Bin(1, \pi_i)$, where Ω_i denotes the hypothetical information set containing all the relevant information. As a matter of fact, the odds quoted on the online fixed-odds betting market represent the 'best' available (ex-ante) forecasts of the likelihood of the outcome of match i. Let o_i be the bookmaker odd for a particular outcome of match i (e.g. home win) and $p_i = \frac{1}{o_i}$ be the corresponding implied probability forecast. The bookmaker's probability forecast is $p_i = E(y_i \,|I_i) + k_i$, where $I_i \subset \Omega_i$ is the (actual) information set available to the bookmakers on match i and $k_i > 0$ is the bookmaker's margin, that is supposed to be equally distributed among all possible outcomes. The bookmaker's forecast error for the outcome of match i is represented by the equation $\varepsilon_i = y_i - p_i$. Under the null hypothesis of market efficiency, in general, p_i overestimates π_i, i.e. $p_i > E(y_i \,|\Omega_i)$, and, as a consequence, the conditional expectation of ε_i is not null but equals (minus) the bookmaker commission and possible price distortions resulting from bettor's bias exploitation, i.e. $E(\varepsilon_i \,|I_i) = -k_i$. Market efficiency for league $j = 1, \dots, J$ can thus be evaluated by estimating the following model:

$$\varepsilon_{i,j} = \alpha_{1,j} \sum_{t=2}^{T} \alpha_{t,j} d_t + \beta_j p_{i,j} + v_{i,j} \,, \qquad v_{i,j} \sim i.i.d.\left(0, \sigma_{i,j}^2\right), \quad i = 1, \dots, N_j$$

where N_j denotes the number of matches considered for league j and d_t is a dummy variable which assumes value 1 for season t and 0 otherwise, for t = 2, . . . , T, so that $\alpha_{1,j}$ captures the average bookmaker commission for the j-th league in season 1 (as example) and $\alpha_{t,j}$, for t = 2, . . . , T, captures the possible development over time of the bookmaker margins. They therefore imply that the bookmaker commission may vary over time and across leagues. Since the regression coefficient β_j in the equation captures the possible effect of the probability $p_{i,j}$ on the forecast error $\varepsilon_{i,j}$, the market efficiency of league j can be evaluated by investigating its statistical significance. In detail, once accounted for the bookmaker commissions (measured by the α), market efficiency would imply that the conditional expectation $E(\varepsilon_i \,|I_i)$ is zero so that a rejection of the null hypothesis H_0 : β_j = 0 would imply that market j is not unbiased.

Actually, the authors have decided to jointly consider home and away win odds, without considering draws. This is because the analysis of draws performed on the dataset confirms the findings of Pope and Peel (1989) and Deschamps and Gergaud (2007), revealing no significant relationships between draw odds, that have a low volatility, and draw outcomes for all the leagues considered. If betting markets are efficient, then the conditional expectation of the forecast errors should be equal to minus the bookmaker commissions. Therefore, from the estimation of the model it would be reasonable to expect that the estimate for the coefficient $\alpha_{1,j}$ may be significantly negative, as this parameter captures the average bookmaker margin, and not to reject the null hypothesis $H_0 : \beta_j = 0$.

The results show that, considering the mean of the odds offered by the 41 online bookmakers analysed, we do not reject the null hypothesis of market efficiency for all the leagues, except for Italian Serie A and Portuguese Primeira Liga at 5% significance level, and Greek Super League even at a 1% level of significance. All the regression slopes (except for German Bundesliga and Dutch Eredivisie) are positive thus implying that, on average, the bookmaker's forecast error tends to increase as the forecast probability increases. This is consistent with the well-known favourite-longshot bias. The estimates of $\alpha_{1,j}$ are all negative, as expected.

To improve the power of the test, it is possible to simplify the model presented above by imposing the restriction of time-invariant intercept and re-estimate the following model:

$$\varepsilon_{i,j} = \alpha_j + \tilde{\beta}_j p_{i,j} + \tilde{v}_{i,j}.$$

On average, the bookmaker commission is significantly lower than zero at least at the 5% significance level for all the leagues, except for Germany, and ranges from 2.19% for Spain to 5.24% for Portugal. The tests for unbiasedness seem to be not affected by this restricted model in terms of significance of the regression coefficients and regarding the implications of evidence of deviation from unbiasedness in Italy, Portugal, and Greece, which may be a signal market inefficiency. If bettors follow a rational behaviour, they will tend to choose the best price offered in the market, therefore, it is interesting to evaluate the degree of market efficiency not only considering mean odds, but also taking maximum odds. As in the first case analysed (with average odds), the authors did not find evidence of time-varying intercepts and, hence, they focused on the restricted model. Compared to the case of mean odds, bettors can substantially reduce the bookmaker margins as only three leagues reveal significant (negative) estimates of α (Italy, Portugal and Greece). These findings are in accordance with the results of Forrest et al. (2005) who show that, using the best available odds, the commission is virtually eliminated. For the same three Mediterranean leagues, Angelini and De Angelis also found evidence of significant estimates of $\tilde{\beta}_j$ at least at 5% significance level, which confirms that these markets are not unbiased, as found in the case of mean odds.

These two authors gave an evaluation also to the degree of market unbiasedness and whether biases' magnitude allows for profitable opportunities for bettors, which in turn would imply market inefficiency. They considered the fitted values from the estimation of the restricted models for all the possible probability values, and, for the j-th league, they derived the following expression named 'efficiency curve':

$$\hat{G}_j(p_G) = \hat{\alpha}_j + \hat{\tilde{\beta}}_j\, p_G, \quad p_G \in (0,1)$$

where $\hat{\alpha}_j$ and $\hat{\tilde{\beta}}_j$ denote the estimates of the parameters in the restricted model. When $\hat{G}_j(p_G) \neq 0$ for a fixed value of p_G, there is evidence of bias and the sign of $\hat{G}_j(p_G)$ suggests to us which side might profit from this bias. In particular, $\hat{G}_j(p_G) > 0$ would imply market inefficiency since bettors can achieve positive returns, whereas $\hat{G}_j(p_G) < 0$ would entail profits for bookmakers. Since $\hat{\tilde{\beta}}_j > 0$, $\hat{G}_j$ tends to increase with the outcome probability p_G. Therefore, there is evidence that probabilities implied in the bookmakers odds of underdogs (favourites) overpredict (underpredict) on average their empirical probabilities. This implies that longshots are under-priced and that wagering on favourites is more profitable for bettors. These results support the favourite-longshot bias. However, all the efficiency curves are below the zero line and, except in the case of the largest values of p_G for Italy, Portugal and Greece, no significant positive values of $\hat{G}_j(p_G)$ can be achieved. This empirical conclusion implies that bettors cannot systematically obtain abnormal positive returns. Therefore, although there is evidence of biases, when mean odds are considered the online betting markets are found to be economically efficient, because it is not possible to systematically earn from them.

In the second case, thus when the maximum odds posted in the market are considered, $\hat{G}_j$ suggests lower profitability for bookmakers as the values for $\hat{\alpha}_j$ are now close to zero. According to the 95% confidence level, neither bookmakers nor bettors seem capable of achieving significant returns in eight leagues. Therefore, if the best odds are considered, the online betting markets can be defined as efficient, in the sense that the biases are too small to overcome the bookmaker margin and consequently there are no profitable opportunities for bettors. This is valid for all the leagues analysed, except for the Italian Serie A and the Greek Super League, which allow for significant returns for punters when p_G is close to one as well as significant profits for bookmakers when p_G is close to zero. Spain is an isolated case, as it is profitable for central values of p_G.

A rather simple betting strategy, which consists in defining a profitable range when the probability of a j-th league is above the lower bound of the confidence interval and when the latter is higher than zero, delivers positive mean returns for all the three Mediterranean leagues which were found to be inefficient. By blindly wagering on odds inferior to 1.67 and 2.08, the authors reached mean returns of 2.09% and 2.71% for Italian Serie A and Greek Super League respectively, while the Spanish Liga turned back a positive mean return of 2.12% by selecting odds included in the range from 1.09 to 3.12. Hence, this strategy shows that abnormal positive returns are achievable, demonstrating that biases can be exploited in order to profit from betting, deleting any doubt about the inefficiency of these markets.

Another contribution to the discussion on betting market efficiency is given by the Elaad et al. (2019) paper, which studies the prices, i.e. the odds, posted by fifty-one online bookmakers, for the final results in over 16,000 football matches belonging to the four major English professional leagues. The matches were collected from 2010/11 to 2017/18 seasons. The authors made tests to verify the version of semi-strong efficiency, i.e. whether there is any significant possibility to achieve positive abnormal returns just by exploiting some predictable bias in the quotes. The first issue they verified is whether the odds systematically over or under-predict any particular outcome. A psychological research by Na et al. (2018) have demonstrated that individuals are inclined to under-predict significantly the draw outcome, which is consistent with the concept of 'black and white' thinking. The second point is whether there is evidence of the favourite-long shot bias. The work by Elaad et al. extends the approach of the model by Angelini and De Angelis (2019) by modelling the heterogeneity between bookmakers in their profit margins and considering whether odds also imply too few drawn matches. The model is the following:

$$\varepsilon_{i,j} = \beta_h h_{ij} + \beta_\alpha a_{ij} + \beta_z z_{ij} + \varphi_{t(ij)} + \alpha_j + v_{ij}, \qquad E\left[v_{ij} \middle| h_{ij}, a_{ij}, z_{ij}, \varphi_{t}, \alpha_j\right] = 0$$

where h_{ij} and a_{ij} are dummy variables indicating whether the odds are for a home or away win. α_j gives bookmaker fixed effects, in order to address heterogeneity in the bookmaker margins and how this might be correlated with their tendency to reflect a favourite-longshot bias in their odds. The authors have attributed any general change of these margins to the estimation of the season fixed effects in $\varphi_{t(ij)}$, where t(ij) indicates the season when a match took place. The residual term v_{ij} contains the remaining heterogeneity.

They have estimated the model using weighted least squares, thus increasing efficiency and partially attributing heteroskedasticity to the forecast errors. The null hypothesis, that the betting market is efficient, must satisfy the following sufficient condition, H_0 : $\beta_h = \beta_a = \beta_z = 0$, otherwise it would imply that home, away or draw results are under or over-predicted by bookmaker odds, respectively. They centred their study on the English Premier League market. If $\hat{\beta}_h$ and $\hat{\beta}_a$ have negative signs mean that home and away wins are over-predicted. In fact, relative to the draw, the odds implied forecast error is on average 1.7% points less for an away win, and bookmakers find lots of difficulties at predicting the draw, as previously suggested by Pope and Peel (1989). Unexpectedly, there is a negative favourite-longshot bias, generated by the fact that bookmakers tend to underestimate the longshot rather than the favourite team. In conclusion, the authors could not find any evidence against the null hypothesis, at standard levels of significance, hence they concluded that the Premier League betting market has been efficient over the past eight seasons.

Also in the other leagues analysed (second, third and fourth English leagues) over this period, there is no significant evidence to reject the Efficient Market Hypothesis in this context, despite apparently large positive coefficient estimates on the favourite-longshot effects in the second and third league. Further than that, the commission rates significantly decreased by 0.3-0.5 points by the end of the last season considered, probably because betting markets are becoming more and more competitive, primarily because the technologic evolution is progressively moving markets online, thus lowering transaction costs. Looking at different bookies, Elaad et al. found that for most of them a linear combination of the odds from competitors in the market would significantly predict their own odds-implied forecast errors. In this way, they could declare the inefficiency for eight of the thirteen bookmakers at the 0.1% level, eight at the 1% level, and twelve at the 5% level of significance. Consequently, they could conclude that the odds set by fixed-odds bookmakers for the main English football championships matches are not capable of incorporating readily available and timely information which must be used to forecast the event outcomes. In other words, in their opinion (p.6): "the increases in market efficiency should lead every bookmaker to incorporate each other's odds and it should be statistically significant, but in magnitude are almost certainly tiny". This may indicate that to detect the best odds on an outcome, the bettors must face some small economic costs, for example due to the time taken to switch funds and attention between online accounts.

Further than the weak efficiency test, Kuypers (2000), using a similar data set that includes also the Scottish championships, presented also a semi-strong efficiency test, by attempting to achieve abnormal returns using an ordered probit model and publicly available information. The publicly available information variables are some simple statistics such as teams' average points per game over the season, teams' cumulative points over the season, teams' league position etc. For each of these ones the model considers the difference between the two teams, which just refers to the value of the variable for the home team minus the value of the variable for the away team. As the variables described above increase, the probability of a home win would increase with them, while a reduction of their value would raise the probability of an away win. As the variable approaches 0 the probability of a draw would be expected to be greatest. A betting rule was then developed: place £1 on the outcome of a particular match whenever the probability ratio of the probability predicted by the model over the implied probability from odds is bigger than a certain value. Imposing a threshold of 1.1 or above, positive returns can be made. This implies that the betting market for football is not semi-strong efficient.

3.2 The efficiency in the exchange betting market

Empirical research on the prediction accuracy of bookmaker odds is well established in the literature. Consequently, the rising of the exchange platforms has shed a light on the efficiency of this market, and thus to its forecasting accuracy. Further than the evaluation of its predictive power in itself, there has been an increasing interest in comparing the accuracy of the fixed-odds and the exchange market, and the consequences of this new business model in the overall market efficiency.

About this topic, by Smith et al. (2006) have declared that betting exchanges have improved significantly the markets efficiency by lowering transaction costs for consumers. This is accentuated in the horse races sector, since regular bookmakers are used to impose a margin around 20%, that is very high, because the possible outcomes are numerous, thus they try to protect themselves by increasing their margin. Instead, Betfair exchange charges always a fixed percentage on the net cash-inflow, that is usually around 5%. The authors confirm their hypothesis working on data on almost 800 UK horse racing during 2002. These values were taken first from traditional bookmakers, from which they calculated the mean prices for each runner in each race, highlighting the most favourable price for each horse (the outlier), as this is an important competitive benchmark against which betting exchange prices are compared by bettors. After that, they matched the bookmaker data with the corresponding betting exchange prices, collected at the same time each day. In order to quantify the information intensity, they divided the information classes in four groups, according to betting volume. They chose this criterion because it is highly correlated with other relevant qualitative criteria such as racecourse grade, information on runners, media coverage and prize money. They classify the groups based on these categories: races with low, moderate, higher than average and very high betting volume.

The results suggest that the established favourite–longshot bias is evidently lower in person-to-person (exchange) betting than in traditional betting markets. Further than that, in both exchange and fixed-odds betting markets, the level of bias is lower the greater the amount of public information that is available to traders. Additional empirical support for an information-based model is found by employing an alternative methodology which enables to arbitrate between information-based and risk preference models of the favourite–longshot bias in relation to the data. The risk model is specified as:

$$\log(p_i) = \alpha + \beta \log(\pi_i)$$

where p_i stands for the subjective probability of horse i winning, and π_i is the corresponding objective probability. The information model is presented as follow:

$$p_i = \alpha + \beta\pi_I$$

In both cases β measures the favourite–longshot bias, with 0< β<1 indicating over-betting of longshots relative to runners at short odds, and β >1 indicating an opposite favourite–longshot bias. The information model appears to fit the data much more closely than the risk alternative and it explains the favourite–longshot bias better.

The paper by Franck et al. (2010) is the first that compares the forecast accuracy of the bookmaker market with that of a betting exchange platform. Using a dataset covering all football matches played in the major leagues of the "Big Five" European countries (England, France, Germany, Italy, Spain) from season 2004/05 to season 2006/07, with 5478 games in total, they contrasted the forecasting power of eight different traditional bookmakers' odds with the prediction accuracy of the corresponding odds traded at Betfair, the most popular exchange market. The correlations between the probabilities incorporated in the Betfair odds and a random bookmaker are 0.917 for draw bets, 0.978 for visitor win bets and 0.981 for home win bets. Thus, the odds traded on the betting exchange must be similar to the fixed-odds. The differences appear to be unsystematically distributed, and a closer inspection reveals that, for home and away win bets, the bookmaker probabilities seem to be higher than the Betfair probabilities in the area of low probability outcomes, and vice versa. The authors estimated the following model to explain the realized outcome (win or loss) of a certain bet $Y_{ei} \in \{0,1\}$ for a given match i using the implicit probabilities of the different markets $P_{ei\,j}$:

$$Y_{ei} = \mathrm{G}(\alpha_{ej} + \beta_{ej}P_{ei\,j} + \varepsilon_{ei\,j})$$

For each event e (home wins, draws and away wins) and every market j (eight bookmakers and Betfair), the coefficients $\widehat{\beta_{ej}}$ are estimated using a probit model. The probit model has the task of relating the probability of occurrence of discrete events to some set of explanatory variables, where $G(\cdot) = \varphi(\cdot)$ is the standard normal cumulative distribution. The prediction accuracy is examined using various goodness-of-fit measures. The model that employs as explanatory variable the probabilities retrieved from Betfair odds has better goodness-of-fit scores than the regressions using either the probabilities implied from the posted fixed-odds taking each bookmaker individually or taking the aggregate average. Therefore, the authors concluded that the betting exchange platforms are better predictors of the final outcomes of a football game than the traditional bookmakers. Moreover, it is worth to notice that the prediction accuracy of draws is remarkably worse than that for the other two possible outcome. The marginal effects of the probability incorporated in the odds are substantially greater than one for all bookmakers and events, and the ones belonging to the exchange probabilities are closer to unity than any regular bookmaker. Hence, the actual chances of winning increase disproportionately with the ones implied in regular bookmaker odds. This fact is evidence that a form longshot bias lies in this market. Moreover, since the marginal effects are higher in Betfair, the P2P market is less affected by this bias, and thus more accurate and efficient, as stated by Smith et al. (2006).

To conclude, the sports betting market shares an analogous nature with more complicated financial markets, that allows tests of economic theory such as the Efficient Market Hypothesis, whose validity is traditionally hard to test. The advantage of the betting market is that there is very little space for information lack which, conversely, affects financial market efficiency. Researchers studied mainly in the European football betting market, using both statistical tests, in the form of regressions, and economic tests, in the form of betting strategies, to discern the efficiency of the market. Over time, these results have been mixed, depending on the statistical tools employed, the market under analysis and the extent of data. Modern software has lowered the barriers to entry in the sports book market, allowing more sports books to operate, thus increasing the competition, and lowering margins. The betting market on aggregate seems to be efficient, but, on the other hand, specific inefficiencies are demonstrated. The principal causes are the irrationality and biases of some players ‘behaviour and the bookies price setting mechanism, which aim more at maximizing their profit than offer efficient odds. Moreover, the exchange platforms seem to better predict future outcome probability, thus they are more efficient. These findings are the foundation on which we can build, also using modern machine learning techniques, profitable betting strategies that aim to exploit market inefficiency. Next chapter will present the statistical arbitrage strategies currently available in the betting market.

4 Arbitrage opportunities

This chapter will present the main arbitrage strategies proposed in the scientific literature. In fact, once several studies have evidenced the presence of different types of biases in the betting market, several research studies have been conducted through the last decade about this topic. Their purpose is to build profitable arbitrage strategies by exploiting these inefficiencies or by modelling a machine learning forecast approach capable of beating the bookies regularly. Before starting to demonstrate the effectiveness of these investments, we must make the idea of arbitrage as clearer as possible. The concept of arbitrage is fundamental in financial literature and has been used in classical analysis of market efficiency, whereby arbitrage opportunities are quickly exploited by investors. However, pure arbitrage opportunities are unlikely to exist in a real trading environment. An arbitrage trade typically involves some kind of risks. In the specific case where these risks are statistically assessed, it is appropriate to use the term statistical arbitrage (SA). SA has been widely investigated in literature focusing on two main approaches, a theoretical one that focus definitions, and a practical one, that aim to develop profitable investment strategies. In particular, several studies have introduced other definitions extending the concept of arbitrage through statistics but without tempting to convert it in practice. On the other hand, research on statistically arbitrage strategies have addressed their effort on building models and investment opportunities with little or no discussion on definitions and theoretical framework.

Thus, the definition of statistical arbitrage remained ambiguous, until the paper by Lazzarino et al. (2018), who pooled all the existing literature about this argument, collecting all the definition of statistical arbitrage and analysing each aspect of them (lexical, operational and conceptual). At the end, they came up with a definition which is able to summarize all these aspects (p.901): "a statistical arbitrage strategy is a relative value strategy with a positive expected excess return and an acceptably small potential loss". The part of the definition about the expected positive excess return incorporates two features. The first one is given by the fact that the strategy focuses on the expected return. This differs from the definition of arbitrage where the strategy has no admissible possible negative outcomes. Losses are allowed in the definition of SA. The second one is given by the excess return. This reflects the fact that every arbitrageur begins a strategy that involves some risk only if the expected returns are higher than the risk free rate, whenever an initial investment is required. The last requirement is given by the acceptably small potential loss, that is the element that distinguishes a statistical arbitrage from a simple investment strategy. To be called arbitrage, a strategy must have a constrained loss profile. Considering the sports betting market, there are several strategies that correspond to this definition. In particular, the most popular are: matched betting, sure betting and value betting. The first two can be considered closer to a pure arbitrage investment plan, since their risk can be approximately to zero. Value betting, instead, comprehends a set of strategies that incorporate a risk aspect, but in the medium-long term they have demonstrated to be profitable. Next paragraphs will explain in detail each one of these arbitrage strategies, showing the mechanism, profitability and flaws.

4.1 Matched betting

The first strategy to consider is the matched betting one, that is one of the simplest. As the name suggests, it consists of matching two bets on two opposite outcomes. Banks (2013) explains that profits are generated from bonuses and free bets that online bookmakers offer to attract new and maintain existing customers. A fundamental condition to apply it is that it is necessary to possess an account at an online bookmaker that offers these promotions. To efficiently run this strategy, it is recommended to another account at a betting exchange, and not only at another bookmaker. By placing a bet on a particular outcome at the bookmaker offering the incentive and simultaneously placing corresponding bets against these outcomes in another website, a profit will be generated regardless of the outcome.

Typically, bookmakers require customers to place a bet using their own money before a free bet is awarded. For this reason, a bet is placed at the bookmaker on a particular outcome. Then, a second bet against this particular outcome is placed (usually) at Betfair. The aim oh this second bet is to offset any loss if the outcome of the first bet does not occur, resulting in a small loss whatever will be the final outcome, but without risking to lose the money wagered. After having placed a bet with real money, the free bonus bet is now unlocked, and the same process of bet and counter-bet is followed. Irrespective of the outcome, a profit is generated because the bet was free.

All bookmakers offer different discounts with their terms and conditions, but the same strategy can always be implemented. At the end, the final goal is to make the bonus money withdrawable. Typically, the counter-bet is placed in the exchange platforms, because it is more convenient than other bookmakers and, as in the case where we wager the opposite outcomes in the same website, it does not alarm the bookie, who can detect our not risky activity and choose to suspend our account. There are several companies that offers, behind the payment of a relatively small amount, a service that compares bookie’s odds with exchange’s odds for every match available in the bet schedule, in order to find the best combination of bet and counter-bet and spend as less money as possible for this procedure, which is perfectly legal. Usually, back an event on a fixed-odd market and lay the same event on Betfair is costly, most frequently around 2%-5% of the first bet. Hence, exploiting the best combination is important, because it allows to make the bonus almost entirely withdrawable.

Two examples can clarify the mechanism. The first one refers to a typical welcome bonus. In Italy, Bet365.it offers to its new client a welcome bonus that equals the amount of the first deposit on the website, until a maximum of €100. The conditions establish that the value of the bonus must be wagered at a minimum odds of 1.5 and the money won by betting with it are locked, they will stay in the bonus balance, until a value of six times the bonus is wagered. After that, the entire bonus balance will become withdrawable. Thus, for us it will be more convenient to win the counter-bet at the first chance, in such a way that the money will be immediately withdrawable in the other website, otherwise we have to keep betting until we generate a total amount of €600 betted on Bet365, that would lead to decrease our profit, since we must spend more money by keeping betting.

For example, supposing that we are a new client and that we deposit €100, the bookmakers will immediately accredit €100 of extra bonus in our account. Looking at the bet schedule, there is a good opportunity to bet on a draw of a Spanish Primera Division between Mallorca and Getafe. Bet365 offers this outcome at 3.10, and in the Betfair exchange platform the same odd is available. Now, we bet €200 on the fixed-odd market, and we lay the draw on a peer-to-peer bet, specifically laying €203.28 and consequently risking €426.89[14]. At the end of the match, if either Mallorca or Getafe win, we will have a net cash-inflow on Betfair of €193.12 (supposing a commission rate of 5%). Subtracting our €100 lost on Bet365, we will have a €93,12 of net profit without risk. In case a draw occurs, we must keep on betting in this way until we zero the bonus balance on Bet365 or until we bet a total amount of €600. In this latter case, the profit will be withdrawable directly from the bookie. However, it will be lower, because we needed to bet more times, but still significant. In these cases, the arbitrageur must be aware that, in case things do not go in the most convenient way, a certain level of liquidity is necessary before being able to get the money.

Another frequent type of promotion is the reimbursed bet. For instance, we suppose that a bookmaker offers us a bet on which we could receive a reimbursement of 50% of the amount wagered, until €30, if the bet results lost. In a Champions League match between Atletico Madrid and Milan played in Spain, the away win is given at 5.5 in the bookie and the lay odds

[14] Remember that a lay bet works in a mirror way. The liability is: gross outcome * (odd - 1).

is 5.9. Again, by betting €60 in the fixed-odd market, and laying €51.28 with a liability of € 251.27, a sure profit of €18.72 can be insured whatever will be the final result. If Milan wins, the net cash-inflow will be: (€60 * 5.5) - €60 - €251.27 = €18.73; if Milan does not win, the computation will be: (€51.28 – 5%) – €60 + €30 = €18,72 since, in this case, the bookmaker will reimburse us the 50% of €60, thus €30.

This type of arbitrage does not have possible losses embedded in the strategy, unless the player mistakenly places the bets, and has good returns. But, on the other hand, it is subjected to the willingness of the bookies to propose promotions. In fact, they can unilaterally limit your account and ban you from every discount, but, in reality, it is quite rare, since all the providers of this market are interested in increasing the number of clients. Further, its simple and safe structure has led to its diffusion across all betting markets.

4.2 Sure betting

The second arbitrage strategy is the sure bet. It consists in placing multiple bets that cover all the possible outcomes on a match, resulting in a positive sure profit. The stakes placed on each side must be chosen such that the return on the combined bet does not depend on the actual outcome of the match. Since all bookmakers impose their margin on their odds, it is impossible to reach a positive return by simply betting on all the possible outcomes at the same time. Nevertheless, in some cases by taking the maximum quote available for each outcome even negative profit margins might be achieved in the market on aggregate. Similar to finance, where arbitrage refers to the practice of making a risk-free profit by taking advantage of the simultaneous purchase and sale of the same, or essentially similar, asset in two different markets for advantageously different prices, sure bet arbitrage refers to betting on sports events so that a sure profit is guaranteed by combining odds offered by different bookmakers. Formally, a bet is an arbitrage bet if $\sum_{j=1}^{n} \frac{1}{\delta_j} < 1$ where n denotes the number of outcomes and δ_j is the odds quoted for the j_{th} outcome. The occurrence of this phenomenon depends on two factors: the divergence of odds between different bookmakers or the different behaviour of their respective clients, and the commissions applied by these bookmakers. The larger the divergence of odds and the lower the commissions, the more arbitrage opportunities there will be. For example, if a bookmaker offers an odd of 5.25 for an over 5.5 goals (six or more goals) in a football match, and another bookmaker offers, for the same match, an under 5.5 goals odd of 1.25, a sure profit of 5% can be reached. In fact, by betting €100 on over 5.5 and €420 in the second bookmaker we will earn €5 without risk[15]. A simple way to calculate the money to invest on each side is the following: initially, we have to choose the amount we want to invest. Supposing that we choose €100, to calculate how much money we need to investment in the j_{th}outcome in order to reach a sure profit whatever is the final result, we can just look at this simple formula: $\left(\frac{€100}{\sum_{j=1}^{n}\frac{1}{\delta_j}}\right)/\delta_j$.

In this case the strategy has been pursued on an intra-market base, thus using two fixed-odds markets. Although arbitrage betting seems to provide risk-free profits at first sight, it includes some practical challenges that must be carefully considered. For instance, it could be that after having placed a bet on a website, the other one has changed its odds in the meanwhile, or that it refuses or limits the amount you can bet on that outcome.

Frank et al. (2012) wrote the first paper that analyses the inter-market arbitrage in the betting industry. This strategy consists in matching bets at the bookmaker market and at the exchange market. The low transaction and trading costs of the inter-market strategy permits to define it as a better arbitrage opportunities seeker than the intra-market strategy (betting only on fixed-odds bookies). The different business model allows the peer-to-peer platforms to apply smaller commissions, as they do not sustain the counterparty risk. Moreover, the exchange mechanism is easier and more immediate. For example, to counter bet on a specific outcome it is sufficient to lay it, and it does not require to back all the other possible results. Another advantage of this arbitrage strategy is that it is unquestionably easier to find greater differences in prices looking at different markets, instead of just focusing the attention on a specific one. Thus, rather than casual random fluctuations, the distance between bookies' abilities to incorporate new relevant information in their odds is the origin of this type of arbitrage opportunities. The authors agree with this conclusion (p.321): "rather than the bet exchange market, it is the bookmaker who gives rise to the arbitrage opportunities".

Empirical studies presented in the previous chapter have shown that bet exchange odds perform better than bookmaker odds in predicting the outcome of sporting events. The paper by Frank et al. (2012) confirms these findings. The Inter-market arbitrage strategy works by implementing two main steps: by a mispriced odds from a traditional bookie and sell them at a lower odds on Betfair or a similar company. To confirm the theory, these scholars selected the quotes offered by ten bookies and matched them with the corresponding odds traded at Betfair.

[15] If the match finishes under 5.5, the cash-flows would be +€425 - €420 = €5.
In the opposite case -€100 + €105 = €5

They analysed precisely 11,933 soccer matches played in the “Big five” European leagues, the English, French, Italian, German and Spanish first league, among a period of time of seven seasons (from 2004/05 to 2010/11). Contrary to the fixed-odds market, in P2P platforms odds are traded continuously in a “double-auction process”, hence quotes are constantly moving. This aspect was an issue, since to give a sense to this study, the authors had to make sure to record in the dataset odds valid simultaneously in both types of market. The data from Betfair were extracted dropping the beginning and the last quote negotiated. However, odds in this type of market are “cross-sectional” and not “longitudinal”, thus they do not have information about how the odds have moved in the trading period. Relying on a standard deviation of 0.146, they picked up the most traded quote in terms of volume, because it was logically the most probable odds a player could have found available in the market.

Assuming a Betfair commission of 5%, in the intra-market analysis, out of 12,782 matches, they found 102 arbitrage opportunities with an average return of 0.9%, that is one arbitrage opportunity every 125 matches, or 0.8%. Alternatively, the inter-market strategy reached an average positive return of 1.7% and it was worth to bet on 5.0% of the matches considered. At the end, the inter-market arbitrage strategy has demonstrated to offer a higher number of arbitrage opportunities, as well as a greater return. In any case, this study underlines that this type of arbitrage opportunities is far from being rare in the football betting market, although the data selection criteria is rather prudent. In fact, by choosing football matches, they considered only intensively traded bets. Secondary sports and championships are less followed, hence for them there are less publicly available information. This aspect accentuates the disagreement among bookmakers and bet exchange platforms, implying more arbitrage opportunities for bettors. Secondly, the odds are compared at just one exact moment among the entire negotiations period, for this reason the number of valuable opportunities could substantially increase by monitoring the odds continuously. The same reasoning could be applied if the strategy would have considered also other type of bets, such as bets on the correct score, sending offs, the final rank order, as well as different sports.

4.3 Value betting

The dream of every bettor is to find a strategy that systematically beats the bookmakers, but, since they implicitly charge a commission in their odds, it is not an easy task. Moreover, the majority of players do not bet following statistical or mathematical rules, they simply choose the team they think is going to win, often regardless of the offered quote. In this way, they are only playing the bookmakers' game. Above all in recent years, with the implementation of machine learning and artificial intelligence techniques, experts and researchers have tried to create several betting rules capable of achieving a safe profit in the medium-long term. What each approach has in common is a shared goal of finding "value" in the odds, where the true chance of an outcome is greater than the one estimated by the bookmaker. Thus, in general, a value bet is placed whenever the punter's notion of the objective probability exceeds the probability implied by the odds. A value bettor does not have to think about the clubs that are playing, their rank position, his expectation etc, but he only has to focus on probabilities.

In Buchdal's (2003) opinion, the important question a punter should ask to himself is whether the true chance of a team to win is greater than the one which the bookmaker has unfairly (in his mind), but potentially mistakenly (in the punter's mind), estimated it to be. In other words, is the bookmaker's odds greater than that which the punter considers to be the fair price? If it is the case, he has detected a value bet. The author views successful betting as a practice of understanding and managing probabilities and describes value betting as the only way to overcome bookmakers' odds, providing an accessible measure of a bettor's expectation to make a profit. Neither statistical probability can often be applied to a sports event because each such event is unique in a set of factors that are stochastic by their very nature, thus the exact probabilities of outcomes are impossible to calculate. This is valid for every type of sport competition.

Stekler et al. (2010) have presented all the issues regarding sport prediction across horse racing and several team sports, such as basketball, American football, soccer etc. In their opinion, there is no evidence that, in general, either statistical systems or experts consistently outperform the market. Even if the player's estimated probability comes out to be different from the bookies' one, it could be estimated wrongly. Only in the medium long term he can know if his model is a good outcome predictor or not, because this type of investment is volatile for its nature and it could lead to losses in the short run. For this reason, different forecasting tools have been proposed in the literature. The "value" has been found through different strategies. This chapter will present them singularly, focusing mainly on the machine learning approach.

4.3.1 Value found through the exchange market

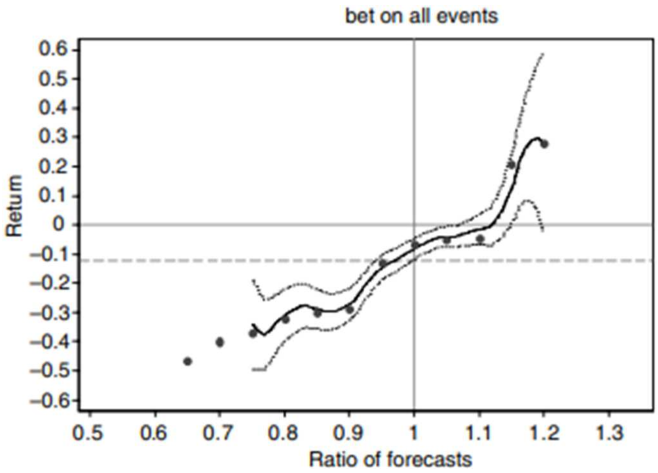
Section 3.5 has explained that the exchange platforms are on average more efficient than the traditional fixed-odds market, hence they are also a better predictor toll. Smith et al (2012) attribute the greater relative efficiency of the exchanges in reflecting objective outcome probabilities to differences in the nature of traders and trading activities. The peer-to-peer bets offer for these players opportunities to trade which are not available in fixed-odds bookmaker markets. For example, skilled traders, insiders, and bettors seeking hedging opportunities are all more attracted by the exchanges, because they can lay odds and cash-out whenever they want. In these circumstances we might expect the proportion of turnover attributable to casual bettors to be lower in the exchanges than in bookmaker markets, with a consequent tendency for exchange odds to reflect more closely objective probabilities, resulting in a higher level of efficiency.

Beside the statistical analysis, Frank et al. (2010) created a betting strategy, based on their findings, which aims to test the economical relevance of the fact that the betting exchange market is a better estimator of future outcomes probabilities compared to the fixed-odds market. Based on this claim, it should be possible to build a betting strategy able to generate abnormal returns by exploiting the forecasting differences between the two markets. The betting strategy consists of betting on an outcome whenever the probability implied in the exchange odds is higher enough than the one incorporated in the fixed odds. Alternatively, a bet is placed if the average bookmakers' quote is greater than the one available at the most popular exchange platform in the web, thus Betfair. The outcome considered is always the final result of a football match (1, X, 2) and the bookies analysed were eight in total. By applying the explained strategy, the authors yielded above-average returns in all cases, in the sense that the average returns are less negative or positive than the returns yielded by wagering on all the outcome of a specific event. The best outcome turned out to be the visitors win, where they could achieve stable positive returns.

In the second part of the article, the scholars have examined in depth their betting rule, to have a better understanding of their first findings. In particular, they divided all the value bets according to the level of disagreement between the two types of market. At this point, they calculated the observed average returns for each group. In general, they discovered that as wider is the distance between the two market's odds, as greater is the expected return on a bet closed with a traditional bookmaker. Thus, to confirm the previous results, they expected a direct proportionality between the observed return and a value R, intended as fraction where the numerator is the exchange probability and the numerator is the average bookmakers probability.

To investigate this relationship, they ordered all bets according to their ratio R. The criteria used to run this study are that every bin should contain at least 50 observations, and the categories are separated by a bandwidth of 5%. Figure 9 shows a locally weighted polynomial regression with 95% confidence intervals (dotted lines), with this result:

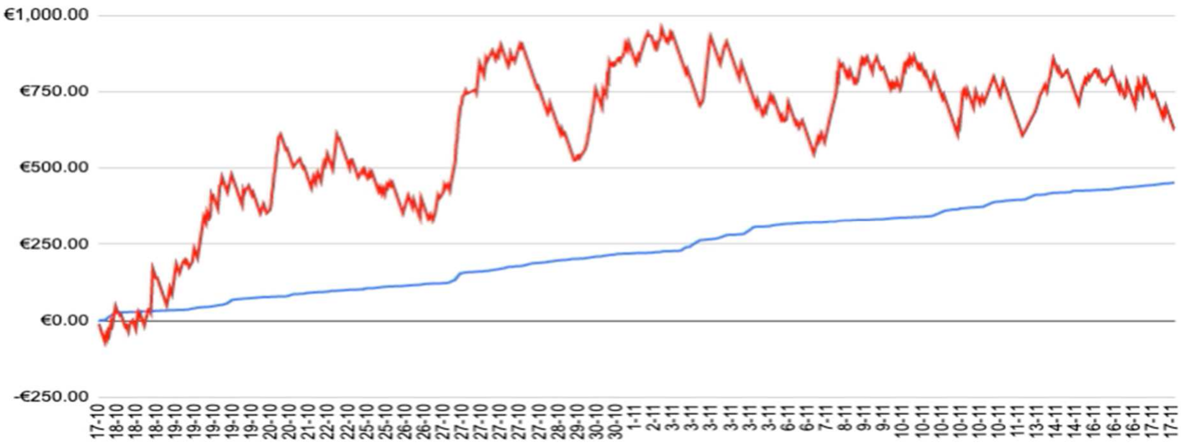
Figure 9 Regression result



It can be noticed that when the Ratio (R) = 1, the observed mean returns are approximately at the level of normal returns, represented by the dashed horizontal line, and that they increase as the R increases. The authors made also an analysis considering each outcome singularly, noticing that the relationship is steeper for visitors than for home win bets. The ratio R is positively correlated also with the average returns for the different groups of R, the dots in Figure 9, and the local polynomial smoother, denoted by the solid line, in all cases. Moreover, for some positive values of R, Figure 9 shows that the betting strategy yields positive returns. For instance, betting against the random bookmaker on all events in the top 5%-quantile of R yields an average return of +10% out of 821 bets, while betting in the top 10%-quantile (547 bets in total) yields +3%, on all home wins and+7% on away wins (547 bets in total). Thanks to all this evidence, it is reasonable to affirm that the exchange odds are good instruments for estimating the real probability of an outcome, and therefore for detecting mispriced quotes in the fixed odds betting market. Thus, the Betfair traded odds enable the bettor to find "value" just by comparing them with the ones posted by bookies.

It is possible to find online websites that provide software capable of comparing instant by instant odds of the same event by all the bookmakers available and the Betfair platform. Ninjabet.com is the widest community across Europe that provides this type of services for its clients. It guides its bettors through the implementation of the matched betting strategy, the sure betting strategy and the value betting one. The subscription prices are respectively growing, since they are as high as the quantity of profit is bigger. Regarding the value betting, they find the value with this exchange approach. On the website[16], a demonstrative one-month simulation is available. Its results are summarized in Figure 10. With a number of 1012 bets, each one of €10, they realized a return of 6.30% out of an expected return of 4.47%, where the expected return is computed as: $1 - \frac{Exchange\ odd}{Bookmaker\ odd}$. Hence, supposing a fixed-odd of 6.5 and an exchange lay odd of 6.25, the expected return would be 4%.

Figure 10 Value betting results



The blue line represents the expected return, while the red line is the realized return. As we expected, the red line is pretty jagged, that means that the volatility is high. Anyway, these results confirm that this way of searching "value" is able to guarantee a profit, even if one month is not a big-time interval.

4.3.2 Value found through the bookmakers' disagreement

The second tool that can be used to find the "value" is the direct analysis of the fixed-odds offered by all the bookmakers active in the market, without considering the exchange platform. To correctly predict the outcome of sports events, bookmakers usually employ several professionals in the sector of data analysis to scan decades of sports data and in order to provide accurate forecasted probabilities. Although bookmakers' way of running their business seems to suggest that it is impossible to elaborate a strategy capable of systematically generating profits, as they have complete power in setting prices and commissions, some research aim to demonstrate the opposite.

Kaunitz et al. (2017) designed a strategy to beat football bookmakers using their own odds. The basic idea is to compete against bookies not trying to be better than them in forecasting sports

[16] https://www.ninjabet.it/istruzioni-strumento-valuebet

events probabilities, but beat them using the probability incorporated in their posted odds, to find bets with mispriced odds. The betting system uses a different approach from classical betting strategies, in the sense that instead of trying to create a model to outdo the bookmakers' forecasting abilities, they used the average posted market odds as a proxy of the true probability of a sports event, relying on a wisdom of crowd effect. With these proxies they searched for games with quotes offered above the estimated true probability, thus looking for mispricing. From chapter 3.1.1 about the bookmakers setting price mechanism, we know that sometimes bookmakers do not find convenient to offer odds that reflect their estimated probability, mostly to attract as many gamblers as possible or to balance their book, in order to avoid getting overly exposed on one side of the market. This implies that bookmakers might offer inefficient odds. This is the key factor that they exploited in this strategy.

To validate their idea, they measured the predictive power of bookmakers' models through a historical analysis of football game results. They employed the historical closing odds from 32 bookmakers of football games from January 2005 to June 2015 (479,440 games in total) from online sports companies operating in Internet. As first step, to quantify the accuracy of bookmakers' outcomes probabilities in an aggregate sense they calculated the probability implicit in the market odds as a whole, calling it consensus probability as follows: $p_{cons} = {}^{1}/_{mean(\Omega\,)}$ where Ω is a set containing the odds across bookmakers for a given event and a given game result (home team win, draw, visitor team win). They put as a condition a minimum of 3 odds available for calculating the consensus probability. Then, according to it, they divided the data in 80 bins from 0 to 1 and in steps of 0.0125. Within each group they computed the consensus probabilities across games using closing odds, the quotes provided by bookmakers right before the start of a match, and the mean accuracy in the prediction of the football game result, that is the proportion of games ending in home team victory, draw or visitor team victory for that bin. Then, these data were used to run a preliminary linear regression analysis that evidenced a strong correlation between the bookmakers' consensus probability and the final results of the match for home victory (R^2= 0.999), draw (R^2= 0.995) and away victory (R^2= 0.998). This effect can be interpreted as a wisdom of the crowd effect, where the aggregate predictions of a group of odds produce more accurate probabilities than those of each bookie taken singularly.

In light of these results, being aware that fixed-odds markets can be considered a good tool to estimate the probabilities of football games, they decided to build the betting strategy. They based their investment plan on the expected payoff of each bet. Assuming Π to be the payoff of the bet (a random variable), p_{real} the actual underlying real probability that the outcome comes true in reality, and ω the pay-out paid by the bookmaker in case that the outcome occurs, the expected payoff of betting €1 is:

$$E(\pi) = p_{real} * (\omega - 1) + (1 - p_{real}) * (-1) = p_{real} * \omega - 1$$

In a preliminary data analysis, they discovered that:

$$p_{real} \cong p_{cons} - \alpha$$

Where p_{cons} is the consensus probability as calculated above and α is an adjustment term that incorporate the intercept estimated in a regression analysis, where the dependent variable was the probability of a classic "1, X, 2" outcome. The α assumed the values 0.034, 0.057 and 0.037 for home victory, draw and away victory, respectively. Then, rearranging:

$$E(\pi) = (p_{cons} - \alpha) * \omega - 1$$

Under these conditions, it is reasonable to place a bet whenever the expected payoff is greater than 0, that is:

$$\omega > \frac{1}{(p_{cons} - \alpha)}$$

Following this way of thinking, they determined the betting strategy, and decided to place a bet whenever the maximum odds offered for a given outcome satisfied the following condition:

$$max(\omega) > \frac{1}{(p_{cons} - 0.05)}$$

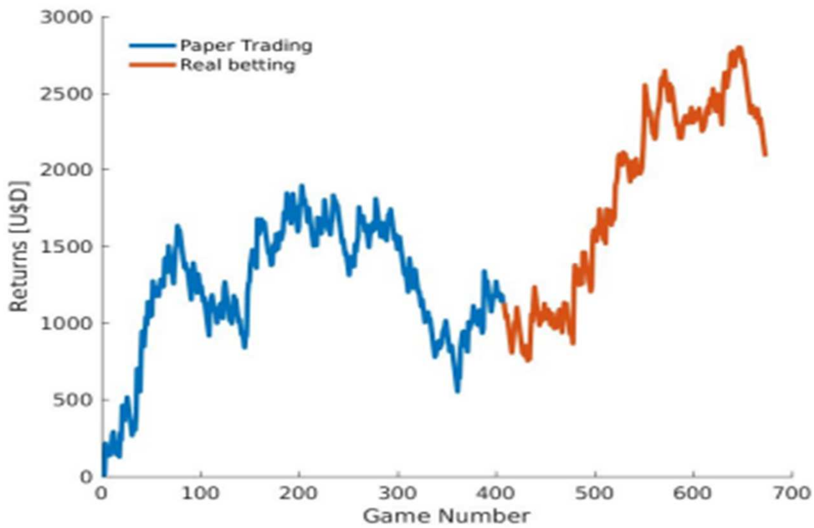
As the α parameter increases, the expected value of each bet increases with it, while the number of games available for betting decreases. This happens because the condition becomes more stringent and narrower, but also safer. In other words, the α parameter regulates the "quality" of the bets placed, in the sense that a higher α implies a wider discard between the consensus and the maximum odds. To select an appropriate value for the α parameter, they tried several simulated strategies by varying the value of α from 0.01 to 0.1. An α of 0.05 turned out to be the best value, since it produced the optimal payoff with the largest number of games.

In summary, the betting strategy consists of placing bets whenever the odds offered by some bookmakers deviate from the average market odds and when this difference is large enough to guarantee a positive expected payoff. It is worth to underline the main advantage of this strategy, that is its simplicity: to identify maximum odds capable of fulfilling the betting condition did not require a sophisticated prediction model. To put the theory in practice, they used a virtual machine to run a program that looked for the odds of every football game from 5 hours before the beginning of a match. Once it collected the odds from the 32 bookmakers considered, at any time within this time interval it calculated, for every match, whether the maximum odds satisfied the strategy's threshold for placing a value bet. In this way, they could draw up the list of bets recommended by the system and place a bet of a fixed amount with the bookmaker individuated. They began the analysis by applying the betting strategy to the closing odds of each game, that are the quotes posted by bookmakers at the onset of a match. They simulated placing bets when the betting condition was satisfied. With this approach, the betting strategy achieved an accuracy of 44.4% and yielded a 3.5% return over the analysis period.

Encouraged by the positive results of the initial analysis, the authors decided to conduct a more realistic simulation in which they placed bets at in the time period from 1 to 5 hours before the onset of each match. This simulation is more realistic because in real life it almost impossible to use the bets available one instant before the beginning of a match. Since to collect data containing the time series of odds movements before the beginning of each match is not an easy task, they wrote a new set of scripts to gather information in real time as they became available online. In total, they obtained data for 31,074 matches, from the 1st of September 2015 to the 29th of February 2016. Then, the started to apply the strategy according to the same rules determined before. Under these conditions, the strategy selected odds with a mean value of 2.32 (STD=0.99.), with an accuracy of 47.6% and yielding a 9.9% return across 6,994 bets.

Once they determined that the betting strategy was profitable, they decided to test it under more realistic betting conditions. To this purpose, they implemented a technique called "paper trading", a virtual trading process in which gamblers can "practice" placing virtual bets, without risking real money. Paper trading allowed to empirically check whether the odds were indeed available at the bookmakers at the time of placing a bet. This passage is necessary, because there was often a time delay between the moment when bookmakers published their quotes and the time it took for the scripts to display that information on the dashboard. In fact, almost one third of the odds that were showed on the dashboard had already been changed, but, nevertheless, the strategy remained profitable. Therefore, the authors decided to keep betting, and after three months of paper trading they reached an accuracy of 44.4% and a return of 5.5% in 407 bets. At this point they tried to execute the strategy with real money maintaining all the procedures unvaried. In 5 months, they obtained an accuracy of 47% across 265 bets, with an 8.5% return. Looking at the number of bets placed with the simulated and real procedure, the authors recognized a problem. In fact, the former can count a number of bets equal to almost ten times the bets placed in a real environment. The explanation for this phenomenon is that the player cannot bet on all available opportunities 24 hours a day and, thus, he misses many of the bets that satisfied the betting conditions. Nevertheless, the paper and the real betting activity confirmed the profitability of the strategy, as shown in Figure 11:

Figure 11 Value strategy returns



Although they wagered according to the sports betting industry rules, just a few months after they start the committing real money, they started to see their accounts limited by bookmakers. Some of their bets were limited in the stake amount they could lay, and some others were denied by the bookie, who often suggested a value lower than the desired one. Under these circumstances they were forced to suspend their betting strategy. This practice was not unexpected. As explained in the first chapter, bookies try to compensate the inefficiency of the football betting market by implementing restrictive practices. At the end, this strategy has demonstrated to be able to find “value” and to be profitable, but, while actuating it, the punters can encounter some difficulties such as the limitation of their accounts.

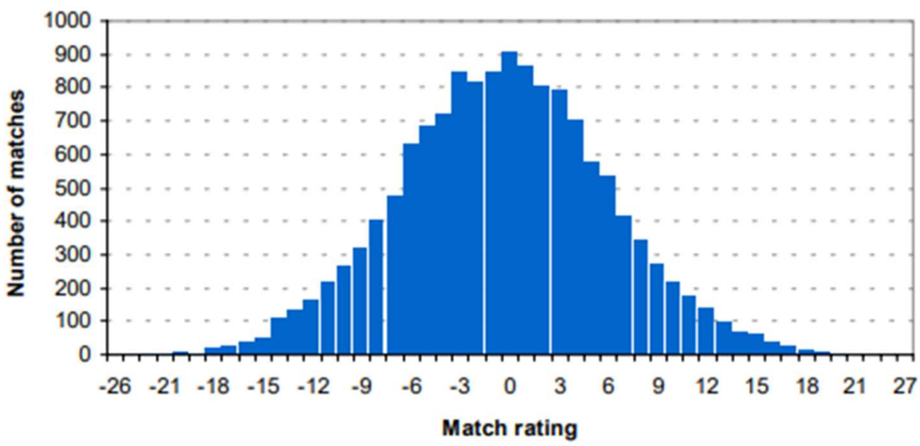
4.3.3 Value found through analysis of historical data

Another useful way to estimate the real probability in order to find mis-priced odds to bet on, is the analysis of past data. To predict football matches accurately several things are needed. First of all, a set of features containing the most valuable information about previous matches must be collected. After that, it is necessary to build a probability distribution of each possible outcome, from which we can calculate the probabilities. At the end, this strategy proceeds as the previous ones: bet on the most valuable outcomes, that are the ones where the bookmakers’ odds move away from the true probability. The way to compute the probability distribution of each outcome is not unique. Thus, there are several techniques to build the probability distribution. One of them is using the rating system.

The website https://football-data.co.uk is one of the most popular sources of historical data regarding football sports betting, and many paper have downloaded the odds from it. Besides being an important data store, it also provides several guide and explanations to its visitors. In particular, it supplies also a general guide regarding the rating system taken from Buchdal's book (2003). It defines it as (p.1): "a quantitative measure of the superiority of one football team over its opponent in a match, which is determined by analysing and comparing one or more statistics of past performance for each of the sides". Such superiority can be assessed in different ways, but usually each method calculates a points difference between points rating for the away side and points rating for the home side in a football match. The home and away team points ratings are computed by selecting different aspects of a team's strength and scanning them through a quantitative analysis. The statistics employed do not have to necessarily be complex. For instance, they could be either league points, league positions or goals conceded and scored, whilst more complex ratings might be based on elaborate match statistics including shots on goal, corners, and even ball possession if such data are available.

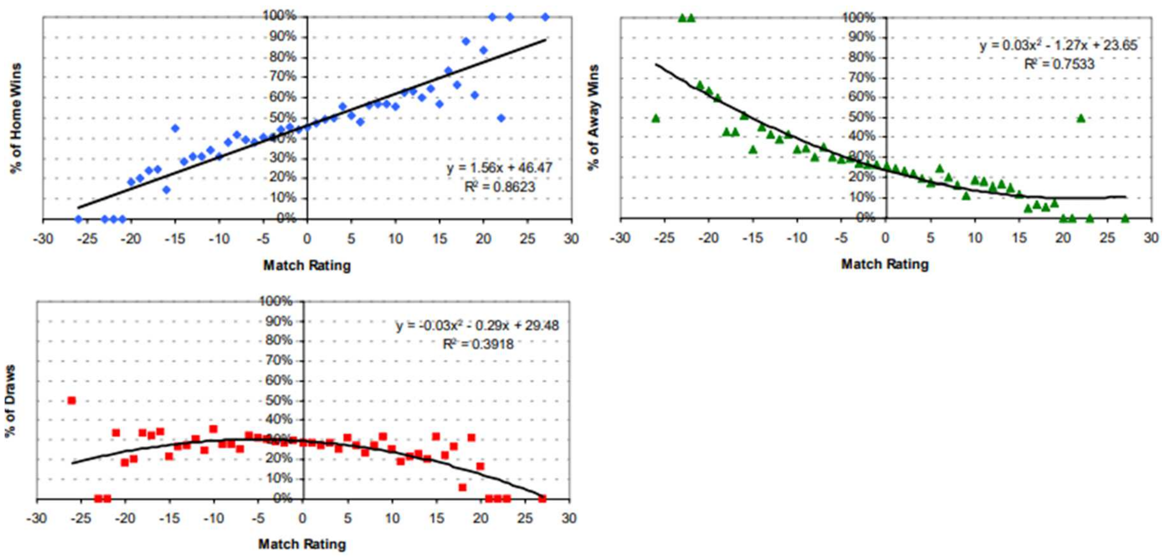
Buchdahl (2003) proposed a simple rating system based on goals superiority. The basic assumption behind this approach is that teams who score more goals and have a better defence over the course of a championship are have grater chances to win their next game. This rating value is calculated simply by taking the goals difference in the last 6 matches, using data from the English Premier League as well as the second, third and fourth British divisions for seasons 1993/94 to 2000/01. For example, if a team has scored 11 goals and conceded 8, its rating would be +3. In the time interval considered, 14,002 games were eligible for a rating calculation. The number of matches with each rating level is shown in Figure 12.

Figure 12 Distribution of games according to match rating



The main objective is to see how frequently a match with a certain rating value ends with a 1, X or 2 result. In the chart above, ratings vary from between -26, that implies a very weak home side and strong away performance, to +27, with vice versa considerations. Looking at past data with a rating of +3, the implied probability of a home win is 49.7%, 28.2% for a draw and 22.0% for an away victory. In general, as greater is the match rating, as higher is the expectancy of a home win. Conversely, when the rating is very negative, the visitor team victory is the most likely outcome. The draw is another time difficult to infer, because the relation between the match rating and the likelihood of a drawn is not well defined. With a rating of +15, only 57% of the games ended with a home victory, and with a rating of -15 only 34% of games processed finished with an away success. The author explains that (p.4): “these discrepancies arise because the relationship between the match rating and the probability of a result is inherently “noisy” and imperfect”. Moreover, these discrepancies appear greater in magnitude for the extreme ratings, for which, due to the limited number of matches in those groups, one or two results can singularly determine and shape the results probability distribution. To manage this variance, they had to standardise their forecasting model. By doing so, they could make a practical attempt at defining the fair odds for a football game. The first issue they faced was to consider each result independently and recognize the "best-fit" relationship with the match ratings. By plotting the relation between rating and specific probability of each outcome (Figure 13), they retrieved the best-fit equation. For each equation, the dependent variable y, the probability of a particular result, is some function of x, the match rating. The value of R^2 attributed to each one of the three estimated equations, is a statistical measure of how closely the real data match the best-fit lines.

Figure 13 Distribution of home (blue), draw (green) and away (red) victory by match rating

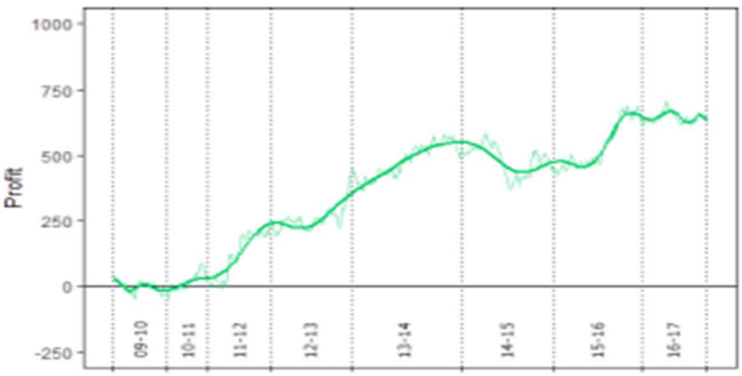


With each equation it is possible to determine the expected probability and hence the fair odds of a home win (blue), draw (green) and away win (red) occurring for any match handled. Considering the previous example, where the match rating was +3, the probability of a home win, for example, can be determined using this equation: y = 1.56x + 46.47 where y = the probability of a home win and x = the match rating. Hence, in this case the home win probability is 51.35%, that is close to the previous 49.7%. Consequently, the fair odds according to this forecasting model is 1/0.5135 or 1.95. We can define fair odds for the away win and draw in the same way, using the equations for away wins and for draws. By comparing the model's fair odds with the bookmaker ones, that incorporate the margin, it is possible to declare whether we are facing a value bet or not. If the bookmaker's odds are higher that the odds estimated with the rating system, we have the possibility to exploit this difference by betting on the fixed-odds, assuming that the ratings system is effectively able to give a trustworthy expectancy of the result. If it is the case, an investor should make a profit over the long-term. With access to historical results and betting odds, the author tested its model using English Premier League games played during the 2001/02 season, thus retrospectively. By taking the best available odds (from 6 bookmakers), the goals supremacy rating system would have returned a profit of 2.1% for all match ratings. Thus, this system seems entitled to be called an effective "value" betting strategy.

Verkerk (2018) employed a Poisson distribution to build a goal distribution model. In his paper, data are collected for 13 seasons (2004/05-2017/18) in the Dutch Eredivisie. The complete dataset consists of 3,832 matches. The data included in the model are: the date and the season at which the match is played, the home and the away team, the full time number of goals scored by the home and the way team, and the full time result (1, X or 2). These data lack a strength indicator per club able to quantify the ability of all teams at a certain point in time and to shape the basement of a statistical model. In this research study, the increasingly popular Elo rating system is engaged. This system was designed in 1960 as a rating system related to chess games, but nowadays it is used as a football rating system as well and a weekly updated publicly available Elo system for football teams already exists. Another strength indicator used is the annual budget per club, assumed to be constant every year. The Elo rating of a club can take into account also matches played years ago, that consequently still have an influence on the present. Therefore, it might be useful to create a set of variables that measure recent form only. For example, an appropriate dimension can be the overall form over the last 2, 5, ..., 17 matches, measured looking at goals scored, goals conceded, and points won. However, just counting all the recent form measures is not adequate enough, since this would be affected more by the quality of the opposing teams instead of assessing the result obtained weighted for the rivals 'quality. Verkerk solved this problem by multiply the results of all clubs by their predicted probability, acquired from the bookmaker odds, of winning the match. With this adjustment short-term luckiness gets assigned lower scores than a favourite club with a solid superiority. Hence, teams with a higher probability of continuing their current form are rewarded with higher scores.

The goals model is based on Maher's (1982) inspiration that the number of goals scored by both teams can be estimated with the Poisson distribution. In a Poisson regression it is assumed that the independent response variables Yi, i = 1, ..., n are Poisson distributed with mean μi. In football matches there are two teams that score their goals following a Poisson distribution. In this paper, the author assumes that the goals of both teams are reciprocally independent because of the small correlation coefficient, thus they can be estimated separately. The size of the train set is h matches, so that the train set for some match i contains the range of matches $\{i - \mathrm{h} - 1$, $i - \mathrm{h}$, ..., $i - 2$, $i - 1\}$. The betting rule is always the same: place a bet if the expected profit is greater than zero. This occurs whenever the predicted probability is greater than the bookmaker's one for one of the three possible match outcomes. The difference between both predictions for some match i is the expected edge, named E_i. The strategy is tested for different threshold level T, with $T \geq E_i$. It seems that the ROI increases with a more conservative (increasing) threshold. In fact, as shown in Figure 14, the most remunerative model is the one with the threshold imposed as $T \geq 7.5\%$.

Figure 14 Growth of the betting profit over all seasons for the Poisson model



However, since T is a measure of the bets' "quality", by enhancing it less matches would meet the requirements of a greater "value", therefore the variance between the ROI per season would significantly increases as well. The average amount wagered is a little over €6. The number of bets placed per season is not constant, which is shown by the width of each interval, that stands fora season, in Figure 14. A larger range between the vertical dotted lines indicates that more value bets were detected in that particular season. To conclude, this paper has demonstrated that it is possible to compete with bookmakers also in the Dutch Eredivisie.

A further analysis of the application of the Poisson model on goals distribution has been extended by Angelini and De Angelis (2016). In fact, their paper focuses on football matches prediction using the Poisson Autoregression model with exogenous covariates introduced by Agosto et al (2016), called the PARX model. This model is an extension of the Poisson Autoregression model, that aims to include covariates in its specification. It has been successfully used to predict corporate defaults and Angelini and De Angelis discovered that it fits particularly well also in the framework of football betting, since the PARX model has the quality to be able to account for the overdispersion in the marginal distributions and the autoregressive persistence, a typical feature of the intensity of the goals scored by a team in a football match. The main difference of this method with respect to the one used by Verker is that the PARX model is capable of capturing the dynamics of the intensity of the goals distribution, hence it exploits additional information by including exogenous covariates in the model specification which can substantially improve the prediction accuracy. In fact, also this model makes use of covariates of strength and form that round the attack and defence ability oh the teams involved, which are useful for predicting football matches results.

To begin, the authors preliminarily analysed the goals distribution of four teams which played the English Premier League from season 2005/2006 to season 2014/2015. By comparing the marginal empirical distributions and the corresponding Poisson distributions, the authors evidenced the presence of overdispersion for all the four clubs, where the average number of goals is smaller than the variance.

To estimate the outcomes probability for each match, Angelini and De Angelis employed two PARX models, one for the home and one for the visitor team. They named y_t as the number of goals scored by a football team at time t, where t = 1, ..., T, λ_t as a measure of the intensity of the model, while H stands for the home team and A for the away team in a specific match, so that $y_t^H|F_{t-1} \sim Pois(\lambda_t^H)$ is the distribution of the goals scored by the home team when it plays at its stadium, and $y_t^A|F_{t-1} \sim Pois(\lambda_t^A)$ is the number of goals scored by the visitor team when it plays away, conditional on the available information set F_{t-1} at time t-1. They firstly estimated two conditional Poisson distributions for home and away goals using the PARX model, and then they derived the two forecast distributions. In this case, the points of interest are one-step ahead forecasts, $y_{T+1}^i|F_T$, for i = H, A, which denotes the number of goals scored by a team in the next match, always given the information set available at time T. The conditional distribution of one-step ahead forecasts is a Poisson of parameter λ_{T+1}^i, thus for having a prediction of the number of goals in the upcoming game, it is indispensable to compute

λ^i_{T+1}. Conditional on all the available information at time T and given the vector of parameters θ, the value $\lambda^i_{T+1}|T$ is equal to:

$$\lambda^i_{T+1|T}(\theta) = \omega + \sum_{j=1}^{p} \alpha_j \lambda^i_{T+1-j}(\theta) + \sum_{j=1}^{q} \beta_j y^i_{T+1-j} + \gamma x_T$$

where x_T denotes a vector of m exogenous positive covariates. The parameters $\omega > 0$ and α_1, ..., α_p, β_1, ..., β_q, $\gamma \geq 0$ are time invariant and, when $\gamma = 0$, the PARX reduces to the PAR model. After having computed a point forecast of the underlying intensity, $\lambda^i_{T+1|T}(\theta)$, the subsequent step is to forecast the distribution of λ^i_{T+1} as:

$$\hat{P}\left(\lambda^i_{T+1} = y^i | F_T\right) = Pois\left(y^i \middle| \lambda^i_{T+1|T}(\theta)\right), \qquad y \in \{0, 1, 2, \dots\}$$

For the purposes of this project, it is recommended to firstly derive the forecast distribution for the home and away teams, and then, imposing the independence of the two distributions, derive the joint forecast distribution:

$$\hat{P}(y^H_{T+1} = y^H, y^A_{T+1} = y^A | F_T) = \hat{P}(y^H_{T+1} = y^H | F_T) * \hat{P}(y^A_{T+1} = y^A | F_T)$$

In other words, they used the estimated intensities $\lambda^H_{T+1|T}$ and $\lambda^A_{T+1|T}$ to derive the probability associated with any possible outcome. For instance, the probability of a home team victory is given by $\hat{P}(y^H_{T+1} > y^A_{T+1} | F_T)$, the probability of a draw is $\hat{P}(y^H_t = y^A_t | F_T)$ and the probability of an away win is $\hat{P}(y^H_t < y^A_t | F_T)$.

In their analysis, the two scholars considered the mean of goals conceded by the opponent team as the only covariate, hence it plays a crucial role since it can be interpreted as a clear indicator of the defence ability of the team. In fact, a club with a low value of x_{t-1} is supposed to have good defence skills, therefore it would be more likely to concede a smaller number of goals than a team with a high value of x_{t-1}.

Once the joint distribution is computed to derive the probability associated with any possible outcome for each match, it is easy to develop a betting strategy for the 1, X, 2 market. Each possible outcome has a quote associated, computed as the average of a series of odds from a group of bookmakers. The betting strategy proposed is based on two conditions: select the result associated with the highest probability and evaluate if the odds is available for that result is high enough to make it worth to bet on it, thus if there is a considerable positive expected value. Let P^0_b be the probability retrieved making the inverse of the odds associated with result b, for b = 1, X, 2. The expected value of a bet is then given by $[B_b] = \frac{P_b}{P^0_b} - 1$. In the case when $E[B_b] >$

0, it is reasonable to place a bet on the most convenient bookmaker. This betting strategy actuated on the matches played in the 2013/2014 and 2014/2015 English Premier League seasons, led to a return of 43.27% and 44.96% by placing 56 and 48 bets respectively.

The authors proposed an additional alternative strategy, under which they only bet on the match outcomes whose profitability is higher than a specific threshold τ. As one might expect, the return of the betting strategy improves for higher values of τ and the number of bets decreases. In fact, the number of bets decreases to 36, 30, 14 for the first and 29, 19, 12 for the second season when $\tau = 0.1$, 0.2 and 0.3, respectively. $\tau = 0.3$ led to the best performance in terms of percentage return, that is 76.36% and 124.08% for 2013/2014 and 2014/2015 seasons. The interesting feature is that the percentage of winning bets did not decrease with the value of τ, but it remained rather constant. This result is particularly interesting because it is clear evidence that this approach is able to detect the mispricing of the odds offered by the betting market, without any loss in the forecasting ability. Indeed, the higher the value of τ, the higher the under-pricing of the odd.

4.3.4 Value found through a machine learning approach

Machine learning (ML) is a recent area of study for predicting sport outcomes. It is an appropriate methodology for sport prediction since it generates predictive models that can predict match results using predefined features in a historical dataset. In the recent period, it has been increasingly used both by bookmakers, who have improved their efficiency, and arbitrage seeking investors. Despite the other techniques presented in this chapter, this way of forecasting requires high informatic competencies and it needs a huge volume of relevant factors and data. Thus, it is not an easily implementable tool. The increasing amount of data related to sports that is now electronically and often publicly available, has meant that there has been an increasing interest in developing intelligent models and prediction systems to forecast the results of matches.

One of the first attempts to use artificial intelligence and machine learning algorithms to predict one of three classes for each game (win, draw, or loss) was done by Ulmer and Fernandez (2014). From historical data they created a feature set that includes whether a team is home or away, features for each team (similar to the Elo rating used by Verkerk (2018)), and form, measured looking at recent results. They based their model on a training dataset containing the results of 10 seasons (from 2002-03 season to 2011-12 season) of the English Premier League, and then they applied the model on a testing dataset composed of the results of 2 seasons (the 2012-13 season and the 2013-14 season) of the same league. While building their model, they complained two main difficulties: a lack of data, and the randomness of the data. Since the data covered a gap of 12 years, there was always a difference in the amount of information available between each season. As such, they were limited to the features from the 2002 season in order to align them. Then, they also found a high entropy (a measure of randomness) that made their data hard to classify. They tried a variety of classifiers to predict outcomes probabilities: Linear from stochastic gradient descent, Naive Bayes, Hidden Markov Model, Support Vector Machine (SVM), and Random Forest. Their best error rates were with our Linear classifier (.48), Random Forest (.50), and SVM (.50) and overall, every model under-predicted draws to some degree. The same authors admitted that their project still lags behind leading industry methods, mainly due to the limited amount of data from the earlier seasons of the dataset.

Nowadays, informatic technologies allow to store an infinite amount of data regarding every aspect of a match. That's why this field is now attracting a large research effort. Bunker and Thabtah (2017) wrote a paper that provides a critical analysis of the previous literature in ML, focusing on the application of Artificial Neural Network (ANN) to sport results prediction. They identified the learning methodologies employed, data sources, appropriate means of model evaluation, and specific challenges associated with predicting sport results. All these things taken together, entitled them to propose a general sport prediction framework based on machine learning techniques, in order to transform this information in a profitable betting strategy. Artificial Neural Networks (ANNs) are perhaps the most commonly applied approach, among ML mechanisms, to the sport result prediction problem. An ANN usually contains interconnected components, named neurons, that convert a set of inputs into a desired output. The ANN can be considered as a powerful tool thanks to its non-linearity of the hidden neurons in adjusting weights that determine the final decision. ANN output relies on input features and other variables related with the network, such as these weights. The ANN model is constructed after processing the training dataset that contains the features used to build the artificial classification model. In other words, the ANN algorithm continuously updates the weights associated with interconnected components in order to reach high levels of predictive accuracy. As warned by the authors, this may lead in some cases to the problem of overfitting, as well as wasting computing resources such as training time and memory.

The scholars ‘intelligent architecture for sport results prediction is now presented, considering each of the six steps of their ML framework, describing the characteristics of the data used for sport results prediction, and how they fit within the framework.

1. Domain understanding: understanding the problem, the goal of the modelling, and the specific characteristics of the sport itself. This involves having some comprehension of sport regulation and which factors are potentially involved in determining the outcome of matches.
2. Data understanding: first of all, find a reliable source of data and possibly automating the data collection process. The level/granularity of the data must be considered. For example, deciding whether or not to include player level data, that would give the possibility to investigate whether specific players' actions or presence can influence thr final result. The definition of the possible outcomes needs to also be considered. Most prior work has treated the sport prediction problem as a 2 (home win or not) or 3 (home win, draw, away win) class values classification problem.
3. Data preparation & feature extraction: the features set must be split into different subsets, for example dividing match-related features, like meters gained, passes made, ball possession, and standings or external features, like recent form or players available. Then, apply feature selection algorithms to select most important variables from original features and feature subsets. Match-related features must undergo a separate averaging process before being re-merged with the external features, since the latter are known prior to the upcoming match to be played.
4. Modelling: the first step is to choose which candidate models to engage in the experimentation. Then trail each model on each feature subset, and subsets that have been selected by the previous passage and choose the best combination of classifier and feature selection technique.
5. Model evaluation: to evaluate the model performance, the easiest thing to do is to count the numbers ended in home wins, away wins and draws and then check for the number of matches that the model has correctly identified. An appropriate training-test split needs to be decided. Usually, professional sport competitions are organized in rounds. When considering a particular season, the number of rounds that used for training and then testing the model must be determined. For example, in a data set with 10 rounds of data, the first 7 rounds of the competition could be used for training the model and the other 3 rounds could be used for testing.

6. Model deployment: the new round of data is obtained from the web and added to the match database. The training data and test data are then adjusted, the model is retrained with the new training data, and the loop restarts.

This model is a general guideline that can be applied to different sports. Due to the fact that each sport has its specific nature of match-related features, results across different studies in this application cannot generally be compared directly. In the end, the authors conclude that Machine Learning seems an appropriate methodology for sport prediction since it generates predictive models that can efficiently forecast match results using predefined features in a historical dataset.

Stubinger and Knoll (2018) are of the same opinion. With machine learning, football league matches hold enormous potential for developing betting strategies, since an enormous quantity of matches are played every week throughout Europe and the other continents. In their study, they constructed a betting strategy for football league matches based on machine learning algorithms. In particular, they benchmarked their strategies with trading variants which rest upon fixed betting odds. Moreover, they performed a large-scale simulation study based on a data set composed of 39 variables for 8,082 football matches of the "Big five" European football leagues.

First of all, they collected odds from the online bookmaker Bet365 and data from the Premier League, Ligue 1, Bundesliga, Serie A, and Primera Division from season 2013/14 to 2017/18. For each of the 8,082 football matches analysed, the data set contains general information about the game, ball possession, pass behaviour, disciplinary measures, defence and attack abilities for both the home and the away team. Since the principal focus of this research paper is to predict which team would win a specific football match, the difference between the number of goals scored by the two teams plays a decisive role. In total, 3,723 home team wins (46%), 1,967 draws (24%), and 2,392 away team wins (30%) were observed. This is in line with the previous finding, that the distribution of the difference between home team goals and visitor goals is asymmetric, which means that there are more matches with a positive goal difference, hence a home victory, than with a negative goal difference, or away win. Surprisingly, the most frequent outcome of the football match was the draw, thus a zero difference between home and away teams.

Different machine learning approaches are applied to forecast the goal difference in a football match (dependent variable y) based on the game covariates observed up to that point (independent variables x). The data set is divided into overlapping study periods, each one shifted by one match day. Each study period includes, as the previous research presented, a training set, which contains all the past matches, and a test set, which represents the considered match day. To model the relation of the dependent variable y as a function of the 39 independent variables x, different common machine learning approaches were employed:

- Random forest (RFO): one of the most popular machine learning framework, that works by building a variety of decision trees to output the final result, i.e. the mean prediction of each tree. The decision trees must be constantly updated.
- Boosting (BOO): boosting is a meta-algorithm for machine learning that merges several weak classifiers into a single strong classifier. This technology reduces bias and variance as well as memory requirements and runtime.
- Support vector machine (SVM): it divides a set of objects into classes in such a way that as wide an area as possible remains free of objects around the class boundaries.

Only information before the match starts was included to avoid any look-ahead bias. Therefore, x is determined according to the team characteristics of the last matches. Finally, the dependent variable was estimated using the models presented above (RFO, BOO, SVM). The betting strategy proceeds as follow: if $\hat{y} > 2$ means the model predicted that the home team will win, while a value of $\hat{y} < -2$ intends that the away team is the most likely to win. In these cases, the strategy would bet 1 monetary unit on home win and away win respectively. The trading thresholds of ±2 was introduced to avoid betting on matches with an expected profit close to zero. A greater distance from 0 guarantees higher chances of success. Assuming to be a conservative investor, the thresholds to ±2 should lower the risk of losing high portion of capital, as we are betting only on clear predictions. In fact, when $\hat{y}$ is inside this interval, it means that the model is uncertain whether the match will end in a home team win or away team win. Consequently, we do not execute any bets.

From a statistical point of view, RFO achieved the highest accuracy, then the lowest root mean squared error (RMSE), while the lowest was mean absolute deviation (MAD). RFO is followed by the other two machine learning methods BOO and SVM. Not surprisingly, the random bet RAN clearly led to the poorest performance. The accuracy of the different approaches was confirmed by the corresponding average payoff. In addition, all the machine learning approaches reveal average payoffs greater than 1, in contrast to the random method. An average payoff greater than 1 is entitled to be recognized as profitable, as it beats the bookmaker on a long-term perspective. Moreover, betting on the home outcome is clearly preferred by all approaches, and, in general, a higher number of bets leads to lower average payoffs. It could be carefully concluded that the machine learning methods are better at selecting the most valuable opportunities out of the pool of all possible matches. Statistical analysis is also reflected in the economic context: a high prediction quality corresponds to meaningful profits and vice versa. The approaches based on machine learning algorithms provide positive returns ranging between 0.5% per match for SVM and 5.42% per match for RFO. As expected, the naive strategy RAN produces a clear loss of −6.30% per match. Thus, this paper is another prove that a data-driven approach for predicting the outcome of football league matches can beat the bookmakers and generate positive returns, and that risk-averse strategies perform better than risk-taking ones.

Stubinger, Mangold and Knoll published an analogous study in 2019, but in this case, they took in consideration more seasons of the same championships (from 2006/2007 to 2017/2018, for a total of 47,856 football matches) and more players characteristics, 40 in total. Using the same machine learning techniques, they obtained slightly different results, but the main findings are confirmed. Random forest (RAF) achieved again the highest accuracy (81.26%), the minimal root mean squared error (RMSE) as well as the lowest mean absolute deviation (MAD). In general, a large number of bets seems to negatively affect the accuracy, which indicates that a risk averse betting behaviour is more convenient in this context. Random forest and Boosting ML approaches provided returns of 0.43% and 0.72% per bet, respectively. The random strategy was not able neither in this case to generate positive returns. Combining all the different machine learning algorithms achieved economically significant returns of 1.58% per match. Thus, these results confirm the statement of the existing literature that forecasting football league matches results in positive payoffs.

A particular research has been conducted by Hubacek et al. (2019). The first innovative aspect concerns the data set: it is composed of NBA[17] data during the period between 2007–2014. Compared to football, basketball has only two possible outcomes as a final result, since the draw does not exist. Moreover, the authors introduced three main innovations in the field of sports prediction with machine-learning. Firstly, previous attempts to build models for match-outcome prediction were focused on maximizing the model's predictive accuracy as the single criterion. They found out that for each level of constant accuracy, increasing the correlation between the model and the bookmaker corresponds to a profit drop in all settings. Thus, unlike the classical approaches, they also reduce the model's correlation with the bookmaker's predictions available through the published odds. They show that such an optimized model achieves better profit, and the approach is thus a way to 'exploit' the bookmaker. The second

[17] The American National Basketball Association

novelty concerns the application of convolutional neural networks, that are a suitable model to leverage player-level data for match outcome predictions. Convolutional networks are defined as: "a specialized type of neural networks that use convolution, a mathematical operation on two functions that produces a third function that expresses how the shape of one is modified by the other, in place of general matrix multiplication in at least one of their layers"[18]. Thirdly, they design a strategy for bet distribution according to the odds and model predictions, trading off profit expectation and variance optimally. In other words, beyond maximizing the profit's expectation, they also aim to minimize its variance. The strategy that turns out to be the best is the one that maximize the Sharpe ratio, intended as: $\frac{\hat{E}[P]-R}{\widehat{\sigma_P}}$, where the numerator is the estimated expected profit minus the profit from a risk-free investment of the disposable wealth, such as through banking interests, and the denominator is P's estimated standard deviation, thus the square root of the estimated variance. Overall, this strategy generated a positive profit of more than 10%, from a number of 500 bets placed.

By combining the power of machine learning methods and the large set of data, Nielsen and Sandøy (2019) aimed to create a data-driven system able to generate models for predicting the outcomes of football matches and generating a profit on the sports betting market. Their report explores different machine learning methods. Two match prediction models using artificial neural networks are developed, where one is a fully data-driven adaptation of a successful Bayesian network model utilizing domain expert knowledge. The effect of utilizing domain knowledge in the structure of neural networks is explored by grouping related input features into separate sub-networks. The profitability of the models is evaluated over two seasons of the English Premier League using different money management strategies. As a second tool, they performed experiments in using reinforcement learning methods to train a money management agent to explore the possibility of building a complete end-to-end betting system. Some model-strategy combinations were able to generate a profit over both test seasons, showing that it is indeed possible to profit from football betting using artificial neural networks.

Value betting is not only theoretical, but, as matched and sure betting, is practically implemented. In recent years, several companies have started to offer a service that provides to their customers their estimated odds, comparing them on the betting market, looking for value opportunities. The number of clients is growing rapidly. One of the most interesting realities in this market is Mercurius, an Italian online platform operating mainly in England. Their mission is to obtain long-term profit by spotting inefficiencies in the offered odds, primarily in the

[18] https://en.wikipedia.org/wiki/Convolutional_neural_network

Betfair exchange market. Although on average, the odds offered in the exchange are more efficient than the ones provided by traditional bookmakers, Mercurius' software fits better in the exchange market due to the larger liquidity available and the lack of restrictions. Moreover, with a large number of bets, it is possible to lower the commission up to 2%.

The way of operating is similar in this type of company, and it can be divided in three steps. The first one consists in compiling odds via fundamental analysis which relates to the process of measuring a security's intrinsic value to find its 'fair market price'. This is made by collecting a large number of data and variables that can help to improve the forecast accuracy. For example, Mercurius takes into account for every match data like the starting and finishing position of each ball movement, the types of ball movements, whether it is a cross, shot, or header, the outcome of each movement, the players involved in each ball movement etc. Then artificial intelligence and statistical models like the Monte Carlo simulation are used to establish team ratings and to find fair odds. Monte Carlo involves running data through hundreds of thousands of simulations to see all the possible outcomes of a decision. In the second step, the generated fair odds are compared to the ones offered in the fixed-odd or in the exchange market. Finally, bets are placed when the expected value is positive. The main innovation of Mercurius stays in this last step: once the software discovers a value bet, it executes the process automatically according to a specific plan. Thus, the client can just invest some money, and the platform will place his bets automatically, following a predetermined strategy. Since its beginning in 2019, Mercurius has reached for its clients a Return on Capital (ROC) of 28.53% out of 1463 bets[19]. As expected, the volatility is high, around 26%. Therefore, this type of investment must be pursued in the medium-long run. The company usually earns on a maintenance fee, a small percentage of the money bets, and on a commission on profits generated.

All the studies about value betting proceed in a similar way, but to find the value is not sufficient to make this strategy profitable. Money must be managed in a proper way. That's why portfolio maximization theory is a core issue if we want to manage returns and volatility, otherwise we could lose our entire investment in the short run. The next section will explain the most common theories employed.

[19] Data are updated until 05/09/2021. All the bets placed are consultable at https://mercurius.io/en/trader-app/performance.

4.4 Money management

Much of the available research on betting systems is centred around the predictive modelling part, often completely neglecting the need for a betting portfolio optimization strategy. Whether or not a model makes a profit is determined by a combination of the accuracy of the model and the type of betting strategy the player implements. Thus, money management in betting is of crucial importance. The punter must know how much to invest for each betting opportunity, to minimize the volatility and maximize the expected value. The betting strategy has a strong influence on the final measures of profit. Consequently, a worse model with a better strategy can easily outperform a better model with a worse strategy. The general principle applied by companies that are currently operating in this sector, is that the amount stacked for a single bet should never be greater than the 2% of the total capital invested, but usually the percentage invested is around 1%. Besides that, there are two basic streams of research in the formal approaches, stemming from information theory and economics, respectively. The first, and the most widespread, is the Kelly criterion (Kelly, 1956), also known as the geometric mean policy, maximizing the expected long-term growth of wealth. The second is the approach of Markowitz's modern portfolio theory (Markowitz, 1952), balancing the criteria of expected profit and variance as a measure of risk. Generally, Markowitz's approach has traditionally dominated the world of quantitative finance, while Kelly's approach has been more prominent in the sports betting industry.

A detailed analysis of these two methods applied to three different sports (horse racing, football and basketball) is presented by Uhrín et al. (2021). Markowitz's modern portfolio theory (MPT) is a standard economic view of the problem based on the idea of the expected profit, usually translated into a utility function reflecting the user's particular preferences. The general idea behind is that a portfolio $\boldsymbol{f^1}$, i.e., a vector of assets $\boldsymbol{f} = f_1, \dots, f_n$, is superior to $\boldsymbol{f^2}$, if its corresponding expected profit, given $\boldsymbol{b}$ equal to the decimal odds – 1, is at least as great as E[$\boldsymbol{b} \cdot \boldsymbol{f^1}$]≥ E[$\boldsymbol{b} \cdot \boldsymbol{f^2}$], and a given risk measure risk : $R^n \rightarrow R$ of the portfolio, with respect to the given odds, is $risk\ (\boldsymbol{f^1}|\boldsymbol{b}) \leq risk\ (\boldsymbol{f^2}|\boldsymbol{b})$. This rule gives some order among the set of all the possible portfolios. The goal is to select a set of efficient portfolios by selecting the portfolios that no other portfolio is superior to. In simple terms, we trade off the expected profit-risk by maximizing the expected profit minus the risk measure multiplied by a parameter reflecting the user's preference for risk. To select a particular portfolio from the efficient frontier, the maximum 'Sharpe ratio' can be used, as shown in section 4.3.4. The MPT approach is often criticized for the arguable choice of risk level, which can be perceived as a formal weakness of the approach, since in many cases the risk is not easy to define. Moreover, the direct maximization of expected profit can be misleading in matches where the mean profit is very different from the median.

The second money management tool is called the Kelly Criterion. It has proven to work efficiently in betting, for this reason it is commonly used. It was named after an American economist John L. Kelly (1956) and originally designed for information transmission. This strategy suggests that the percentage of your bankroll, the total capital invested, that should be bet is equal to: $\frac{b \cdot P-(1-P)}{b}$, where P denotes the probability of the event and b is again equal to the decimal odds – 1. For example, if a gamble has a 60% chance of winning, and the odds on a winning bet is 2, then the gambler should bet 20% of the bankroll at each opportunity in order to maximize the long-run growth rate of the bankroll. The Kelly Criterion increases the stake size whenever the expected profit increases. Therefore, the stake size is positively related to the probability of success as well. This criterion is based on the idea of expected multiplicative growth, so that a portfolio $\boldsymbol{f}$ is chosen such that the long-term value of the resulting, continuously reinvested, wealth W_t is maximal in an infinite horizon of t. This betting strategy is preferred if you want to grow your bankroll more quickly. However, this comes with more risk of getting bankrupt. In contrast to MPT, there is no explicit term for risk here, as the notion of risk is embedded in the growth-based view of the wealth progression, i.e., the long-term value of a portfolio that is too risky will be smaller than that of a portfolio with the right risk balance. The main flaws of this approach underlined by the authors are its unrealistic assumptions: we are assuming that we know the true probability distribution of game outcomes, we are repeatedly presented with the same games, and we play for an infinite amount of time.

The core issue regarding the mathematical strategies is that their calculations are carried out as if the true probability distribution over the outcomes was known. Further, they are often sensitive to even the slightest error in the estimates. Some adjustments can be made to manage the extra risk stemming from the underlying errors, as well as more sophisticated techniques incorporating the uncertainty of estimates directly into computation of the strategies. The most trivial risk-avoiding technique is restricting the maximum percentage of bankroll that can be bet to a fixed value m. Another efficient practice is the fractioning. It consists in betting only a fraction ω of the calculated portfolio and keep the rest of 1 – ω in the bankroll. The choice of ω should depend on the actual distributions and the player's attitude for risk. The same idea of fractioning can be applied to any strategy, including MPT, and it is overall useful whenever the estimates are erroneous. A third possible technique is the drawdown, which is more invasive since it actually modifies the original optimization problem. The idea of drawdown is to incorporate a special probabilistic constraint into the Kelly strategy so as to push the solution away from the riskier region. The choice of the threshold is then left to the user's preference as an input parameter into the optimization problem. The probabilistic boundary is expressed as the following constraint: $P\left(W_t^{min} < \alpha\right) \leq \beta$, expressing that the probability of our wealth falling below α can be at most β.

The final purpose of this paper by Uhrín et al. (2021) is to assess the performance of the individual strategies and their risk modifications presented, in various realistic settings on real data. The strategies for the experiments were chosen with the aim to represent the diverse portfolio of approaches occurring in practice, with the goal to provide an unbiased statistical assessment of their performance. They collected 3 datasets of different properties from 3 different sports (horse racing, basketball and football), each containing a significant number of 'matches' (races and games) for statistical evaluation. The key properties considered are: dataset size, the accuracy of the bookmaker, the accuracy of the predictive model, the number of possible match outcomes, the range of the offered odds, the average margin present in the odds and the Kullback–Leibler[20] advantage of the player. The latter plays a key role in the performance of the original Kelly strategy, where the growth of profit can is directly proportional to the KL advantage. In total, this study considers 2700 horse races, 16000 basketball games and 32000 football matches.

The models providing the probabilistic estimates were trained following the chronological order of the matches, so that all of their estimates are actual future predictions. The betting investment strategies analysed are based on the views of the MPT and the Kelly criterion, together with a number of their popular modifications aimed at additional risk management in practice, where their original underlying mathematical assumptions do not hold. While many of the chosen strategies contain hyperparameters to be set, they additionally tuned each for the best possible performance via grid-search, too. To provide an unbiased estimate of their actual performance in practice, they also followed a strict evaluation protocol for each of the strategies. This means that they have split each dataset into two groups, the training and testing subsets, found the best hyperparameter setting on the training subset and evaluated the fixed setting on the test subset. In order to make the output profit measures more robust, both the training and testing are evaluated by generating 1000 separate 'runs' through each subset, where the sequence of games is randomly reshuffled and 10% of games are randomly removed each time. They hence evaluate properties of each strategy on 1000 separate wealth investment trajectories through previously unseen games. To select the best possible strategy setting on the train set, they

[20] Statistical measure of the difference of the predictive performances (cross-entropy) of the player and the bookmaker, respectively.

always chose a strategy that reached the maximum median final wealth, given that no more than 5% of the wealth trajectories did not fall below 90% of the final wealth. Hyperparameter settings that did not satisfy the required condition were removed.

To express a final evaluation of the strategies on the test set, they chose different metrics to provide more insights into the properties of the individual strategies and game settings. The metrics are: the median and mean final wealth position, the lowest and maximal wealth position, the standard deviation of the final wealth positions and the ruin percentage of wealth trajectories, for which they defines a ruin situation whenever the initial bank W_0 reached a value below 0.01% during the entire investment period.

The conclusions from the experiments regarding the discord between the two approaches roughly follow the initial intuitions. The strategies based on the Kelly criterion result in a generally higher median final wealth, while the contrary happens with the mean final wealth, that is higher on the MPT. For common practical purposes, the most suitable option out of the strategies reviewed seems to be the fractional Kelly, given that the fraction hyperparameter has been properly tuned to reflect the amount of uncertainty in each particular problem setting. This methos achieved the best performance when evaluated by the chosen metrics in most of their experiments while being comparatively simpler than the other strategies. The findings thus further support its common use in betting practice. The other common practice of setting a maximum bet limit was inconclusive as it improved the overall results in some cases while decreasing the profits in others. The distributional robust Kelly strategy then proved to be the safest in all of the experiments and can thus be suggested to extremely risk-averse investors. The second safest strategy was then to incorporate the drawdown constraint, which also proved quite efficient in trading of the security for profit.

To conclude, different ways of managing money have demonstrated to influence the final result of the investment in sports betting. Thus, a good strategy is as important as a good value detector. In a betting environment, the Kelly approach has been demonstrated to be the one that fits better, even if some restrictions, in line with the player's risk profile, are appropriate. Combining these two instruments, profitable arbitrage betting strategies have been proved to be implementable.

5 Personal value betting experiment

This research project has been inspired by the work made by Kaunitz et al. (2017), and it has been developed using Microsoft Excel. Its purpose is to applicate the value betting strategy, where the "value" is detected through the analysis of the odds offered in the market. In other words, the aim is to make a profit by exploiting the bookmakers' disagreement, as explained in section 4.3.2. To sum up, this strategy assumes that the average odds calculated from the market is as a good predictor of the actual probability of an event to occur. When a quote posted by a certain bookie is far over a certain threshold with respect to the average odds, betting on this odds has a positive expected value. Keeping betting using this rule, should lead to a profit in the medium-long run.

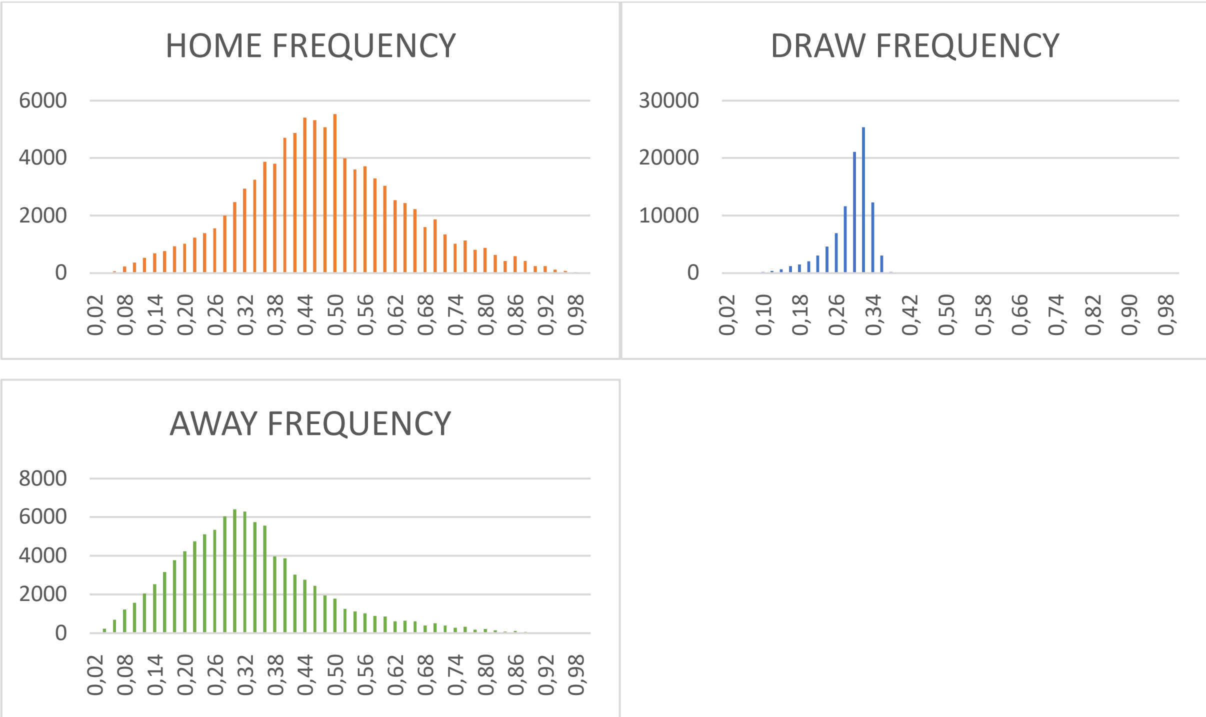
Data for this experiment were downloaded from *https://www.football-data.co.uk/data.php.* In total, the data set comprehends 94,122 football matches played between March 2012 and the end of December 2021. The championships considered are 32, coming from 27 different countries. In fact, for the principal European countries (England, Germany, France, Spain and Italy) also the second league has been included. The other championships are European, except for the Chinese, Japanese, Mexican and American ones. The odds collected for this study are the closing ones published right before the beginning of each match by more than 30 bookmakers. The most popular are: 'Interwetten', 'Bwin', 'bet-at-home', 'Unibet', 'Stan James', '10Bet', 'William Hill', 'bet365', 'Pinnacle Sports', 'Betsafe', 'Betway', '888sport', 'Ladbrokes', 'Betclic', 'Sportingbet', 'myBet' and 'Paddy Power'.

5.1 The strategy

The first step consists of computing the average odds among all the ones available for the analysis and isolating the highest one. The probability implicit in the average odds (P_B) is expressed using this formula: $P_B = {}^{1}/_{Avg(\theta)}$, where θ is a set containing the odds across bookmakers for a given event and a given game result (home team win, draw, away team win). This process must be repeated for each outcome of each match, thus resulting in a total of 282,366 probabilities (94,122*3).

Once I had retrieved the probability implicit in the market, I verified if this probability is indeed affordable, thus if it is a real good estimate of the chances of an outcome to occur, as the wisdom of the crowd effect would suggest. In order to do it, I binned all the data according to the consensus probability from 0 to 1 in steps of 0.02 (i.e. 50 bins). The first interesting thing to notice is the frequency distribution related to each type of outcome. Figure 15 summarizes them, where in the x axis are stored the probalities, while the y axis counts for the number of matches that belong to that bin.

Figure 15 Odds frequency distribution

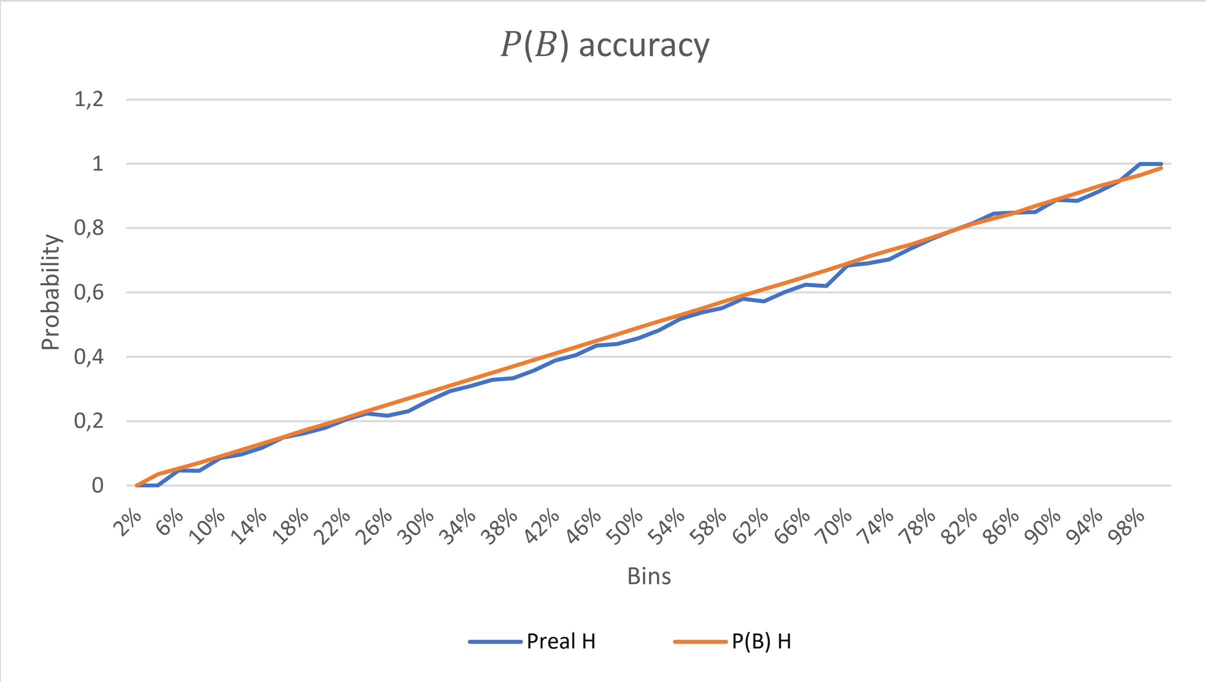


These distributions inspire some consideration: in general, the home victory is the most probable outcome, demonstrating that there is a home advantage effect. Moreover, the draw distribution appears very different with respect to the others. The odds are all concentrated around a probability of 30%. This indicates that the draw is almost never the favourite outcome, and its probability is often difficult to estimate.

Once I divided all the probabilities related to a specific outcome in each one of the 50 bins, I calculated the average of all the probabilities belonging to each group, naming them $P_{B(i)}$, where i stands for the group number (from 1 to 50). Then, I measured how many matches on a single group are indeed ended with the outcome in consideration. The most convenient outcome to take as example is the home victory, because it is well distributed across the groups, therefore we can test for the accuracy of every bin. In order to do it, I counted the number of matches ended with a home victory for every specific group. Dividing it by the total number of matches belonging to that group, I obtained the percentage of home victory, that I called $P_{real(i)}$.

To measure the $P_{B(i)}$' prediction accuracy, I firstly built a linear regression model, where P_B was the independent variable and P_{real} the dependent variable. Through an OLS estimation, I obtained the value of the intercept (-0.023) and the slope (1.015). Moreover, the R^2 was equal to 0.997. These results suggest that $P_{B(i)}$ is an extremely accurate proxy of the actual probability of occurrence of a home victory. Figure 16 confirms this finding:

Figure 16 P(B) accuracy



Knowing that the x axis represents the groups and the y axis the probability, the lines are drawn by connecting the points representing the probability obtained in each group from the method explained. The two lines are almost coincident. The orange one appears a little bit above P_{real} line. In fact, the linear regression has a negative intercept and the slope is positive, hence the P_{real} will be always lower that the probability implies in the odds. This is explained by the fact that the average odds used to calculate P_B are still including the margin, thus the implicit probability should be lowered by a small percentage. By doing it, the two lines would be even more similar. In short, the average of the odds offered in the market has demonstrating to be an affordable tool to estimate the real probability of an outcome to occur.

The idea behind the betting strategy is to exploit the implicit information contained in the bookmakers' aggregate odds (P_B) to systematically take advantage of mispriced events. In other words, bets are placed whenever a quote available in the market is higher enough than the average market odds, thus when its implied probability is lower than the P_B. If the maximum odds available is ω, the expected profit is:

$$E(\pi) = P_{real} * (\omega - 1) + (1 - P_{real}) * (-1) = P_{real} * \omega - 1$$

Assuming that the real probability is equal to the one implicit in the average market odds discounted by a constant value α that accounts for the margin included in the odds, thus:

$$P_{real} \cong P_B - \alpha$$

The updated expected profit becomes:

$$E(\pi) = (P_B - \alpha) * \omega - 1$$

The strategy rule establishes that a bet must be placed whenever the expected value is positive, hence where the expected profit $E(\pi)$ is more than zero, therefore when:

$$\omega > \frac{1}{(P_B - \alpha)}$$

The value α should be set as the intercept of our linear regression model. In this way we would minimize the difference between P_{real} and P_B. Nevertheless, the player can choose to set it arbitrarily. In fact, the α can be interpreted as a tolerance regulator: the expected value of each bet increases with it, while the number of bets executed is inversely proportional to it. Therefore, it can be modified according to the investor's risk tolerance, but, in general, it should not be lower than the intercept, otherwise it could signal misleading positive expected profit opportunities.

Given this betting rule, the next step is to organize the capital investment. In order to give a complete overview of the investment, I decided to evaluate the returns by supposing to begin with a certain capital, established in €1,000, and to bet a fixed percentage of it whenever the bet condition is satisfied. In accordance with chapter 4.3.2 regarding the money management, the percentage of capital wagered must not be higher than 2% of the total capital available. Together with the α parameter, the portion of capital bet is a tool to regulate the level of risk we are willing to bear. A bigger percentage invested increases with the risk of losing a wider part of the capital, together with an increased possible maximum return. In this way, the capital available is updated after every match, thus the amount wagered for each bet depends on the previous results, but it is always the same percentage of the capital available. After every match, the capital has three possibilities: it remains unchanged if the bet is not placed, it decreases by the fixed percentage if the bet is placed but lost, or, if the bet is placed and won, it increases following this formula:

$$C_{t+i} = (C_{t+i-1} * \beta) * (\omega_i - 1)$$

The first parenthesis indicates the amount bet, equal to the capital available C_{t+i-1} multiplied by the fixed percentage β, while the second one shows the maximum market odds minus 1, to avoid including the amount staked in the net profit. For example, if the maximum odds is 1.55, we could calculate the net profit just by multiplying the amount staked (K) for 0.55 (1.55 - 1), otherwise we should subtract the amount wagered (K) from the pay-out. Thus:

$$K * (1.55 - 1) = (K * 1.55) - K$$

5.2 Results

In my Excel worksheet, matches are ordered chronologically. Hence, it is possible to see how the capital is changing across the 10 years considered. Comparing the final capital with the initial one is not a proper way to make a complete evaluation of this strategy. To solve this problem, I have applied an evaluation scheme to all the three possible outcomes.

To begin, I have divided my sample monthly, from March 2012 to December 2021. Groups' size is different, as the number of matches varies among the year, according to the season or special championship's features. For example, July is always the month with the smallest number of matches. I decided to put together in a single group April, May and June 2020, because, due to the pandemic, only a small number of matches were played in that period. After this division, I calculated the monthly return for every month in the sample, supposing to begin with a capital of €1,000, by applying this formula:

$$R_{m(i)} = \frac{C_{m(i)}}{C_{m(i)-1}} - 1$$

where $C_{m(i)}$ is the capital at the end of the month (i), and $C_{m(i)-1}$ is the beginning capital available, which is the ending capital of the previous month. Once I obtained the monthly returns, I calculated the average monthly return. Later, in order to align it with the standard return representation, I annualized it using this formula:

$$R_A = [(1 + R_m)^{12}] - 1$$

As last step, I computed the standard deviation of each series of monthly returns, and, using it with the average monthly returns in the Excel normalization function, I plotted the normal distribution.

To correctly evaluate the results of this strategy, I made several simulations by changing the two risk tolerance regulators: α and β. Moreover, the results must be presented considering one outcome at a time.

The first one and the most interesting is the home victory. Table 2 summarizes the results for different values of α (columns) and β (rows). The first percentage on each cell represents the average of the monthly returns, while the second value is its annualized version.

Table 2 Home victory results

β \ α	0.01	0.02	0.03	0.04	0.05	0.06
0.50%	0.89% 11.25%	1.11% 14.14%	1.16% 14.83%	0.83% 10.43%	0.13% 1.62%	-0.16% -1.94%
1%	1.84% 24.39%	2.61% 36.29%	2.76% 35.58%	1.83% 24.38%	0.38% 4.63%	-0.22% -2.56%
2%	5.19% 83.53%	8.4% 163.14%	7.82% 146.72%	4.51% 69.76%	1.25% 16.09%	0.12% 1.39%
3%	14.53% 409.39%	23.25% 1128.76%	16.98% 556.44%	8.34% 161.63%	2.76% 38.64%	1.25% 16.02%

The best combinations appear to be the ones with a high percentage of capital invested and with a medium level of α, but it is not necessarily the case. In fact, a higher return corresponds to a higher level of risk, thus at the end the strategy could not be profitable. If at the beginning the returns are negative, the capital could be result eroded only after a few months, without having the possibility to recover. To give a final evaluation of this investment, it is necessary to have a complete overview of the economic results. Table 3 shows the final capital for every combination of α and β[21]. All the numbers inside the table must be intended in Euro (€).

Table 3 Final capital

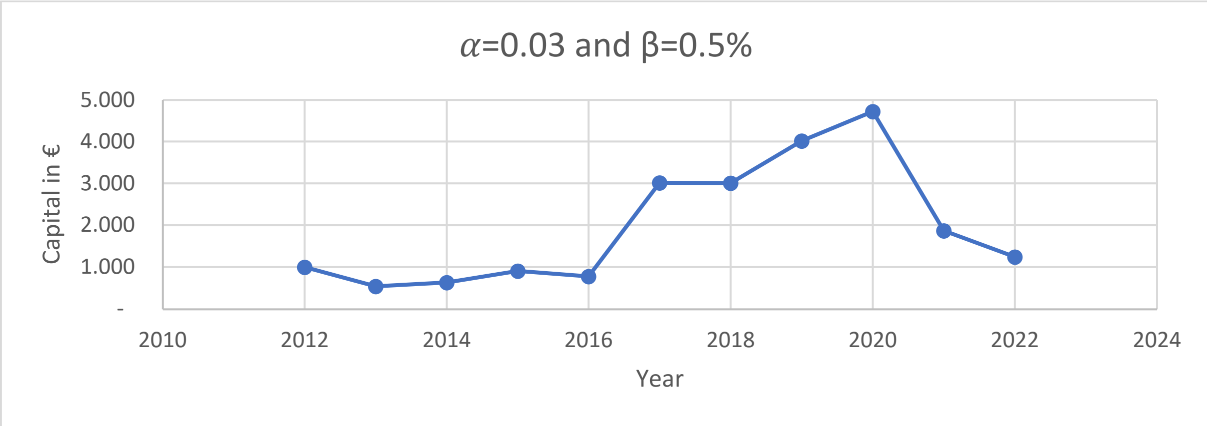
β \ α	0.01	0.02	0.03	0.04	0.05	0.06
0.50%	396.46	693.02	1,243.57	1,199.62	583.35	418.24
1%	3.03	24.87	261.18	368.52	97.92	51.10
2%	0	0	0.06	0.65	0.07	0.02
3%	0	0	0	0	0	0

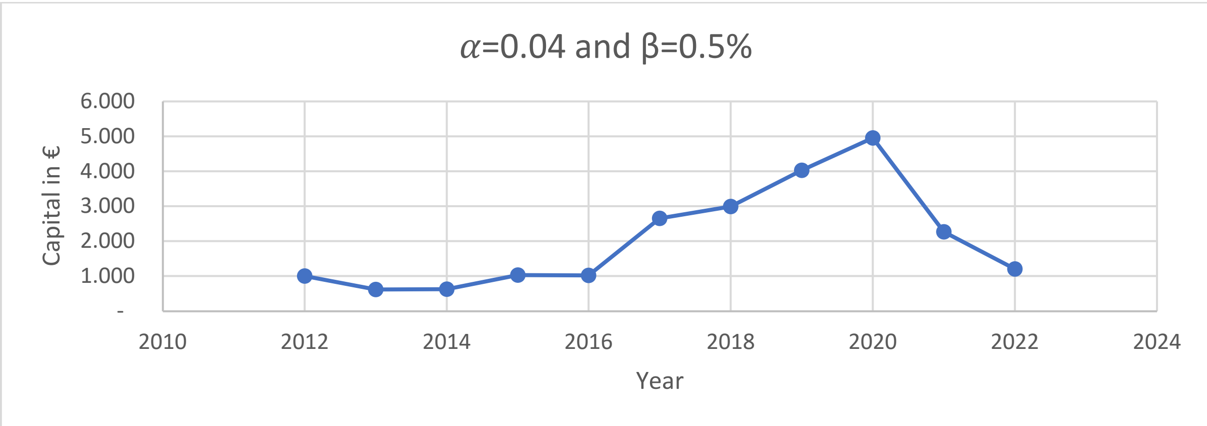
The best combination of α and β depends on the investor's risk aversion function. In this research case, the best strategy is to behave prudently. Apparently, a final profit can be reached only imposing α=0.03 or 0.04 and β=0.5%. To understand the investment properties, it is useful

[21] Remember that the starting capital was €1000.00.

to analyse the capital evolution. For example, with α=0.03/0.04 and β=0.5%, the capital evolution is:

Figure 17 Capital evolution





In both cases there is a huge drawdown after the end of 2019. In fact, before 2020 the initial capital was more than tripled. This point deserves a reflection, as 2020 is a very significant year that has generated big turbulences in the system due to the pandemic. The lockdown has deeply transformed the variables affecting a football match. For example, the line-ups were influenced by the impossibility to employ infected players, the matches were played without home supporters, in unusual periods of the year and with a remarkable higher frequency. The FIGC "Report 2021" has registered an unusual increase in the away victories of about 4%. Thus, not surprisingly the "home" bet has been deeply penalized. Therefore, given the variated conditions that have affected the outcomes and consequently our strategy, the investor should have paused his investment to prevent himself from these predictable losses, until a regular condition would have been re-established. In this case, thus before the beginning of the pandemic, the final capital in Euro would have been the following:

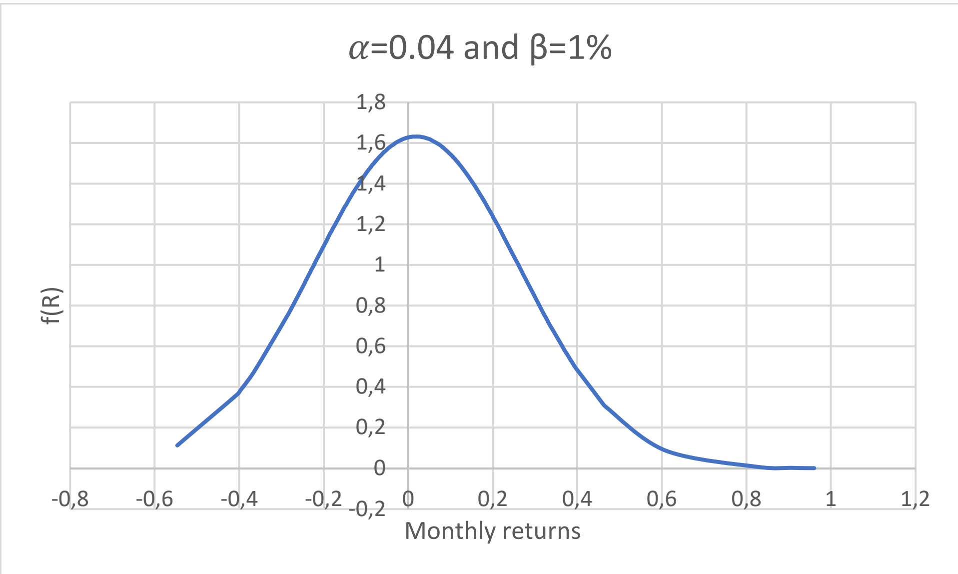
Table 4 Final result before 2020

β \ α	0.01	0.02	0.03	0.04	0.05	0.06
0.50%	3664.72	2967.00	4,722.65	4,953.49	2,424.11	1,742.69
1%	600.92	790.6	5265.87	8,363.97	2228.41	1,175.69
2%	0	0.05	95.24	1,025.43	110.8600	33.74
3%	0	0	0.01	2.13	0.14	0.03

From this point of view, the scenario looks differently. The strategy would present a positive result with 12 combinations of α and β. Again, the best results are obtained with low values of β, thus investing a small amount of the capital for each bet, in line with the money management theory presented in chapter 4.3.2. The best strategy would have been α=0.04 and β=1%, where we would have earned €7,363.97, while by investing high percentages of our portfolio we would have incurred in bankrupt after the pandemic, or even before with low values of tolerance.

Another representation of the characteristics of this investment can be displayed by plotting the normal distribution of the monthly returns. In fact, by computing their average and standard deviation, and using them in the Excel normal distribution function, it is possible to obtain the normal distribution value for each one of the monthly returns. Pooling all the points together gives the chart of the normal distribution in Figure 18:

Figure 18 Home monthly returns normal distribution



This type of representation can give a clear idea of the riskiness and possible profit of each investment. In this case, it seems worth to start the betting strategy.

Regarding to the draw, the results show negative returns with every combination of α and β. For example, keeping the previous prudent conditions (α=0.04 and β=1%), the average monthly return would be -17.66%. This result is strongly negative, but it is not a big surprise. In the previous literature, this outcome is always recognized as the most difficult to estimate, thus the evaluation of the expect profit of the investments on this final result could be misleading. Several papers do not even consider it while modelling a betting strategy, since it is the less affordable outcome. These results are in line with the conclusions of Pope and Peel (1989), who detected a low variance on draw odds, as displayed by the draw section of Figure 15. This might suggest that little attempt is made to adjust the odds set against draw outcomes to reflect differing perceptions about the probability of this outcome. Pope and Peel (p.328) wrote: "this behaviour could simply reflect a general inability to predict draw outcomes with any degree of reliability". In light of these considerations, using the average odds as predictor of the draw probability is not a good strategy, but, given the bookmakers less successful ability at predicting the draw, it could be a good opportunity to exploit using other instruments to forecast the real probability of a draw, such as a machine learning processes. For this type of research, the draw outcome has turned out to be not profitable, as expected, therefore it would have been more convenient to not consider it as an outcome to bet on.

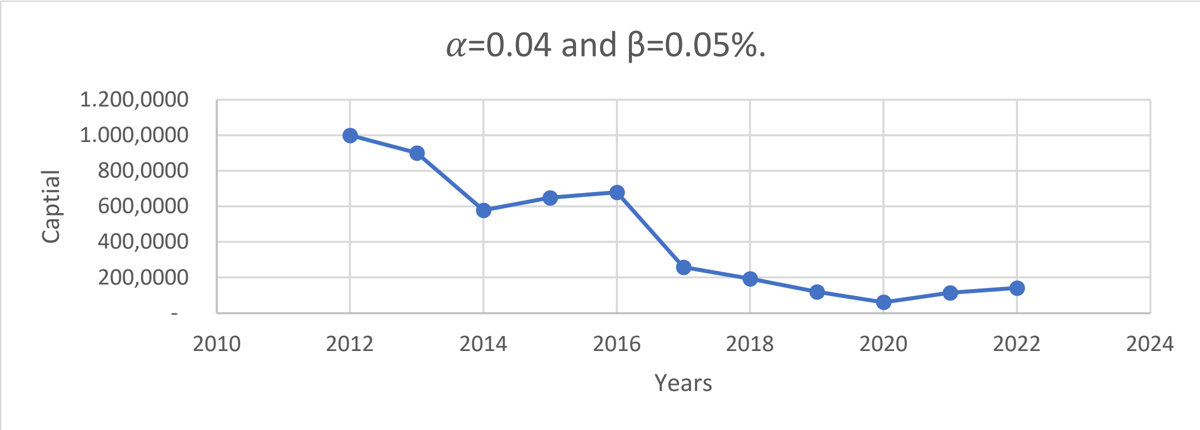
The last outcome to consider is the away win. The returns are embedded in Table 5, where the average monthly returns are above the annualized value for each set of α and β.

Table 5 Away victory results

β \ α	0.01	0.02	0.03	0.04	0.05	0.06
0.50%	-6.99% -58.11%	-3.67% -36.17%	-0.94% -10.68%	-0.78% -8.93%	-1.28% -14.35%	-1.98% -21.31%
1%	-12.91% -80.95%	-6.79% -56.99%	-2.26% -24.02	-1.89% -20.48%	-3.08% -31.4%	-4.61% -43.22%
2%	-22.11% -95.01%	-10.43% -73.34%	-6.43% -54.94%	-5.34% -48.25%	-8.00% -63.23%	-11.56% -77.11%
3%	-29.62 -95.84%	-8.56% -65.82%	-12.64% -80.25%	-10.35% -73.06%	-14.09% -83.83%	-20.07% -93.20%

In this case, the strategy does not seem to be profitable, since all the returns show a negative sign. The best combination of α and β is again a prudent one: with α = 0.04 and β = 0.05% the average monthly return is -0.78%. In this subcase, the capital evolution would be:

Figure 19 Capital evolution



The capital is eroded, and it never reaches a value higher than the beginning investment. By analysing the capital evolution in every combination, I noticed that the only β that is able to save a portion of the capital after 10 years of betting is β=0.05%; with other values, the capital results entirely wasted. Due to the terrible results in the first year, in some cases it is lost very quickly.

Interestingly, there is a mirror effect in 2016 and in 2020. The strongly positive returns of the home victory are reflected in a negative period for the away victory. The same phenomenon, but with an opposite direction, happens from 2020 to the end. The strongly negative home victory returns are reflected in a good period for the away victory, where the average monthly returns with the pandemic are +4.42%. Anyway, in both case the magnitude of this mirror effect is lower than the home returns.

5.3 Final considerations

After all these results, we can make some considerations about how to efficiently run this strategy. The first conclusion is that the draw in not a good outcome to bet on, hence it should not be considered. This assertion is in line with several papers presented in this dissertation, where the draw odds were found to reflect misleading probabilities rather often. The only outcome that seems to be profitable is the home win, but this is true only before the pandemic. The opposite happens with the away victory. Thus, in general, betting on the home victory is the most favourable option. There is like a favourite-longshot bias similar to the one detected by Deschamps and Gergaud (2007). In fact, home odds are the most convenient ones, and betting on them gives back abnormal returns in the medium-long run. Further, the number of betting opportunities are higher in the home victory, confirming that it is the most attractive outcome. With α=0.04, the home win presents 33,750 positive expected returns bets, while the away outcome 30,480 opportunities. Considering the period before 2020, betting without being too risked would have returned good results. As suggested in the results section, a smart investor should have had recognized that the changed conditions were more favourable for the away teams, thus from that moment he should have moved the strategy to the visitors clubs, increasing his final profit.

In this betting strategy there is systemic risk, attributable to the pandemic, but the advantage is that it is possible to suspend the betting activity whenever a distortion factor comes in, and maybe to exploit its consequences.

Another interesting conclusion is retrieved by analysing championship by championship. In short, the highest returns come from the less important and followed championships, like the North European ones. In this case, odds are more discordant because just a few numbers of information are available to compute the odds, therefore it turns in more valuable opportunities for the players.

This study project has the advantage of being simple and easily implementable. The probabilities are collected from the average market odds, and not from sophisticated machine learning techniques. Thus, everyone can build his strategy under these assumptions. The main defect is its limited ability to detect opportunities, because odds are collected only in the instant that precedes the beginning of a match. A continuous supervision of the odds movements would increase the number and the quality of our bets. Moreover, the number of bets in this study are not possible to be executed by only one investor. Thus, the player should elaborate an automized system that place his bets for him, or he should choose only the most valuable bets. The last problem is that bookmakers could limit the accounts, thus limiting our strategy. A possible solution is to include among the bookmakers analysed the exchange platforms, that does not limit the players' possibility to bet. As a whole, excluding the draws, the strategy would have been profitable before the pandemic, where it would have been appropriate to pause the strategy on home win and start betting on the away side. To conclude, this research has demonstrated that it is possible to profit from a value betting strategy like this one, and his maximization depends on the player's attitude to risk and his ability to interpret new external factors that could affect the theoretical basement of this strategy.

Conclusion

This dissertation has analysed the principal characteristics of the sport betting market, presenting where and how it is possible to invest on it.

The gambling world has the right to be recognized as a real economic sector, for its numbers and the characters that operate in it, that are big companies, often listed, and a wide group of players. In the last two decades, this environment has known a great period of development, mainly thanks to the technological revolution. The rising of the exchange platforms has instituted a parallel market to the usual fixed-odds one, creating new investment opportunities and a new starting point for innovative areas of research. These are the reasons why this theme has been widely debated recently.

Besides all these considerations, the principal purpose of this thesis is to propose the sports betting as a new asset class to consciously invest on. In fact, among the vastity of players that bets just for fun, only the 2% of them results in a profit. Thus, after an explanation of the functioning and a detailed collection of the scientific literature about the efficiency analysis of this market, this study work has presented how it is possible to rationally invest, and not gamble, on this asset so far. The first activity is the sport trading, where the player can speculate on the odds movements, making a profit by buying and selling them during the course of a match or even before its start. The second implementable activity is the arbitrage betting, divided in three main options: the matched betting, the sure bet and the value betting. The last one is the most interesting, because of its theoretical background and because it does not necessarily require a direct effort from the investor, since it can be automated. The key factor in this strategy is to detect mispriced odds offered in the market. This process is not unique, in the sense that there are several ways to estimate the real probability of an event to occur. The most accurate system is the machine learning approach, where thousands of data are pooled to obtain an affordable representation of the future probabilities of an event. On the other hand, this is by far the most complicated approach, because it requires thousands of data and advanced informatic skills. The alternatives assume as an efficient predictor the exchange market, or the analysis of past results, or the probability implied by the average of the odds offered in the market.

The last process has been used to conduct the research project presented in chapter 5. Its advantage is the simplicity of its mechanism. This work has confirmed that in the medium-long run this strategy can be profitable, excluding the unreliable Covid period, but also very volatile.

As underlined in the paper by Thukral and Eleuterio (2016), sports trading is a perfect diversification tool, since it is completely uncorrelated with any other type of assets, except for the listed teams and companies. Even the matches inside the sports portfolio are mutually independent. They showed that a simple betting strategy based on consistently laying the 4 favourite horses in the UK from 2010 to 2016 can outperform hedge fund managers and the S&P 500, achieving excess returns with low volatility. As expected, these returns are uncorrelated with the S&P 500.

By now, the sport betting market is still known as source of entertainment more than an asset class, but it is constantly growing and evolving, therefore in the next future it can become a great diversification tool and a popular asset to invest on. So far, the connected ethical issues and the lack of a common and shared regulation are slowing down its development. Since each country has its own rules and taxes, the betting markets are always operating nationally and they are closed to foreign players, therefore, terms and conditions vary from country to country.

Gomber et al. (2008) have proposed three development scenarios for the sports betting market, in order to make it more accessible and reliable, and to overcome the current geographic limitations. The first proposal is to design sports bets as financial instruments on a regulated market that guarantees for market integrity and investor protection. In this retail-to-intermediary market organisation, the new asset class should instantly benefit from the developed environment of a regulated market, also by lowering transaction costs through economies of scale in transaction services. The second option is an intermediary-to-intermediary market development, where an inter-bookmaker platform provides trading services to exchange position risks between wholesale market participants on an international basis and enables risk hedging and diversification. In other words, they propose to integrate all the geographically fragmented markets into a wholesale international market. Both bookies and retail clients would benefit from this new setting, because the formers could reach new counterparties and additional business opportunities would arise, while the market participants would have a more liquid and price competitive environment. The last idea consists in instituting a central counterparty (CCP) clearing service, which would establish multilateral risk management capabilities and reduce systemic risk. Moreover, it would improve the operational efficiency through a centralization and standardization process and increase the market liquidity.

In general, I believe that to effectively become a new accessible and convenient asset to invest on, the sports betting market needs a decisive impulse through two main innovations. Firstly, it must experience a process of internationalisation, that would make it more competitive and that would harmonize the regulation, at least at a continental level. Secondly, the differences with the financial value chain should be reduced, by developing a product design that would transfer the new asset class to a financial market environment, to benefit from established distribution channels, infrastructures, and investors protection.

To conclude, sports betting has demonstrated to deserve the attention of the economic and technological community, both for its theoretical fundamentals and for its investment opportunities. In the next decades, it could become a widely implemented diversification tool and its evolution presents a fascinating field for future research.

Bibliography

Agosto A., Cavaliere G., Kristensenc D. and Rahbek A. (2016), *Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX),* Journal of Empirical Finance, Volume 38, p.640–663. Available at: http://dx.doi.org/10.1016/j.jempfin.2016.02.007

Alberola J., García-Fornes A. and Espinosa A. (2012), *Using a case-based reasoning approach for trading in sports betting markets*, Appl Intell, Volume 38, p.465–477. Available at: https://doi.org/10.1007/s10489-012-0381-9

Alberola J., García-Fornes A. and Espinosa A. (2010), *Price Prediction in Sports Betting Markets*, MATES 2010, Leipzig, Germany (2010). Available at: http://dx.doi.org/10.1007/978-3-642-16178-0_19

Angelini G. and De Angelis L. (2018), *Efficiency of online football betting markets*, International Journal of Forecasting, Volume 35 (2), p. 712-721. Available at: https://doi.org/10.1016/j.ijforecast.2018.07.008

Angelini G., De Angelis L. (2016), *PARX model for football matches predictions,* Research paper (2), Università di Bologna, Dipartimento di Scienze Statistiche. Available at: http://dx.doi.org/10.13140/RG.2.2.27404.72325

Ashiya M. (2015), *Lock! Risk-Free Arbitrage in the Japanese Racetrack Betting Market,* Journal of Sports Economics, Volume 16 (3), p.322-330. Available at: http://www.lib.kobe-u.ac.jp/handle_kernel/90003552

Banks J. (2013), *Edging your bets: advantage play, gambling, crime and victimisation*, Crime, Media, Culture, Volume 9 (2), p.171-187. Available at: http://orcid.org/0000-0002-1899-9057

Beer F. and Lin F. (2019), *Sports Sentiment and Stock Returns: The Bombay Stock Exchange*, ACRN Journal of Finance and Risk Perspectives, Volume 8, p.56-70. Available at: https://doi.org/10.35944/jofrp.2019.8.1.003

Brown A. and Yang F. (2014), *The Role of Speculative Trade in Market Efficiency: Evidence from a Betting Exchange*, Review of Finance, European Finance Association, Volume 21 (2), p.583-603. Available at: https://doi.org/10.1093/ROF%2FRFW027

Buchdahl J. (2003), *Fixed Odds Sports Betting: The Essential Guide*, In: *Rating System for Fixed Odds Football Match Prediction*, Football-Data.co.uk (2003). Available at: https://football-data.co.uk/ratings.pdf

Bunker R.P. and Thabtah F. (2017), *A machine learning framework for sport result prediction*, Applied Computing and Informatics, Volume 15 (1), p.27-33. Available at: https://doi.org/10.1016/j.aci.2017.09.005

Buursma D. (2010), *Predicting sports events from past results*, 14thTwente Student Conference on IT, 2011, Enschede, The Netherlands.

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

Castellani M., Pattitoni P. and Patuelli R. (2013), *Abnormal Returns of Soccer Teams: Reassessing the Informational Value of Betting Odds,* Journal of Sports Economics, Volume 16 (7), p.735-759. Available at: https://doi.org/10.1177/1527002513505285

Choi D. and Hui S. K. (2014), *The role of surprise: Understanding overreaction and underreaction to unanticipated events using in-play soccer betting market*, Journal of Economic Behavior & Organization, Volume 107, p.614-629. Available at: http://dx.doi.org/10.1016/j.jebo.2014.02.009

Constantinou, A. C. and Fenton, Norman E. (2013), *Profiting from arbitrage and odds biases of the European football gambling market*, The Journal of Gambling Business and Economics, Volume 7 (2), p.41-70. Available at: http://www.ubplj.org/index.php/jgbe/article/view/630

Croxson K. and Reade J. J. (2014), *Exchange vs. dealers: a high-frequency analysis of in-play betting prices*. Discussion Papers 11-19, University of Birmingham, Department of Economics.

Croxson K. and Reade J.J. (2012), *Information and Efficiency: Goal Arrival in Soccer Betting,* The Economic Journal, Volume 124 (575). Available at: http://dx.doi.org/10.1111/ecoj.12033

Daunhawer I., Schoch D. and Kosub S. (2017), *Biases in the Football Betting Market*, SSRN Electronic Journal. Available at: http://dx.doi.org/10.2139/ssrn.2977118

Deloitte Italy (2019), *Online Gaming, Il settore del gioco online: confronto internazionale e prospettive*, Deloitte Creative Team.

Deloitte UK (2021), *Testing times Football Money League,* Deloitte Sports Business Group.

Deschamps B. and Gergaud O. (2007), *Efficiency in Betting Markets: Evidence From English Football,* The Journal of Prediction Markets, Volume (1), p.61–73. Available at: http://dx.doi.org/10.5750/jpm.v1i1.420

Dixon M. and Cole S. (2002), *Modelling association football scores and inefficiencies in the UK football betting market,* Journal of the Royal Statistical Society, Applied Statistics, Volume 46 (2), p.265-280. Available at: https://www.jstor.org/stable/2986290

Dixon M.J. and Pope P.F. (2004), *The value of statistical forecasts in the UK association football betting market,* International Journal of Forecasting, Volume 20, p.697 – 711. Available at: http://dx.doi.org/10.1016/j.ijforecast.2003.12.007

Dzalbs I., Kalganova T. (2019), *Forecasting price movements in betting exchanges using Cartesian Genetic Programming and ANN*, Big Data Research, Volume 14, p.112-120. Available at: https://doi.org/10.1016/j.bdr.2018.10.001

Edmans A., García D. and Norli Ø, (2007), *Sports Sentiment and Stock Returns,* The Journal of Finance, Volume 62 (4), p.1967-1998. Available at: http://dx.doi.org/10.1111/j.1540-6261.2007.01262.x

Elaad G., Reade J.J. and Singleton C. (2019), *Information, prices and efficiency in an online betting market,* Finance Research Letters. Available at: https://doi.org/10.1016/j.frl.2019.09.006

Fan Q., Lei W. and Zhang X. (2017), *The impact of sports sentiment on stock returns: A case study from professional sports leagues*, IEEE Global Conference on Signal and Information Processing, 2017. Available at: http://dx.doi.org/10.1109/GlobalSIP.2017.8309094

FIGC (2021), *Report Calcio 2021*. La Federazione, area istituzionale. Available at: https://www.figc.it/it/federazione/federazione-trasparente/reportcalcio/

Forrest D. and Simmons R. (2008), *Sentiment in the betting market on Spanish football, Applied Economics*, Volume 40, p.119–126. Available at: http://dx.doi.org/10.1080/00036840701522895

Forrest D., Goddard J. & Simmons R. (2005), *Odds-setters as forecasters: the case of English football,* International Journal of Forecasting, Volume 21, p.551–564.

Franck E., Verbeek E. and Nuesch S. (2010), *Prediction accuracy of different market structures, bookmakers versus a betting exchange,* International Journal of Forecasting, Volume 26, p.448–459. Available at: http://dx.doi.org/10.1016/j.ijforecast.2010.01.004

Franck E., Verbeek E. and Nuesch. S. (2010), *Sentimental preferences and the organizational regime of betting markets*, Southern Economic Journal, Volume 78(2), p.502-518. Available at: https://doi.org/10.4284/0038-4038-78.2.502

Frank E., Verbeek E. and Nuesch S. (2012), *Inter-market Arbitrage in Betting*, Economica, Volume 80, p.300–325. Available at: https://doi:10.1111/ecca.12009

Goddard J. and Asimakopoulos I. (2004), *Forecasting Football Results and the Efficiency of Fixed-odds Betting*, Journal of Forecasting, Volume 23, p.51–66. Available at: https://doi.org/10.1002/for.877

Gomber P., Rohr P. and Schweickert U. (2008), *Sports betting as a new asset class—current market organization and options for development*, Swiss Society for Financial Market Research. Available at: http://ssrn.com/abstract=2794937

Gonçalves R., Ribeiroc V. M., Pereiraa F.L. and Rocha A.P. (2019), *Deep learning in exchange markets,* Elsevier, Information Economics and Policy, Volume 47, p.38-51. Available at: https://doi.org/10.1016/j.infoecopol.2019.05.002

Hubacek O., Sourek G. and Zelezny F. (2019), *Exploiting sports-betting market using machine learning,* International Journal of Forecasting, Volume 35, p.783-796. Available at: https://doi.org/10.1016/j.ijforecast.2019.01.001

Ioulianou M., Christofi I. et al. (2011), *Automatic Identification of Low-Risk Trading Opportunities in Sport Betting Exchange Markets*, University of Macedonia, Department of Applied Informatics, Greece.

Kaunitz L., Zhong S. and Kreiner J. (2017), *Beating the bookies with their own numbers - and how the online sports betting market is rigged, s.n. Available at: https://github.com/Lisandro79/BeatTheBookie*.

Klein C., Zwergel B. and Heiden S. (2009), *On the existence of sports sentiment: the relation between football match results and stock index returns in Europe*. Review of Managerial Science, Volume3 (3), p.191-208. Available at: http://dx.doi.org/10.1007/s11846-009-0031-8

Knoll J. and Stübinger J. (2019), *Machine-Learning-Based Statistical Arbitrage Football Betting,* Künstliche Intelligenz*,* Volume 34, p.69–80. Available at: https://doi.org/10.1007/s13218-019-00610-4

Kotlyar V. Y. and Smyrnova O. V. (2012), *Betting market: Analysis of arbitrage situations*, Cybernetics and Systems Analysis, Vol. 48 (6), p. 912–920. Available at: https://doi.org/110.1007/s10559-012-9472-1

Kuypers T. (2000), *Information and efficiency: an empirical study of a fixed odds betting market,* Applied Economics, Volume 32 (11), p.1353-1363. Available at: https://doi.org/10.1080/00036840050151449

Landi G. (2015), *Betting Exchange: la rivoluzione del Trading Sportivo*, Betting Exchange, p.1-26.

Lazzarino M., Berrill J. and Šević A. (2018), *What Is Statistical Arbitrage? Theoretical* Economics Letters, Volume 8, p.888-908. Available at: http://www.scirp.org/journal/tel

Lee M. and Smith G. (2002), *Regression to the mean and football wagers*, Journal of Behavioral Decision Making, Volume 15 (4), p.329-342. Available at: http://dx.doi.org/10.1002/bdm.418

Levitt S. D. (2004), *Why are Gambling Markets Organized So Differently from Financial Markets,* Luckner S., Schröder J. and Slamka C. (2006), *On the Forecast Accuracy of Sports Prediction Markets,* The Economic Journal, Volume 114, p.223–246. Available at: https://doi.org/10.1111/j.1468-0297.2004.00207.x

Luckner S., Schröder J. and Slamka C. (2006), *On the Forecast Accuracy of Sports Prediction Markets,* International seminar: Negotiation, Auctions, and Market Engineering, 2006, Dagstuhl Castle, Germany, p.227-234. Available at: http://dx.doi.org/10.1007/978-3-540-77554-6_17

Maher M. J. (1982), *Modelling association football scores*, Journal of the Royal Statistical Society, Applied Statistics, Volume 46 (2), pp. 265-280. Available at: https://www.jstor.org/stable/2986290

Mangold B. and Stübinger J. (2020), *Investigating inefficiencies of bookmaker odds in football using machine learning,* 3rd International Conference on Advanced Research Methods and Analytics, 2020, Valencia, p.173-179. Available at: http://dx.doi.org/10.4995/CARMA2020.2020.11619

Marttinen N., (2001), *Creating a Profitable Betting Strategy for Football by Using Statistical Modelling*, Master Thesis, Trinity College Dublin, Department of Statistics, p.65-71.

Uhrín M., Šourek G., Hubácek O. and Železný F. (2021), *Optimal sports betting strategies in practice: an experimental review,* Journal of Management Mathematics, Volume 32, p.465–489. Available at: https://doi.org/10.1093/imaman/dpaa029

Na S., Su Y. and Kunkel T. (2018*), Do not bet on your favourite football team: the influence of fan identity-based biases and sport context knowledge on game prediction accuracy,* European Sport Management Quarterly, Volume 0, p.1–23. Available at: https://doi.org/10.1080/16184742.2018.1530689

Nielsen E. and Sandøy S. (2019), *Profiting from Football Betting using Artificial Neural Networks,* Master Thesis, Norwegian University of Science and Technology, Department of Computer Science.

Nordsted P. (2009), *Mastering Betfair*, Harriman House.

Occhipinti D. (2016), *Betting exchange: a market maker process*, Master Thesis, LUISS University, Department of Economics and Finance.

Oikonomidis A. (2013), *Weak form efficiency and pricing dynamics in a competitive globalised market setting*, PhD Thesis, University of Southampton, Faculty of Business and Law.

Ottaviani M. and Sørensen P.N. (2007), *The Favorite-Longshot Bias: An Overview of the Main Explanations*, Handbooks in Finance, Volume 6, p.83-101. Available at: https://doi.org/10.1016/B978-044450744-0.50009-3

Palomino F., Renneboog L. and Zhang C. (2005), *Stock price reactions to short-lived public information: the case of betting odds,* CentER Discussion Paper, Volume 62. Available at: https://research.tilburguniversity.edu/en/publications/059428e3-2ed6-42e2-8d3c-2cf48caf0ec6

Paton D. and Williams L.V. (1988), *Do betting costs explain betting biases?* Applied Economics Letters, Volume 5, p.333-335. Available at: https://doi.org/10.1080/758524413

Pitt L.F., Watson R.T., and Shapiro D.M. (2005), *www.Betfair.com: World Wide Wagering,* Communications of the Association for Information Systems, Volume 15, p.149-161.

Pope P.F. and Peel D.A. (1989), *Information, Prices and Efficiency in a Fixed-Odds Betting Market,* Economica, Volume 56 (223), p.323-341. Available at: https://www.jstor.org/stable/2554281

Rajkovic I. (2013), *Applied Financial Time Series Modelling: Forming a betting strategy based on the identifiable anomalies of European football fixed-odds betting markets*, Master Thesis, University of Durham, Department of Economics, p.1-21.

Renneboog L.D.R. and Vanbrabant P. (2000), *Share Price Reactions to Sporty Performances of Soccer Clubs listed on the London Stock Exchange and the AIM*, CentER Discussion Paper; Vol.19. Available at: https://research.tilburguniversity.edu/en/publications/0907f95d-bfed-4fe9-a13d-fd52cf07ea86

Scholtens B. (2009), *Scoring on the stock exchange? The effect of football matches on stock market returns: an event study*, Applied Economics, Volume 41 (25), p.3231-3237. Available at: https://hal.archives-ouvertes.fr/hal-00582192

Smith J.S. (2021), *Bookies who lose*, Business life. Available at: https://www.betasia.com/how-to-win/459822/bookies-who-lose

Smith M. A., Paton D. and Williams L. V. (2006), *Market Efficiency in Person-to-Person Betting,* Economica, Volume 73, p.673–689. Available at: https://doi.org/doi:10.1111/j.1468-0335.2006.00518.x

Smith M. A., Paton D. and Williams L. V. (2009), *Do Bookmakers Possess Superior Skills to Bettors in Predicting Outcomes?* Journal of Economic Behavior and Organization, Volume 71 (2), p.539. Available at: https://hal.archives-ouvertes.fr/hal-00684229

Snowberg E. and Wolfers J. (2010), *Explaining the Favorite-Longshot Bias: Is it Risk-Love or Misperceptions?* Journal of Political Economy, Volume 118 (4), p.723-746. Available at: https://doi.org/10.1086/655844

Spann M. and Skiera B. (2009), *Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters,* Journal of Forecasting, Volume 28, p.55–72. Available at: https://doi.org/10.1002/for.1091

Spinosa C.L. (2014), *Testing the Efficiency of the NFL Point Spread Betting Market*, CMC Senior Thesis, Claremont McKenna College. Available at: http://scholarship.claremont.edu/cmc_theses/986

Stadtmann G. (2006), *Frequent news and pure signals: the case of a publicly traded football club,* Scottish Journal of Political Economy. Available at: https://ssrn.com/abstract=839104

Stekler H.O., Sendor D. and Verlander R. (2010) *Issues in sports forecasting*, International Journal of Forecasting, Volume 26, p.606–621. Available at: http://dx.doi.org/10.1016/j.ijforecast.2010.01.003

Stübinger J. and Knoll J. (2018), *Beat the Bookmaker – Winning Football Bets with Machine Learning,* Artificial Intelligence XXXV, p.219-233. Available at: https://doi.org/10.1007/978-3-030-04191-5_21

Stübinger J. and Knoll J. (2019), *Machine-Learning-Based Statistical Arbitrage Football Betting*, Künstliche Intelligenz, Volume 34, p.69–80. Available at: https://doi.org/10.1007/s13218-019-00610-4

Stübinger J., Mangold B. and Knoll J. (2019), *Machine Learning in Football Betting: Prediction of Match Results Based on Player Characteristics,* Applied Science, Volume 10 (46). Available at: http://dx.doi.org/10.3390/app10010046

Thaler R. H. and Ziemba W. (1988), *Anomalies Parimutuel Betting Markets: Racetracks and Lotteries,* Journal of Economic Perspectives, Volume 2 (2), p.161–174. Available at: https://doi.org/10.1257/jep.2.2.161

Thukral L. and Eleuterio P.V. (2016), *Sports betting as a new asset class: Can a sports trader beat hedge fund managers from 2010 - 2016?* SSRN Electronic Journal. Available at: https://dx.doi.org/10.2139/ssrn.2794937

Tiitu T. (2016), *Abnormal returns in an efficient market? Statistical and economic weak form efficiency of online sports betting in European soccer,* Master Thesis, Aalto University, Department of Finance.

Uhrìn M., Šourek G., Hubácek O. and Železný F. (2021), *Optimal sports betting strategies in practice: an experimental review,* Journal of Management Mathematics, Volume (32), p.465–489. Available at: https://doi.org/10.1093/imaman/dpaa029

Ulmer B., Fernandez M. (2015), *Predicting Soccer Match Results in the English Premier League,* Computer science.

Verkerk B. (2018), *Beating the bookie in the football betting market without using extensive match data*, Research Paper, Amsterdam University.

Vlastakis N., Dotsis G. and Markellos R. N. (2009), *How Efficient is the European Football Betting Market? Evidence from Arbitrage and Trading Strategies,* Journal of Forecasting, Volume 28, p.426–444. Available at: https://doi.org/10.1002/for.1085

Williams R.J., Rehm J. and Stevens R. (2011), *The Social and Economic Impacts of Gambling*. Final Report prepared for the Canadian Consortium for Gambling Research, p.35-51.

Websites

https://www.ilsole24ore.com/art/slot-e-scommesse-2020-l-erario-perde-45-miliardi-e-il-lockdown-cresce-gioco-illegale-ADx8YiBB

https://igamingbusiness.com/italy-fy-revenue-top-10-operators-product/

https://corporatefinanceinstitute.com/resources/knowledge/finance/gross-gaming-revenue-ggr/

https://mercurius.io/en/trader-app/performance

https://www.statista.com/statistics/1186231/casino-and-online-gambling-industry-market-size-global/

https://www.football-data.co.uk/data.php

https://en.wikipedia.org/wiki/Convolutional_neural_network