Louvain School of Management

THE ROLE OF PROFESSIONAL FOOTBALL AS A TRIGGER OF IRRATIONAL INVESTOR BEHAVIOR IN CAPITAL MARKETS

Author(s): Marc Massuet i Ardil
Supervisor(s): Dr. Leonardo Iania
Academic year 2022-2023
International Management
Behavioral Economics

ABSTRACT

The purpose of this thesis was to extend the empirical knowledge about investors' irrational behavior, and to do so by putting professional sports and football at the center of my research. During the first phase of this work, I reviewed literature that showed evidence of instances where investors may behave against the economic rationality principle and where investor sentiment may be the driver of behavior.

My objective, therefore, was to observe this behavior in an empiric case. I proceeded to retrieve data of professional football matches in the 2012-2022 period from nine different traded clubs and six different national leagues. I hypothesized that several variables related to these matches could have an effect on investor behavior and, ultimately, be translated into stock price fluctuation. In my research, I defined several independent variables that I hypothesized could be statistically significant: the result of the match, the day of the week where the match took place, the location where the match was held, the singularity of the match and the national league of the teams.

For the statistical analysis, I defined a multiple regression model based on the results of similar studies from authors reviewed in the theoretical block of this thesis. My results showed that in some instances wins and losses could have a positive and negative effect respectively on stock prices. However, the model showed very low robustness and I concluded that I was not able to reject the Null Hypothesis based on my results. Further research would be needed and I suggest several variables that could potentially improve the model.

Key Words

irrational behavior; stock prices; sports; football; behavioral biases; investor sentiment; reactiveness to news

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve
Boulevard Emile Devreux 6, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique
www.uclouvain.be/lsm

Chapter	Title	Page
	Abstract	3
1.	**INTRODUCTION**	**4**
	I. THEORETICAL PART	**5**
2.	**INVESTOR SENTIMENT**	**5**
2.1.	Reactiveness to news	6
2.2.	Loss effect	7
3.	**PROFESSIONAL SPORTS AS A TRIGGER**	**8**
3.1.	Behavioral biases triggered by single events	8
3.2.	Investor sentiment and reactiveness to sports results	9
4.	**IMPLICATIONS AND RELEVANCE OF LITERATURE**	**10**
	II. EMPIRICAL PART	**12**
5.	**APPROACH, OBJECTIVES AND HYPOTHESIS**	12
5.1.	Generic Hypothesis	12
5.2.	Specific Hypothesis	13
6.	**METHODOLOGY AND DATA**	13
6.1.	Material	13
6.2.	Variables in the study	14
6.2.1.	Independent variables	14
6.2.2.	Dependent variables	14
6.3.	The model	14
6.4.	General and Statistical Procedures	16
6.4.1.	Data sourcing	16

6.4.2.	Data cleansing	19
6.4.3.	Data treatment and categorization	19
6.4.4.	Data aggregation and final data set	22
6.5.	Notation	24
7.	**RESULTS**	25
7.1.	Case description	25
7.2.	Initial analysis	26
7.2.1	Clubs analysis	26
7.2.2	Aggregated analysis	30
7.3.	Comparison and analysis of independent and dependent variables	32
7.4.	Effects observed	33
8.	**CONCLUSIONS**	35
8.1.	Review of objectives	35
8.2.	Review of results	36
8.3.	Limitations and next steps	37
9.	**REFERENCES AND BIBLIOGRAPHY**	40
9.1.	Literature and References	40
9.2.	Database sources	41
9.3.	Data analysis tools	41

1. INTRODUCTION

One of the most repeated mantras among Economic students attending their first Microeconomics courses is the assumption that individuals behave in a rational manner when making economically driven decisions. Only those lucky students that are able to reach the latest stages of their degree are able to discover how this economic rationality principle, which is a very useful simplification for academic and pedagogic purposes, does not translate completely to reality.

Many behavioral economists and psychologists have studied phenomena that suggests that, in occasions, individuals may undertake behavior considered irrational within the context of an economic decision. Kahneman and Tversky (1974) and Ariely (2010), for example, provided many insights on customers' decision-making processes and rationality through several experiments.

The purpose of this research project is to extend the empirical knowledge about individual's irrational behavior, putting our focus on investors, specifically investors of traded stock. In order to do so, we aim to identify irrational investing behavior translated into fluctuation of prices of stock for several firms. The differentiating element of this work, however, is the scope taken to do so: sport, in particular professional football. We expect our analysis to provide more insights on what role play professional sports, especially match results, towards the valuation of the stock of the teams.

This work has been structured in two defined blocks: a theoretical and an empirical block. The first block's objective is to revise existing literature on stock reactiveness to news in order to set the foundations of the study and to learn more about what should be considered rational behavior and what diverges from expected behavior. Furthermore, we aim to dive deeper into existing studies and literature focusing on sports and its hypothesized relation towards the capital markets. The second block of this project constitutes a statistical analysis, performed with several econometric tools, in order to identify potential relationships between the variables we have identified and, ultimately, to give and answer to the hypothesis that we formulated.

However, one of the main limitations of this study, which needs to be highlighted, is the impossibility to assess the existence of a causal relationship between our two main variables: stock prices and professional sport matches results. Due to the character of this work as a master thesis and the time, content and extension constraints that it carries, the goal of this research will be limited to the identification of correlation between several variables.

I. THEORETICAL BLOCK: LITERATURE REVIEW

My objective in this theoretical block is to review literature addressing certain phenomena that can potentially have an effect towards asset pricing. My goal is, ultimately, to better identify what can constitute rational and irrational behavior and some stances of over and under reacting. Finally, we will dive further into the role of a specific industry, professional football, and how it potentially relates to the stock market.

2. INVESTOR SENTIMENT

Stock prices, and the capital markets in general, are an expression of supply and demand. If we assume that investors are behaving as economically rational individuals (Bias, Smith, & Jansson; 2012), then we can expect them to evaluate which firms are more likely to bring returns for their investment before they allocate their capital and build a portfolio. The high demand for stock of attractive enterprises with high return expectations should drive up their prices in the capital markets. Expected returns are, therefore, one of the key drivers of stock price and, ultimately, enterprise value. The ability to forecast accurately future returns can provides stock prices with some degree of predictability.

Some authors, however, have dived deeper into the topic by examining whether there can be other significant drivers of stock price. Hirshleifer (2001) describes asset pricing as in *vibrant flux*, observing in his work than a purely rational approach is not the most adequate to understand investor behavior and proposing a broader approach that is strongly influenced by psychology. Schiller (2000) refers to this psychologically driven volatility as *irrational exuberance* and affirms that is present in all asset markets.

2.1. Reactiveness to news

As mentioned above, some academic studies have shown that there may be different drivers for asset pricing that are not directly linked to traditional valuation models.

Mitchell and Mulherin (1994) analyzed the impact of public information on the stock market. To do so, authors attempted to identify potential relations between the number of news headlines and announcements reported by the Dow Jones on a daily basis and some financial references such as market returns for the period of the study. Their work showed a positive relationship between both variables, robust, but weaker and incapable of explaining part of the results.
Other authors have developed further this line of work. Chan (2008) performed a very relevant study putting the focus on over and under reaction by investors. In this study, the prices of several stock are analyzed. To do so, the author selected stock from companies which had been affected by public and relevant news during a specific period of time, and then contraposed it to a control group, companies with substantial price fluctuations that had not experienced such type of news.

The study found that the first type of stock, that is, stock from companies linked to relevant highlighted news, experienced a higher drift when the news were negative. Author categorizes this type of events as examples of under reaction to news by investors. The conclusion of the study is that investors appeared **to underreact to public signals** and **overreact to perceived private signals**. The effect was strongest for new stocks, where very negative returns combined with negative news headlines were directly related with underperformance for a period of time ranging up to 12 months.

Chan follows a statistical analysis and procedure that resembles the one chosen for this study. In order to assess the existence of relations between asset pricing and news, the author used a large sample of news headlines affecting randomly selected firms and the dates where they were made publicly available in several media and then incorporated into the analysis their stock price movements in those particular dates and within a determined period of time.

2.2. Loss effect

At the beginning of this master thesis, I mentioned how behavioral economics have been receiving a more relevant role in the field these past decades. The work of some important authors has brought more light to the psychology of economic agents and their decision-making processes which is valuable for conducting business. One of the most relevant studies on the topic was conducted by Kahneman and Tversky (1984), in which they identify some agent behavior as an anomaly. These phenomena, an expression of irrational behavior, constitute what the authors denominate as *loss aversion*. In some of the cases studied an observed by Kahneman and Tversky, subjects showed proof that **the negative utility of giving up an object** -the loss- **was greater than the utility originated with its acquisition**. It is possible to translate these utilities into a function, which shows an asymmetry of value for wins and losses.

Arkes, Herren and Isen (1988) studied the role of potential loss in the influence of affect on risk-taking behavior. To do so, authors designed an experiment choosing several undergraduate university students as subjects. These were asked to invest in lottery tickets that could lead to a reward. Prices for the lottery tickets were regulated according to supply and demand. During the experiment, it was possible to observe interesting risk-averse behavior and loss control: subjects showed high willingness to overpay for an insurance to cover potential losses. Arkes, Herren and Isen's study, however, was inconclusive due to the difficulty of interpreting the results and the concurrence of several factors that could be drivers for the relations observed. Authors noted though that positive results could potentially lead to both risk-taking and risk-averse behavior.

We can see, therefore, that in some instances economic agents can overreact to losses and engage in what can be considered irrational behavior. It will be interesting to see how this loss aversion can relate to the context of football matches and its results. As we will see in the empirical part of this project, I hypothesize that football match losses will have a greater effect than wins on investor behavior, especially in matches were losses conduct to elimination from the tournament as Edmans, García and Norli (2005) observed in their study.

3. PROFESSIONAL SPORTS AS A TRIGGER

In the previous chapter, we were able to observe how investors can react in different manners when exposed to an event, triggering sometimes irrational economic behavior that can be linked to asset pricing in the stock markets. This chapter will develop further irrational asset pricing, but we will do so by putting the emphasis on a specific field: professional sports. In order to perform our analysis, we will revise first a relevant paper whose aim is to investigate investor sentiment towards asset pricing in the context of professional sports results.

3.1. Behavioral biases triggered by single events

Edmans, García and Norli (2005) identify three main factors to determine whether a single event could have an impact relevant enough on investor sentiment in order to be translated into pricing: (1) strong drive on *mood;* (2) large *reach* in population; (3) *correlated effect* across the majority of individuals in the country.

The authors conclude that professional sports events, in particular international games, are a strong candidate that meets all requirements of behavioral biases triggers and hypothesize that they may have an effect on investor sentiment. They base this statement in the previous work of several authors Arkes, Herren, and Isen (1988); Hirt, Erickson, Kennedy, and Zillmann (1992). It is especially interesting to highlight Wann, Dolan, Mcgeorge, and Allison (1994) which show evidence of the strong connection between the results of sports event and the mood of sports followers. Authors observed that high-identification fans reported an increase in pre- to postgame positive emotions following a win and an increase in negative emotions following a loss.

Edmans, García and Norli provide some data that supports the high follower base for some sports, especially football, fulfilling the large reach requirement. Finally, the last factor they considered was the correlated effect across the nation, for which they chose international matches where national teams participated, making the assumption that the correlated effect would be given in this context.

3.2. Investor sentiment and reactiveness to sports results

As mentioned, in the previous chapter, single events seem to be able to trigger behavioral biases by some individuals given some factors are met. Edmans, García and Norli (2005) suggested professional sports, in particular football, as a driver to this irrational behavior.

In order to assess the hypothesized relations between match results and investor behavior, the study collects data from professional matches from several disciplines, being football match results approximately half of the total observations collected, as well as data regarding stock returns in several indices during that period of time. Results from the study show that match losses lead to a significant market decline, especially for smaller stock and matches with higher exposure.

Other authors have also studied potential relations between asset pricing and professional sport, centering the analysis on one particular sport: football. It is particularly interesting to highlight Palomino, Renneboog & Zhang (2009) whose focus was on the London Stock Exchange. Their study adds a new component into the analysis of investor sentiment: betting odds as a means to assess result expectations. Authors identify strong stock market reactions to news about the match results. As we were able to see in the first chapter, investor reaction to news can lead to irrational investor behavior and can be translated into asset pricing.

The study finds abnormal returns after a match, for one subsequent day in the case of wins and for a three-day period in the case of losses. This finding aligns with Chan (2003) and (2008) and his work on investor sentiment and reactiveness to headlines and news. We can observe, therefore, recurrent evidence on (1) investor overreaction to news, and to match results in particular; (2) higher drift and stronger effect in losses. As we will see in further chapters, one of my hypotheses following the research on loss aversion (Kahneman & Tversky, 1984; Arkes, Herren & Isen, 1988) is that losses have a stronger effect on investor sentiment and can be translated into stronger effect on asset pricing. We will need to analyze in depth the LOSS factor as a variable in order to assess whether my empirical study is aligned with my hypothesis and the findings of Palomino, Renneboog and Zhang, or whether it deviates from them.

Finally, the study shows that there is no strong evidence of investor reaction to betting odds publication by experts, despite the strong accuracy between the odds and the final match results. Authors come with two possible answers to this finding, which seems to be opposed to previous findings on investor reaction to news (Mitchell & Mulherin, 1994; Chan, 2003); (1) investors over react to match results and (2) betting odds information does not have enough visibility and reach despite being public in order to affect investor sentiment.

4. IMPLICATIONS AND RELEVANCE OF THE LITERATURE

During the previous two chapters I was able to review extensive work on my topic of study. Some works like Kahneman & Tversky (1984); Arkes, Herren, & Isen (1988); Hirt, Erickson, Kennedy, and Zillmann (1992) and Ariely (2008) were more theoretical and described some relevant concepts such as loss aversion and investor sentiment which helped to set the foundations for my research. Reviewing these helped me understand how economic agents can deviate due to external drivers from the predictable and rational behavior.

Once the framework of rational-irrational behavior was settled, I decided to pursue more research and define the scope to investors and the scope and to single events triggers. I believe that Chan (2003) and Mitchell, Mark & Mulherin (1994) brought me valuable insights on investor behavior and reactiveness to certain types of stimuli: public news.

After these first mostly theoretical papers, I decided to follow my research with more empirical and statistically driven work. Chan (2003) already showed empirical evidence of investor overreaction to news and headlines and several effects were also appreciated. Authors such as Edmans, García & Norli (2007) and Palomino, Renneboog, Zhang (2009) had previous research done that developed investor sentiment and reactiveness in the context of sports. This line of research was exactly my goal and, thus, their research became very valuable to frame my project. While Edmans, García and Norli's work was more generic and included several sports disciplines, Palomino, Renneboog and Zhang narrowed their researched to football. Findings of both papers are very insightful: while some of the effects appreciated are hard to explain due to concurrence of several mixed effects authors were able to confirm previous research on investor sentiment and overreaction to professional sports matches and results. Moreover, authors were able to appreciate a stronger effect for losses, which required more time on average to be absorbed by the market. These findings are aligned, as mentioned, with the results of previous research not focused on sports so it can be interpreted as a given case within a higher-level model of investor sentiment.

As I have described in the empirical part of this project, some of my ex-ante hypothesis are aligned with the findings from the literature review. I hypothesized, among other, that there is a statistically significant effect on asset pricing after football matches and that losses have a stronger effect than wins on investor sentiment and reactiveness. I will continue my research by narrowing my research to football matches as well as developing the mentioned researched papers. Besides providing further empirical evidence, I aim to introduce some differences that will make my work original and substantially different from Edmans et al. (2007) and Palomino et al. (2009). This can be easily achieved through data source selection. The first authors focus their study on data regarding international football and national teams. The second authors do so by focusing on the London Stock market and the traded teams in it. My goal for this thesis is to include data from several national competitions, and therefore, several stock markets, not only the London Stock Market.

II. EMPIRICAL PART

5. APPROACH, OBJECTIVES AND HYPOTHESIS

The literature review analyzed in the previous chapters focuses on one hand, on stock over and under reaction to news and, on the other hand, on the sport and professional matches as a trigger of irrational behavior.

This study will focus on stock valuation and its evolution for several firms operating in a particular industry: professional sports. **My objectives** for this project are the following:

a) To study the relationship between the results of professional sports matches and the stock price of the firms that manage the professional teams participating in these matches.

b) To study whether temporality is a relevant factor to stock price fluctuations when combined with a particular sport match result.

c) To study whether particular matches and its results may have a stronger effect or a special relation towards stock price fluctuations for the sport clubs participating.

d) To study whether the opposing team may be a relevant factor to stock price fluctuation for the sport teams participating.

Once considered the research approach and objectives, I propose the following generic and specific hypothesis:

5.1. Generic Hypothesis

- There is a relationship between the results of professional sport matches and the stock price fluctuations of the teams participating in them.

5.2. Specific Hypotheses

1. There will be a positive correlation between the result WIN and stock price valuation and a negative correlation between the result LOSS and stock price valuation.

2. The result LOSS will have a stronger effect on stock price valuation than the result WIN will. That is, *Losses* will have a higher impact on the stock price decrease than *Wins* will have on stock price increases.

3. The temporal variable will strengthen the effect. That is, a match taking place in a Weekend period (Friday to Sunday) will have a stronger effect on stock price fluctuation than the same result in another period of time.

4. The geographical variable will show different effects for different locations. That is, certain locations may experiment a stronger effect than others as a reaction to matches.

5. The singularity variable will strengthen the positive or negative effect of *wins* or *losses* on stock price valuation. That is, a win or a loss on specific matches between teams in the same geographic area or with a special rivalry relationship will have a stronger effect on the variable stock price than matches were this special condition does not concur.

6. METHODOLOGY AND DATA

The objective in this chapter is to show the material, variables, the procedure used to evaluate independent and dependent variables, as well as the statistical procedure used in the research.

6.1. Material

The material used during the empirical part of this research project in order to assess the selected independent and dependent variables has been the information obtained through the following sources:

a) Daily stock price data of nine sports clubs during the period [2012-2022] retrieved from the historical data on **https://finance.yahoo.com**, focusing especially on **closing price**.

b) Match result history of nine sports clubs during the period [2012-2022] retrieved from the specialized database **https://www.football-data.co.uk** for First Division **National League competitions**.

6.2. Variables in the study

Once all information had been collected, the variables for the statistical study were determined:

6.2.1. Independent variables

- Result (RES): We obtained 3 values for the result variable: WIN, DRAW, LOSS
- Weekend (WND): We obtained 2 values for the temporal variable: YES, NO.
- Location (LOC): We obtained 2 values for the location variable: HOME, AWAY.
- League (LE): We obtained 6 values for the league variable: POR, IT, NL, SCOT, GER, ENG.
- Singular match: (SING): We obtained 2 values for the singularity variable: YES, NO.

6.2.2. Dependent variables

As dependent variable for this study, that is, the variable whose value we hypothesize depends on the values of the independent variables, we have selected **stock prices (P)** of the nine firms managing the professional football clubs described in *Table 1*.

6.3. The model

Selecting the right statistical model for my analysis is one of the most important steps of my project. After carefully reviewing studies focusing on stock prices and capital markets, I have observed that most authors agree that linear regression models are generally accepted as the best fit (Karim, 2021; Panwar, 2021). Taking a look at my variables, I decided to use an Ordinary Least Squares (OLS) linear regression model.

This take is similar to the following model proposed by Edmans, García and Norli (2005):

$$\boldsymbol{R_{it} = \ 0_{it} + \ 1_i R_{it-1} + 2_{it} r_{Mt-1} + \ 3_{it} R_{Mt} + \ 4_i R_{Mt+1} + \ 5_i D_t + \ 6_i Q_t + \ i_t} \tag{1}$$

In this model, $\boldsymbol{R_{it}}$ represents the continuously compounded daily local currency return on a broadly-based stock market index for country *i* on day *t*, $\boldsymbol{R_{Mt}}$ represents the continuously compounded daily US dollar return world market index on day *t* on the database the authors used for their study, $\boldsymbol{D_t}$ is a set of dummy variables for time (Monday through Thursday) and $\boldsymbol{Q_t}$ is a set of dummy variables for holidays (days for which the previous one through five days are non-weekend holidays).

Their study shares similarities and some hypothesis with mine, although the data and some of the variables differ significantly. For that reason, I believe that the best approach is to follow a similar model as the one suggested in their work with some modifications accounting for the new and different variables I plan to introduce.

In order to estimate the effect of the outcome of professional football matches, I propose the following model:

$$\boldsymbol{R^{*}{}_{cl} = \ \alpha + \ \beta_1 RES_{cl} + \beta_2 LOC + \ \beta_3 WND + \ \beta_4 SING} \tag{2}$$

Where $\boldsymbol{R^*}$ represents the change in stock price between closing price for the first day consecutive to a match and last closing price before the match for each *club*; ***RES*** is a dummy variable representing the result of the match for each *club*; ***LOC*** is a dummy variable representing whether the teams object of study are playing home or away; ***WND*** is a dummy variable representing whether the match is played during a weekend period (which I have determined to be from Friday to Sunday); **SING** is a dummy variable representing whether the match has a singular character or especial relevance and importance. Finally, $\boldsymbol{\alpha}$ is a constant that captures all the effect that cannot be explained through the independent variables.

In following chapters of this thesis, I will explain what were the results I obtained from applying this method and why I decided to undertake one extra step. Model (2) can be applied for each individual club, but I decided to aggregate all the observations from each club into one final dataset that allowed me to apply the following model (3):

$$\boldsymbol{R^{*}_{aggr} = \alpha + \beta_1 RES_{aggr} + \beta_2 LOC + \beta_3 WND + \beta_4 SING + \beta_5 LE_{aggr}} \tag{3}$$

This model is equivalent as model (2) but allows me to assess one additional variable: $\boldsymbol{LE}$, which is a dummy variable representing the geographical location of the professional league for each *club*. The results obtained from this model can provide a more global output, less focalized on the country or national league, but still incorporate this element to analyze potential relationships.

6.4. General and Statistical Procedure

6.4.1. Data sourcing

The procedure of data collection has changed depending on the variable addressed. The table shows the professional football clubs that I selected for the study, the Professional leagues where they participate and the Exchange market where their stock is traded. I want to highlight than in my initial selection, there was an additional tenth team: AS Roma, which was excluded from the study for the reasons that I will describe below.

TEAM	PROFESSIONAL LEAGUE	STOCK EXCHANGE
Benfica	Liga I (Portugal)	Euronext Lisbon
Sporting CP	Liga I (Portugal)	Euronext Lisbon
FC Porto	Liga I (Portugal)	Euronext Lisbon

Lazio	Serie A (Italy)	Borsa Italiana
Juventus	Serie A (Italy)	Borsa Italiana
Ajax	Eredivisie (The Netherlands)	Amsterdam Stock Exchange
Celtic	Premier League (Scotland)	London Stock Exchange
Borussia Dortmund	Bundesliga 1 (Germany)	Frankfurt Stock Exchange
Manchester United	Premier League (England)	New York Stock Exchange

Table 1: Description of Professional clubs of the study. Source: own elaboration, based on researched databases.

The data retrieving process was constituted of two steps. Firstly, I retrieved match result data from the specialized web https://www.football-data.co.uk. The results obtained belonged only to domestic leagues. One of the factors I considered during this work was to include the international factor as an additional independent variable through the matches in international competitions such as *UEFA Champions League* (UCL) and *UEFA Europe League* (UEL), however due to the irregular pattern of participation of some of the teams in the sample and the complexity of adding this extra variable I decided to focus only on domestic competitions, at least during this first stage of the study.

Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	HS	AS	HST	AST	HF	AF	HC	AC	HY	AY	HR	AR	B365H	B365D	B365A	BWH	BWD	BWA	GBH	GBD	GBA	IWH	IWD	IWA	LBH	LB
I1	25/08/12	Fiorentina	Udinese	2	1	H	0	1	A	23	7	8	2	12	14	10	0	2	1	0	0	2.05	3.2	3.75	2.05	3.3	3.9	2.05	3.3	3.9	2	3.2	3.4	2.15	3.1
I1	25/08/12	Juventus	Parma	2	0	H	0	0	D	15	6	6	5	12	21	5	8	2	6	0	0	1.3	5	10.5	1.28	5	11	1.28	5	11	1.35	4.5	7.3	1.3	4.3
I1	26/08/12	Atalanta	Lazio	0	1	A	0	1	A	10	13	3	4	14	19	8	5	1	4	0	0	2.4	3.2	3	2.4	3.2	2.95	2.4	3.2	2.95	2.6	3.1	2.6	2.5	3.1
I1	26/08/12	Chievo	Bologna	2	0	H	0	0	D	6	4	3	1	17	18	1	1	3	3	0	1	2.1	3.2	3.6	2.1	3.15	3.6	2.1	3.15	3.6	2.2	3	3.2	2.1	3.1
I1	26/08/12	Genoa	Cagliari	2	0	H	0	0	D	15	17	4	4	19	12	3	6	0	3	0	0	1.95	3.3	4	2	3.25	3.8	2	3.25	3.8	2	3.15	3.45	2	3.2
I1	26/08/12	Milan	Sampdoria	0	1	A	0	0	D	23	10	6	3	16	11	7	2	2	5	0	0	1.4	4.5	8	1.48	4.1	7.75	1.48	4.1	7.75	1.45	3.9	6.5	1.36	
I1	26/08/12	Palermo	Napoli	0	3	A	0	1	A	16	12	6	3	11	13	6	5	3	3	0	0	3	3.25	2.38	3	3.4	2.4	3	3.4	2.4	2.7	3.2	2.4	2.62	3.2
I1	26/08/12	Pescara	Inter	0	3	A	0	2	A	13	14	5	6	7	12	10	2	3	1	0	0	4.33	3.6	1.8	5	3.55	1.7	5	3.55	1.7	4.3	3.4	1.75	4	3.5
I1	26/08/12	Roma	Catania	2	2	D	0	1	A	18	11	5	4	15	15	10	2	3	2	0	0	1.4	4.5	8	1.35	4.75	8.25	1.35	4.75	8.25	1.5	3.8	5.7	1.4	4.2
I1	26/08/12	Siena	Torino	0	0	D	0	0	D	5	8	0	1	10	17	4	5	0	1	0	0	2.25	3.1	3.4	2.25	3.15	3.25	2.25	3.15	3.25	2.2	3.1	3.1	2.1	3.1
I1	01/09/12	Bologna	Milan	1	3	A	1	1	D	15	17	3	6	18	21	1	8	1	5	0	0	3.6	3.2	2.1	3.8	3.25	2	3.8	3.25	2	3.2	3.2	2.1	3.75	3.2
I1	01/09/12	Torino	Pescara	3	0	H	1	0	H	16	1	9	0	12	20	5	0	1	2	0	1	1.83	3.4	4.5	1.83	3.5	4.6	1.83	3.5	4.6	1.9	3.3	3.7	1.85	3.3
I1	02/09/12	Cagliari	Atalanta	1	1	D	0	0	D	25	10	6	5	24	19	13	3	3	1	0	1	2.3	3	3.4	2.2	3	3.5	2.2	3	3.5	2.1	3.1	3.3	2.1	3.2
I1	02/09/12	Catania	Genoa	3	2	H	0	1	A	20	16	5	6	20	24	9	5	3	2	0	0	2.05	3.25	3.75	2	3.3	3.75	2	3.3	3.75	2	3.3	3.3	2.1	3.2
I1	02/09/12	Inter	Roma	1	3	A	1	1	D	21	12	6	3	17	17	11	2	2	2	0	1	1.95	3.6	3.75	2	3.5	3.8	2	3.5	3.8	1.85	3.3	3.9	1.95	3.5
I1	02/09/12	Lazio	Palermo	3	0	H	1	0	H	21	9	6	2	14	10	7	4	3	2	0	0	1.67	3.75	5	1.67	3.6	5.25	1.67	3.6	5.25	1.7	3.4	4.6	1.7	3.6
I1	02/09/12	Napoli	Fiorentina	2	1	H	0	0	D	5	16	1	3	17	10	7	9	2	2	0	0	1.73	3.5	5	1.75	3.5	4.75	1.75	3.5	4.75	1.7	3.5	4.4	1.73	3.5
I1	02/09/12	Parma	Chievo	2	0	H	1	0	H	11	14	3	4	22	17	3	4	3	3	0	0	2.05	3.2	3.8	2.05	3.2	3.7	2.05	3.2	3.7	2	3.2	3.4	2.05	3.2
I1	02/09/12	Sampdoria	Siena	2	1	H	1	0	H	12	9	5	3	16	22	11	1	3	3	0	1	1.83	3.4	4.5	1.83	3.3	4.5	1.83	3.3	4.5	1.85	3.3	3.9	1.83	3.4
I1	02/09/12	Udinese	Juventus	1	4	A	0	2	A	10	21	1	10	11	19	3	5	3	3	1	0	5	3.5	1.73	5.25	3.5	1.75	5.25	3.5	1.75	3.8	3.4	1.85	4.6	3.4
I1	15/09/12	Milan	Atalanta	0	1	A	0	0	D	17	10	7	4	12	19	8	5	2	4	0	0	1.5	4	7	1.42	4.1	8.25	1.42	4.1	8.25	1.45	3.9	6.5	1.4	
I1	15/09/12	Palermo	Cagliari	1	1	D	1	0	H	7	12	3	4	16	27	0	1	3	2	0	0	2.2	3.2	3.4	2.25	3.3	3.3	2.25	3.3	3.3	2.2	3.2	3	2	3.2
I1	16/09/12	Chievo	Lazio	1	3	A	0	2	A	11	12	2	6	11	20	10	1	3	2	0	0	2.8	3.1	2.6	2.85	3.2	2.45	2.85	3.2	2.45	2.7	3.05	2.5	2.88	3.1
I1	16/09/12	Fiorentina	Catania	2	0	H	1	0	H	15	8	7	4	15	16	5	3	3	2	0	0	1.91	3.3	4.2	1.83	3.45	4.25	1.83	3.45	4.25	1.85	3.3	3.9	1.73	3.4
I1	16/09/12	Genoa	Juventus	1	3	A	1	0	H	13	19	4	5	15	18	1	9	2	3	0	0	5.5	3.75	1.62	5.75	3.9	1.62	5.75	3.9	1.62	5	3.6	1.6	4.2	3.5
I1	16/09/12	Napoli	Parma	3	1	H	2	1	H	12	21	6	5	17	20	1	6	1	4	0	0	1.55	3.8	6.5	1.5	3.85	6.9	1.5	3.85	6.9	1.55	3.7	5.4	1.44	
I1	16/09/12	Pescara	Sampdoria	2	3	A	0	1	A	15	10	6	5	18	18	7	2	3	2	0	0	2.8	3.1	2.6	2.85	3.1	2.5	2.85	3.1	2.5	2.65	3.05	2.55	2.6	3.1
I1	16/09/12	Roma	Bologna	2	3	A	2	0	H	14	15	7	6	17	23	3	8	5	3	0	0	1.4	4.75	7.5	1.37	4.75	7.7	1.37	4.75	7.7	1.37	4.2	7.2	1.36	4.2
I1	16/09/12	Siena	Udinese	2	2	D	0	2	A	14	9	6	4	20	15	3	2	2	2	0	1	2.5	3.2	2.88	2.45	3.15	2.9	2.45	3.15	2.9	2.5	3.05	2.7	2.5	
I1	16/09/12	Torino	Inter	0	2	A	0	1	A	11	8	3	3	14	7	6	1	4	4	0	0	3.1	3.3	2.3	3.3	3.4	2.2	3.3	3.4	2.2	3.2	3.2	2.1	3	3.3

Figure 1: Example of extract of raw data for the Serie A, the Italian Division 1 football league, during the 2012-2013 season. Source: https://www.football-data.co.uk.

The second step in the data retrieving process was to find daily stock price data for the initial ten clubs selected. I was able to find daily data on open and closing prices as well as volume, peak and low prices on the specialized database **https://finance.yahoo.com**.

At this time, I observed that two of the selected clubs presented singular cases: (1) stock price data for one of the clubs (Manchester United) was only available after August 2012; (2) stock price data for AS Roma was limited due to its acquisition in August 2020 by *The Friedkin Group Inc.* After reviewing it, I decided to exclude the latter from the study, limiting the number of clubs within my sample to nine, which still represent six different national leagues and countries.

AJAX.AS

Date	Open	High	Low	Close	Adj Close	Volume
2012-06-01	7.250000	7.250000	7.110000	7.110000	6.796828	532
2012-06-04	7.100000	7.100000	7.100000	7.100000	6.787269	10
2012-06-05	7.150000	7.150000	7.150000	7.150000	6.835067	70
2012-06-06	7.100000	7.250000	7.100000	7.250000	6.930662	1017
2012-06-07	7.110000	7.125000	7.110000	7.110000	6.796828	88
2012-06-08	7.100000	7.100000	7.100000	7.100000	6.787269	119
2012-06-11	7.100000	7.100000	7.100000	7.100000	6.787269	26
2012-06-12	7.110000	7.250000	7.110000	7.250000	6.930662	235
2012-06-13	7.100000	7.110000	7.100000	7.110000	6.796828	180
2012-06-14	7.100000	7.100000	7.100000	7.100000	6.787269	411
2012-06-15	7.100000	7.250000	7.100000	7.250000	6.930662	43
2012-06-18	7.200000	7.200000	7.100000	7.110000	6.796828	138
2012-06-19	7.100000	7.100000	7.100000	7.100000	6.787269	39
2012-06-20	7.100000	7.100000	7.100000	7.100000	6.787269	116
2012-06-21	7.100000	7.110000	7.100000	7.110000	6.796828	391
2012-06-22	7.100000	7.110000	7.100000	7.110000	6.796828	184
2012-06-25	7.100000	7.110000	7.020000	7.020000	6.710793	954
2012-06-26	7.010000	7.020000	7.010000	7.020000	6.710793	11
2012-06-27	7.030000	7.030000	7.030000	7.030000	6.720352	242
2012-06-28	7.020000	7.100000	7.020000	7.020000	6.710793	73
2012-06-29	7.020000	7.020000	7.000000	7.000000	6.691673	1129
2012-07-02	7.000000	7.150000	7.000000	7.150000	6.835067	152
2012-07-03	7.150000	7.150000	7.000000	7.020000	6.710793	550
2012-07-04	7.000000	7.150000	7.000000	7.150000	6.835067	44
2012-07-05	7.000000	7.020000	7.000000	7.010000	6.701233	549
2012-07-06	7.000000	7.030000	7.000000	7.030000	6.720352	134
2012-07-09	7.000000	7.150000	7.000000	7.000000	6.691673	279

Figure 2: Example of extract of data of AFC AJAX NV, traded company at the Amsterdam Stock Exchange, for the 2012-2022 period. Source: https://finance.yahoo.com.

Furthermore, it is interesting to note that all leagues experienced a singular calendar schedule change, showing no match history between the months of April and June, while showing records of matches during July and August, a summer period that is usually considered a vacation period between seasons. This anomaly can be explained by the impact of COVID-19 and the interruption of the regular season due to public health reasons and ultimately had a slight impact on the total number of games and its distribution for some of the leagues.

At the end of this step, I had successfully collected data for both football matches and daily stock prices for the clubs and period object of my study. However, the data I obtained was raw and needed cleansing and treatment before it could be used for my statistical analysis. Financial data regarding stock prices included information and values that I did not want to include in my study, and the football results included more than 50 different metrics and KPIs that could be relevant for determining betting odds, but that had no significant value for the purposes of my study. For that reason, my next step was to select the most relevant data and cleanse into a format that would suit my statistical model and process.

6.4.2. Data cleansing

Once I had the data extracted, the next step was to select the most relevant metrics that would impact my work. Starting with the *football matches data*, I decided to keep data regarding the **time of the match** (*Date)*, which was key in order to create the connection with daily stock prices, as well as being related to one of my variables: *WND*; the **teams** and **location** of the match (*HomeTeam, AwayTeam)* participating in the match; the **score** (*FTHG* and FTAG or *Full Time Home Goals* and *Full Time Away Goals*); and the **result** (*FTR* or *Full Time Result*). In regards to the stock data, I decided to keep **opening** (*Open)* and **closing prices** (*Close*), not including in my final data high and low prices or volumes.

Finally, I aggregated the ten different data documents for match results that I had extracted in seasonal format for each national league into a unique and single working sheet for the entire 2012-2022 period, reducing the total amount of data sources from sixty to six distinct working documents.

6.4.3. Data treatment and categorization

The table below describes the total number of observations of my study within some parameters, that is, the total number of matches retrieved from **all the professional teams** participating in each of the **six national leagues** and each of the **ten seasons** of the 2012-2022 period. The total number of observations is 18805, however, for the purposes of this study only a small part of these observations will be relevant: the observations belonging to the nine professional clubs traded in stock markets that were selected as targets of the study.

SEASON / LEAGUE	POR	IT	GER	ENG	SCOT	NL	TOTAL
2012-2013	240	380	306	380	228	306	**1840**
2013-2014	240	380	306	380	228	306	**1840**
2014-2015	306	380	306	380	228	306	**1906**
2015-2016	306	380	306	380	228	306	**1906**
2016-2017	306	380	306	380	228	306	**1906**
2017-2018	306	380	306	380	228	306	**1906**
2018-2019	306	380	306	380	228	306	**1906**
2019-2020	306	380	306	380	179	232	**1783**
2020-2021	306	380	306	380	228	306	**1906**
2021-2022	306	380	306	380	228	306	**1906**
TOTAL	**2928**	**3800**	**3060**	**3800**	**2231**	**2986**	**18805**

Table 1: Total number of matches by season and national league. Source: own elaboration based on https://www.football-data.co.uk

After the first aggregation, the next step was to identify which matches involved one or several of the selected clubs and, once identified, categorize them based on the independent variables described.

The independent variable ***RES*** has three different values: *WIN, DRAW, LOSS*. I decided to treat this variable as a dummy variable, computing it as follows:

$$RES_{cl} = \{RES_{win} \; ; \; RES_{draw} \; ; \; RES_{loss}\}$$

Each of these variables receives a value of 1 if the observed match belongs to that category or a value of 0 if the match has a different result.

The independent variable ***LOC*** has two different values: *HOME* or *AWAY*. I decided to treat this variable as a dummy variable, assigning the value 1 for those observations where the studied team was playing home, and the value 0 when it was playing away.

The independent variable ***WND*** has two different values: *YES* or *NO*. I decided to treat this variable as a dummy variable, assigning the value 1 for those observations where the studied team was playing during a weekend period (Friday to Sunday), and the value 0 when it was playing during weekdays (Monday to Thursday).

The independent variable ***SING*** has two different values: *YES* or *NO*. I considered singular games, those that took place between clubs that have a historical rivalry or are close competitors for the same objectives. Although the degree of singularity of a game can be of subjective character, by limiting my assumption towards the most relevant matches in terms of supporter bases, viewership ratings and tradition, I believe the final categorization held an objective character. I assigned a value of 1 to those matches that had a singular character and a value of 0 to those that did not.

The independent variable ***LE*** has six potential different values: IT, GER, POR, ENG, SCOT. I decided to compute them as the following dummy variable:

$$LE_{cl} = \{LE_{it} \ ; \ LE_{ger} \ ; \ LE_{por} \ ; \ LE_{eng} \ ; \ LE_{scot}\}$$

Each of these variables receives a value of 1 if the observed match belongs to that specific country's national league, or a value of 0 if it belongs to a different national league. After selecting the categories and values that observations can receive, the cleaned data presented the form shown in *Figure 3*.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Date	Weekday	HomeTeam	AwayTeam	FTHG	FTAG	FTR	CLUB	WND	LOC	LE	RES	SING
11	20/08/2012	1	Everton	Man United	1	0	H	TRUE	NO	AWAY	UK	LOSS	NO
15	25/08/2012	6	Man United	Fulham	3	2	H	TRUE	YES	HOME	UK	WIN	NO
30	02/09/2012	7	Southampton	Man United	2	3	A	TRUE	YES	AWAY	UK	WIN	NO
34	15/09/2012	6	Man United	Wigan	4	0	H	TRUE	YES	HOME	UK	WIN	NO
47	23/09/2012	7	Liverpool	Man United	1	2	A	TRUE	YES	AWAY	UK	WIN	YES
54	29/09/2012	6	Man United	Tottenham	2	3	A	TRUE	YES	HOME	UK	LOSS	NO
68	07/10/2012	7	Newcastle	Man United	0	3	A	TRUE	YES	AWAY	UK	WIN	NO
73	20/10/2012	6	Man United	Stoke	4	2	H	TRUE	YES	HOME	UK	WIN	NO
87	28/10/2012	7	Chelsea	Man United	2	3	A	TRUE	YES	AWAY	UK	WIN	NO
92	03/11/2012	6	Man United	Arsenal	2	1	H	TRUE	YES	HOME	UK	WIN	NO
102	10/11/2012	6	Aston Villa	Man United	2	3	A	TRUE	YES	AWAY	UK	WIN	NO
115	17/11/2012	6	Norwich	Man United	1	0	H	TRUE	YES	AWAY	UK	LOSS	NO
123	24/11/2012	6	Man United	QPR	3	1	H	TRUE	YES	HOME	UK	WIN	NO
135	28/11/2012	3	Man United	West Ham	1	0	H	TRUE	NO	HOME	UK	WIN	NO

Figure 3: Example of categorized data for Premier League matches in the 2012-2022. Source: own elaboration, based on data retrieved from https://finance.yahoo.com.

These values and categories already brought some insights and improved the visibility of the data initially retrieved, but some more treatment was necessary in order to be able to use it as input into the model. For that reason, the next step in the analysis was to transform these categorical variables into *dummy* binary variables that could be used in Multiple regression models.

N	O	P	Q	R	S
WND Binary	LOC Binary	RES win	RES loss	RES draw	SING binary
0	0	0	1	0	0
1	1	1	0	0	0
1	0	1	0	0	0
1	1	1	0	0	0
1	0	1	0	0	1
1	1	0	1	0	0
1	0	1	0	0	0
1	1	1	0	0	0
1	0	1	0	0	0
1	1	1	0	0	0
1	0	1	0	0	0
1	0	0	1	0	0
1	1	1	0	0	0
0	1	1	0	0	0
1	0	1	0	0	0
1	0	1	0	0	1
1	1	1	0	0	0
1	0	0	0	1	0
0	1	1	0	0	0

Figure 4: Example of binary categorical variables for Premier League matches in the 2012-2022 period. Source: own elaboration, based on data retrieved from https://finance.yahoo.com.

T	U	V
P before game	P after Game	P variation
13,06	13,19	0,13
13,26	13,83	0,57
13,30	13,19	-0,11
12,45	12,96	0,51
12,72	13,37	0,65
12,73	12,55	-0,18
12,98	13,00	0,02
12,25	12,38	0,13
12,44	12,59	0,15

The final step was to determine the dependent variable: stock prices. In order to do so, I estimated the difference between the earliest opening stock price after a match and the latest close price before the match. This date was added into subsequent columns on the data sheet.

Figure 5: Dependent variable "Price variation" for Manchester United Premier League matches in the 2012-2022 period. Source: own elaboration.

6.4.4. Data aggregation and final data set

The previous section describes the process to obtain the final data sets for each of the professional clubs that are object of the study. I was able to perform individual regressions on each of the clubs, but I will be describing the results obtained in the its corresponding section.

Due to the results obtained, I considered one additional step: to aggregate all the data from the six different professional leagues. In order to do so, I created six binary variables for each of the professional leagues: IT, POR, SCOT, ENG, NL, GER and cleansed the data to keep exclusively the variables that were relevant for the study: (1) the dependent quantitative variable *P_var* or *R**; (2) the independent variables transformed into binary dummy variables *WND, LOC, RES, SING* and *LE*.

As a result, I obtained a final data set for the 2012-2022 period that covered **six national leagues**, **nine different professional football teams** traded in capital markets and a final sample with a **total of 3179 observations**. This data set is the one that I decided to use for my final statistical study in order to find potential relationships between variables and in order to assess the validity of my hypothesis at the start of my thesis.

It is important to note that there was a significant change to the dependent variable at this point. I planned to use R* or difference between latest closing price and earliest opening price before and after each of the observed matches and for each individual club. However, each club had a different stock price and, therefore, fluctuations in price could be significantly different for each club. Once all of this data was aggregated I needed to find a new variable that could be representative of the stock returns but at the same time would englobe all eight stocks. For that reason, I decided to transform the *P_var* variable -or difference in price- into ***P_var_percent***, a new dependent variable that accounted for the percental increase or decrease in price between latest closing and earliest opening prices before and after the observed matches. With this new variable, I was now able to compare different stocks regardless of the nominal price of each individual stock.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	WND_binary	LOC_binary	RES_[illegible]	[illegible]s	RES_draw	SING_bin	LE_IT	LE_POR	LE_SCOT	LE_ENG	LE_NL	LE_GER	P_var	P_var_percent
2	1	1	1	0	0	0	1	0	0	0	0	0	0,001	0,52%
3	1	0	1	0	0	0	1	0	0	0	0	0	0,0004	0,20%
4	1	0	1	0	0	0	1	0	0	0	0	0	0,001	0,48%
5	1	1	1	0	0	0	1	0	0	0	0	0	0,0006	0,28%
6	0	0	0	0	1	0	1	0	0	0	0	0	-0,009	-4,23%
7	1	1	1	0	0	0	1	0	0	0	0	0	0,0032	1,56%
8	1	0	1	0	0	0	1	0	0	0	0	0	-0,0062	-3,03%
9	1	1	1	0	0	0	1	0	0	0	0	0	0,0042	2,09%
10	1	0	1	0	0	0	1	0	0	0	0	0	0,0004	0,20%
11	0	1	1	0	0	0	1	0	0	0	0	0	0,0019	0,96%
12	1	1	0	1	0	1	1	0	0	0	0	0	-0,0043	-2,16%
13	1	0	1	0	0	0	1	0	0	0	0	0	-0,001	-0,52%
14	1	1	0	0	1	0	1	0	0	0	0	0	0,0028	1,51%
15	1	0	0	1	0	1	1	0	0	0	0	0	-0,0027	-1,40%
16	1	1	1	0	0	1	1	0	0	0	0	0	0,0015	0,76%

Figure 6: Example of final data set with independent variables in yellow and dependent variable in pink for the six national leagues, eight teams and 2012-2022 period. Source: Own elaboration.

6.5. Notation

Abbreviation	Term
D	Draw
ENG	England
GER	Germany
IT	Italy
L	Loss
LE	League
LOC	Location
NL	The Netherlands
P	Price
POR	Portugal
RES	Result
SCOT	Scotland
SING	Singular Match
W	Win
WND	Weekend

7. RESULTS

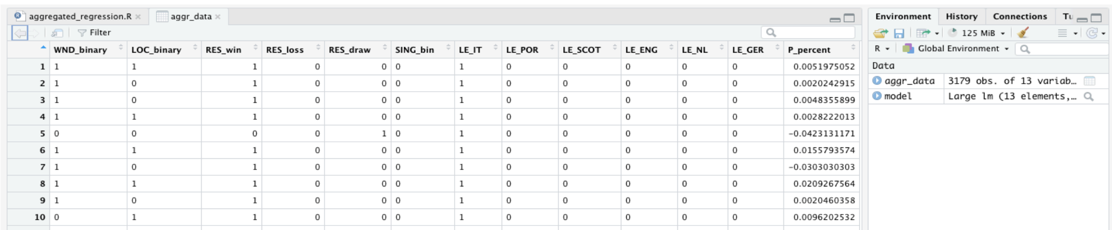
The purpose of this section is to describe the results obtained after the application of several econometric tools to the data set that I retrieved, cleansed and structured. For that purpose, I decided to organize this chapter in several sections. Firstly, we will take a look into the case description followed by an initial analysis of the data and the results obtained after the analysis.

Secondly, we will take a detailed look into the variables that were established in earlier chapters in order to accept or reject the hypothesis proposed in this master thesis. Finally, we will describe the potential effects observed between variables and its implications towards the different variables.

7.1. Case description

My objective with this research project was to assess whether there was any relationship between the match results of football clubs and its stock valuation that could not be explained by the market itself, that is, an effect that could be described as part of irrational investor behavior out of the principle of economic rationality.

For that reason, I proceeded to retrieve data from a sample of publicly traded public clubs and their match results during the 2012-2022 period and I elaborated a final data set (Figure 6) that would allow me to perform a multiple variable regression analysis in order to learn more about these variables. From **the total of approximately 18805 matches and 37610 observations**, **only 3179 observations** were interesting for the purposes of the study. The following figure shows a representation of the data and initial case for my study within the statistical tool I was planning to use:



	WND_binary	LOC_binary	RES_win	RES_loss	RES_draw	SING_bin	LE_IT	LE_POR	LE_SCOT	LE_ENG	LE_NL	LE_GER	P_percent
1	1	1	1	0	0	0	1	0	0	0	0	0	0.0051975052
2	1	0	1	0	0	0	1	0	0	0	0	0	0.0020242915
3	1	0	1	0	0	0	1	0	0	0	0	0	0.0048355899
4	1	1	1	0	0	0	1	0	0	0	0	0	0.0028222013
5	0	0	0	0	1	0	1	0	0	0	0	0	-0.0423131171
6	1	1	1	0	0	0	1	0	0	0	0	0	0.0155793574
7	1	0	1	0	0	0	1	0	0	0	0	0	-0.0303030303
8	1	1	1	0	0	0	1	0	0	0	0	0	0.0209267564
9	1	0	1	0	0	0	1	0	0	0	0	0	0.0020460358
10	0	1	1	0	0	0	1	0	0	0	0	0	0.0096202532
11	1	1	0	1	0	1	1	0	0	0	0	0	-0.0216189040

Figure 7: Case description and clean final aggregated dataset in RStudio ready for multiple variable regression analysis. Source: own elaboration.

7.2. Initial analysis

In order to perform my multiple variable regression analysis, I decided to use the *RStudi*o software. For that reason, I imported the datasets elaborated and described in the previous chapter to the statistical tool. As mentioned in the previous chapter, I initially started by analysis on each individual club, therefore performing eight different regression analyses. After all of these and analyzing carefully the results obtained, I decided to pursue one final analysis on the aggregated data of all clubs in order to have a bigger dataset and in order to incorporate the *LE* variable. It is important to note that for the purposes of accepting or refusing the Null hypothesis in my regressions, I will be taking 0,05 or 5% as statistically significant value.

7.2.1. Clubs analyses

In order to analyze the variables obtained in previous chapters, I started my analysis by focalizing in each individual club. The data I used, therefore, was data on each singular national league and matches, which I subsumed into the following Model (2):

$$\boldsymbol{R^{*}{}_{cl} = \alpha + \beta_1 RES_{cl} + \beta_2 LOC + \beta_3 WND + \beta_4 SING} \tag{2}$$

As a result, I obtained the quotients $\boldsymbol{\beta}$ for each of the following variables: RES, LOC, WND, SING. There are two key factors that I have to highlight: (1) at this stage, and as mentioned previously, the variable LE was still not present in the analysis and will be incorporated in the aggregated regression; (2) in order to avoid collinearity, I did not include the variable *RES_draw* in the regression. The results obtained will, therefore, set the effect of a draw on stock as a basis and provide the positive or negative effect than a win or a loss may have in comparison.

I started my analyses with *Manchester United* and the *Premier League (ENG)*. The following figure shows an extract of the dataset used in the analysis.

Filter

	WND_binary	LOC_binary	RES_win	RES_draw	RES_loss	SING_binary	P_var
1	0	0	0	0	1	0	0.13
2	1	1	1	0	0	0	0.57
3	1	0	1	0	0	0	-0.11
4	1	1	1	0	0	0	0.51
5	1	0	1	0	0	1	0.65
6	1	1	0	0	1	0	-0.18
7	1	0	1	0	0	0	0.02
8	1	1	1	0	0	0	0.13
9	1	0	1	0	0	0	0.15
10	1	1	1	0	0	0	-0.01
11	1	0	1	0	0	0	0.01
12	1	0	0	0	1	0	0.08
13	1	1	1	0	0	0	0.17

86 MiB

R Global Environment

Data	
MANU	380 obs. of 22 va…
MANU_cl…	380 obs. of 7 var…
model	List of 13
model_2	List of 13
model_3	List of 13
uk_data	3800 obs. of 22 v…

Figure 8: Extract of data corresponding to the Manchester United dataset, in Rstudio, clean and ready for a multiple regression analysis. Source: own elaboration.

I proceed to perform a multiple regression analysis with the function *lm*, establishing *P_var* as dependent variable and the binary dummy variables as independent variables. The result of my first regression is shown in the following figure:

```
Call:
lm(formula = P_var ~ WND_binary + LOC_binary + RES_win + RES_loss +
    SING_binary, data = BVB_clean)

Residuals:
     Min       1Q   Median       3Q      Max
-0.69929 -0.05663  0.00390  0.06550  0.56262

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.074110   0.030238  -2.451   0.0148 *
WND_binary1   0.016649   0.028517   0.584   0.5597
LOC_binary1   0.006282   0.015025   0.418   0.6762
RES_win       0.083562   0.019738   4.234 2.98e-05 ***
RES_loss     -0.038253   0.022980  -1.665   0.0969 .
SING_binary1  0.061228   0.024171   2.533   0.0118 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1354 on 334 degrees of freedom
Multiple R-squared:  0.1336,	Adjusted R-squared:  0.1207
F-statistic:  10.3 on 5 and 334 DF,  p-value: 3.343e-09
```

Figure 9: Regression # 1: Multiple regression test performed with RStudio on Manchester United dataset. Source: own elaboration.

The regression test shows that all variables except WND do not pass the p-value test and are not statistically significant. We have to accept the Null hypothesis for this regression. Furthermore, $\boldsymbol{R^2}$ has a value of 0,02795, and therefore, the model can only explain a 2,80% of the effect observed in the dependent variable, that is, stock price variation. This model, therefore, has a low accuracy and lack of statistical significance.

The next step on my research was to perform a similar analysis on a different club and stock in order to compare results. I decided to choose *Borussia Dortmund* and *GER* as my next club and league to study. The contiguous figure reproduces the same approach observed before for the Dortmund dataset.

```
Call:
lm(formula = P_var ~ WND_binary + LOC_binary + RES_win + RES_loss +
    SING_binary, data = MANU_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-1.5307 -0.1665 -0.0101  0.1340  1.4866

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.1186191  0.0475075  -2.497  0.01296 *
WND_binary1   0.1227103  0.0395341   3.104  0.00206 **
LOC_binary1   0.0229790  0.0334401   0.687  0.49240
RES_win      -0.0006821  0.0404243  -0.017  0.98655
RES_loss     -0.0372986  0.0490504  -0.760  0.44749
SING_binary1  0.0002342  0.0548223   0.004  0.99659
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3225 on 374 degrees of freedom
Multiple R-squared:  0.02795,	Adjusted R-squared:  0.01495
F-statistic:  2.15 on 5 and 374 DF,  p-value: 0.05893
```

Figure 10: Regression # 2: Multiple regression test performed with RStudio on Borussia Dortmund dataset. Source: own elaboration.

The regression test shows that both the RES_win and SING variables pass the p-value test and are statistically significant. The other variables, however, do not manage to pass the test and the Null hypothesis for each of them needs to be accepted. Furthermore, $\boldsymbol{R^2}$ has a value of 0,1354, and therefore, the model can only explain a 13,36% of the effect observed in the dependent variable, that is, stock price variation. The model for *Dortmund* stock, therefore, has a relatively low accuracy but significantly higher than the previous model for *Manchester United*. I will analyze these differences and the reasons that could explain the gap between both stocks in further chapters.

My next step, once I observed the low significance of the model, was to attempt to refine or perfect it. I observed that the model I initially suggested, proposed the independent variables as unique and non-related to each other. But, could the real situation be different? For example, could a WIN or LOSS have a stronger effect depending on whether it takes places as a HOME team or as an AWAY team? It seems plausible that a LOSS in a team’s own stadium could have a stronger emotional impact on fans, potentially on investors, and that this strong effect could be translated to the teams’ stock. Furthermore, it seems that the variable SING could potentially be affected by the result of a match. For that reason, I decided to incorporate these new variables into the model.

The result is not significantly different from what I had obtained in the previous regression. *RES_win* remains being statistically significant, but now we can observe that the *SING* variable is no longer statistically significant. However, the *RES_loss* variable passes the p-value test this time. Overall, we can observe that $\boldsymbol{R^2}$ increases slightly in value to 0,1464 or 14,64%. This value is still low and the model cannot explain a significant part of the effect observed on the dependent variable, but it is slightly more accurate than the previous model.

```
Call:
lm(formula = P_var ~ WND_binary + LOC_binary * RES_win + SING_binary *
    RES_win + LOC_binary * RES_loss + RES_loss * SING_binary,
    data = BVB_clean)

Residuals:
     Min       1Q   Median       3Q      Max
-0.68172 -0.05887  0.00365  0.06559  0.55901

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           -0.06496    0.03429  -1.894  0.05907 .
WND_binary1            0.01495    0.02895   0.516  0.60594
LOC_binary1           -0.01932    0.03473  -0.556  0.57843
RES_win                0.07951    0.02640   3.012  0.00279 **
SING_binary1           0.07785    0.04897   1.590  0.11283
RES_loss              -0.06326    0.03049  -2.075  0.03878 *
LOC_binary1:RES_win    0.02581    0.04003   0.645  0.51961
RES_win:SING_binary1  -0.08764    0.06573  -1.333  0.18332
LOC_binary1:RES_loss   0.04510    0.04702   0.959  0.33812
SING_binary1:RES_loss  0.02341    0.06068   0.386  0.69994
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1353 on 330 degrees of freedom
Multiple R-squared:  0.1464,    Adjusted R-squared:  0.1231
F-statistic: 6.287 on 9 and 330 DF,  p-value: 3.235e-08
```

Figure 11: Regression 3: Multiple variable regression performed with RStudio on Borussia Dortmund dataset. Source: own elaboration

The analysis for each of the individual clubs followed similar results as the two individual cases presented above. I will proceed to summarize the results of each individual regression in the table below taking into account (1) the R^2 value of each regression; (2) what variables pass the p-value test and are statistically significant for each regression:

TEAM	R^2 VALUE	WND	RES_win	RES_loss	LOC	SING
Benfica	6,86%	Yes	Yes	No	No	No
Sporting CP	3,67%	No	Yes	No	No	Yes
FC Porto	1,23%	No	No	No	No	No
Lazio	15,64%	Yes	Yes	No	Yes	No
Juventus	1,91%	No	No	No	No	No
Ajax	2,00%	No	No	No	No	No
Celtic	1,72%	No	No	No	No	No
Borussia Dortmund	14,64%	No	Yes	Yes	No	No
Manchester United	2,80%	Yes	No	No	No	No

Table 2: Summary of individual regressions for the nine selected clubs. Source: own elaboration.

My next step in the analysis will be to aggregate all the data from different clubs to increase the number of observations in the sample and to add the *LE* variable into the analysis.

7.2.2. Aggregated analysis

I proceed to perform a multiple regression analysis with the *RStudio* function *lm*, establishing *P_percent* as dependent variable and the binary dummy variables as independent variables. This new model has several differences with the previous one that have been mentioned in the previous chapter.

Firtstly, it incorporates six new binary variables corresponding to each of the professional leagues of the study: *LE_IT, LE_SCOT, LE_PORT, LE_GER, LE_NL, LE_ENG*. In order to perform the analysis and to avoid collinearity problems, I will not include one of them in the regression: *LE_ENG*. This means that the results obtained will be based on the Premier League as a standard benchmark.

Furthermore, the dependent variable has significantly change. Since this new dataset aggregates the data of all nine clubs and stocks, I needed a new dependent variable that could englobe the value of all stocks. For that reason, I decided to transform the pure nominal variation in price used in previous regressions to percental increments and decreases in the prices. As a result, I obtained the dependent variable *P_percent*.

```
Call:
lm(formula = P_percent ~ WND_binary + LOC_binary + RES_win +
    RES_loss + SING_bin + LE_IT + LE_POR + LE_SCOT + LE_NL +
    LE_GER, data = aggr_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.190432 -0.010201 -0.001607  0.009776  0.223713

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0079137  0.0025587  -3.093  0.00200 **
WND_binary1  0.0049488  0.0016806   2.945  0.00326 **
LOC_binary1 -0.0020038  0.0012429  -1.612  0.10703
RES_win      0.0097262  0.0016353   5.948 3.02e-09 ***
RES_loss    -0.0062315  0.0021692  -2.873  0.00410 **
SING_bin1    0.0035085  0.0020841   1.683  0.09239 .
LE_IT1      -0.0009025  0.0021813  -0.414  0.67911
LE_POR1     -0.0020847  0.0021084  -0.989  0.32286
LE_SCOT1    -0.0007257  0.0025587  -0.284  0.77673
LE_NL1      -0.0034015  0.0026364  -1.290  0.19708
LE_GER1     -0.0014003  0.0026007  -0.538  0.59031
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03465 on 3168 degrees of freedom
Multiple R-squared:  0.03191,   Adjusted R-squared:  0.02886
F-statistic: 10.44 on 10 and 3168 DF,  p-value: < 2.2e-16
```

Figure 12: Regression #4: Multiple variable regression performed in RStudio on the Aggregated dataset. Source: own elaboration

The regression test shows that both the RES_win and RES_loss variables pass the p-value test and are statistically significant. The variable *WND* is also statistically significant and passes the test. The other variables, both SING and the LE variables for each geography have high p-values and do not pass the test, therefore, we need to accept the Null hypothesis for each of them. Furthermore, $\boldsymbol{R^2}$ has a value of 0,0319, and therefore, the model can only explain a 3,19% of the effect observed in the dependent variable, that is, stock price variation. The aggregated model for all the stocks, therefore, has a relatively low accuracy and cannot explain a significant part of the effect observed on the stocks. I attempted a similar approach as with the individual stock by adding a potential relationship between RES and SING or LOC, but neither passed the p-value test and the $\boldsymbol{R^2}$ decreased slightly.

We observe, then, that the only variables that pass the p-value and are statistically significant are RES, both as WIN and LOSS, and WND. I performed one last regression analysis incorporating only these statistically significant variables. The results are aligned with our previous findings:

```
Call:
lm(formula = P_percent ~ WND_binary + RES_win + RES_loss, data = aggr_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.192104 -0.010216 -0.001788  0.009672  0.224011

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.009649   0.001956  -4.934 8.47e-07 ***
WND_binary1  0.004983   0.001657   3.007  0.00266 **
RES_win      0.009045   0.001610   5.617 2.12e-08 ***
RES_loss    -0.005859   0.002151  -2.725  0.00647 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03465 on 3175 degrees of freedom
Multiple R-squared:  0.02968,    Adjusted R-squared:  0.02876
F-statistic: 32.37 on 3 and 3175 DF,  p-value: < 2.2e-16
```

Figure 13: Regression #5: Multiple variable regression performed in RStudio on the aggregated dataset, incorporating only statistically significant variables.

As it can be appreciated in Figure 13, the variables WND and RES (both for WIN and LOSS) remain being statistically significant and pass the p-value test. The $\boldsymbol{R^2}$ value, however, decreased slightly from 3,19% to 2,97%. The results clearly show that most of the effect observed is due to a variable not included in the model. I will analyze each of the variables and its implications in the following sections.

7.3. Comparison and analysis of independent and dependent variables

In this section I will analyze each variable and compare them to my specific hypotheses. As mentioned before, $\boldsymbol{R^2}$ can only explain a 3,19% of the effect observed in the dependent variable, and therefore, any findings I make from the results of the model will not have a significant statistical reach. However, I still believe that taking a look into each hypothesis and variable can be interesting in order to make some conclusion of my research.

1. My first hypothesis was that there would be a **positive correlation** between the result WIN and stock price valuation and a **negative correlation** between the result LOSS and stock price valuation. Taking a look into my aggregated regression (4), we can observe that the quotients for this variable are $\boldsymbol{\beta_{win} = 0,0097}$ and $\boldsymbol{\beta_{loss} = -0{,}0062}$. Both values of the variable RES pass the p-value test and are statistically significant. Therefore, we have to **reject the Null Hypothesis** for them and accept the alternative hypothesis that these independent variables have an effect on our dependent variable, stock fluctuations.

2. My second hypothesis was that the result LOSS would have a stronger effect on stock price valuation than the result WIN would. Taking a look into regression (4), we can observe that the negative quotient of RES_loss is inferior to that of RES_win. That means that WIN has a stronger impact on stock price than LOSS. We have to **reject this specific hypothesis**.

3. The third hypothesis was that the temporal variable would have an impact on the effect. That is, a match taking place in a Weekend period (Friday to Sunday) will have a stronger effect on stock price fluctuation than the same result in another period of time. The variable *WND* passes the p-value effect and has a positive contribution to the effect observed in stock price fluctuation with a quotient of $\boldsymbol{\beta_{wnd} = 0,0049}$. We **reject the Null Hypothesis** for the WND variable and accept the alternative hypothesis that WND is statistically significant for R*.

4. My fourth specific hypothesis was that the geographical variable would show a significant and positive effect contribution that would vary depending on the different locations. Regression (4) shows that none of the different binary variables of the variable

LE pass the p-value test and, therefore, are not statistically significant. We have to **accept the Null Hypothesis** and reject the alternative hypothesis that these variables were significant and contributed to the effect observed on the dependent variable.

5. The last specific hypothesis was that the singularity variable would contribute to stock price valuation. Taking a look into regression (4), we can observe that the variable SING does not pass the p-value test and, therefore, is not statistically significant. I attempted a second regression, establishing a potential relationship between the variable *SING* and *RES,* that is, I hypothesized that the effect of a singular match on stock prices would be related to the result of the match itself, not exclusively on the condition of singular match. My line of thought was that singular matches would behave as a booster of the confirmed effect that WIN and LOSS have on R*. However, the additional regression did not pass p-value tests either and had even lower R^2 value. In conclusion, we have to **accept the Null Hypothesis** and reject the alternative hypothesis that singular matches contributed to the effect observed on the dependent variable.

7.4. Effects observed

Performing several analyses allowed me to appreciate some effects and to discern some variables that did not hold the effect that I hypothesized. The main effects observed have been the following:

1. Positive correlation and effect contribution between the independent variable RES_WIN and the dependent variable R* or P_percent.

2. Negative correlation and effect contribution between the independent variable RES_LOSS and the dependent variable R* or P_percent.

3. Positive correlation and effect contribution between the independent variable WND and the dependent variable R* or P_percent.

In conclusion, the final model that synthesizes the analysis, taking the aggregated dataset results, is as follows:

$$\boldsymbol{R^*}_{\boldsymbol{aggr}} = \boldsymbol{\alpha} + \boldsymbol{\beta_1}\,\boldsymbol{RES_{aggr}} + \boldsymbol{\beta_2}\,\boldsymbol{WND}$$

(5a)

Adding the quotients and intercept found in the regression, the model turns into the following:

$$\boldsymbol{R^*}_{\boldsymbol{i}} = -\mathbf{0,0079} + \mathbf{0,0097}(\boldsymbol{WIN}) - \mathbf{0,0062}(\boldsymbol{LOSS}) + \mathbf{0,0049}(\boldsymbol{WND})$$

(5b)

This model, therefore, should be able to predict the percental increase of the stock price of a football club based on (1) the result of a match; (2) whether the match is held during a weekend or not. However, as mentioned in the previous section, the robustness of this model is not significant enough in order to make an accurate prediction.

8. CONCLUSIONS

8.1. Review of objectives

The purpose of this research project was (1) to extend the empirical knowledge about individual's irrational behavior, putting our focus on investors of traded stock and (2) to do so by putting our sport, in particular professional football, at the center of our research. I expected my analysis to provide more insights on what role play professional sports, especially match results, towards the valuation of the stock of the teams.

The thesis was structured in two defined blocks: a theoretical and an empirical block. The first block's objective was to revise existing literature, to comment it and to highlight potential similarities and differences to my research goal in order to extract as much information as possible before the empirical approach. In that sense, I was able to find more literature and research work than I initially expected, thus, I consider this stage very successful in its initial documentation steps. Some of the works I reviewed (Chan, 2003; Edmans, García & Norli, 2007; Palomino, Renneboog, & Zhang, 2009) had similar objectives as the ones I proposed and were a perfect benchmark in order to see previous work in the same line of thought. I attempted, however, to avoid overlapping in the scope in order to bring new insights into the topic and I believe I managed to successfully diverge enough from these authors research in order to gain some degree of autonomous originality on my work. I can conclude, therefore, that I was able to successfully document and review interesting materials that were helpful for my study and that helped create the background knowledge to further dive into my research questions and my statistical approach.

The second block of this project constituted a statistical analysis, performed with several econometric tools such as *RStudio*, in order to identify potential relationships between the dependent and independent variables that I proposed at the beginning of the work, as well as to put the test suggested hypothesis.

The empirical analysis proved more challenging than the theoretical one. The data sourcing phase did not bring any relevant difficulties and I was able to find extensive data on all the clubs I had selected for my study, as well as their financial data. The difficulty was not access to information, but rather the overload of data that I experienced. Being able to carefully select the right data, cleanse it and combine it was a long and exhaustive process that I believe I completed successfully. The next step was to select the right model for my analysis and it was at this point that the work done by previous authors and reviewed in the theoretical block had a protagonist role. I was able to select a model that had been useful for a study with similar goals and adapt it for the purposes of mine.

Overall, I am satisfied with the approach I took along the research process and how I structured my work. I was able to progress at a constant rhythm and my theoretical and empirical work was able to interact and interconnect, making my output more significant and meaningful. I can conclude that, in what concerns my **initial learning objectives**, I was able to **successfully** attain them.

8.2. Review of results

In the previous section, I concluded that my research project was able to attain the learning objectives I had set for myself in terms of progression, structure, methodology and learning experience. In this section, I will reflect on the results of my research and the hypotheses I had suggested.

One of the first elements to review in order to assess the success or not of the research are the results. If we take a detailed look into the results from my regression, it is clear than the models suggested did not accurately explain the effect observed on the stock returns. R^2 values were not high enough and the statistical significance of the models was, therefore, low. Furthermore, as I have described in the previous chapter, most of the variables I proposed did not pass the p-value tests and I had to accept the Null Hypotheses for them, rejecting the alternative hypotheses that they were statistically significant.

Despite the low statistical significance of the models, I was satisfied to observe that, indeed, the variable RES, both for wins and losses, seemed to have an effect on stock returns. I was able to reject the Null Hypotheses for the results variables and for the WND variable.

After making an overall assessment of the specific hypotheses I had proposed and the results of my research, I have to conclude that I cannot accept and confirm my Generic

Hypothesis that *there is a relationship between the results of professional sport matches and the stock price returns of the teams participating in them.* Even though I observed a positive and negative effect of wins and losses, the low R^2 values of the models, both for each individual stock and for the aggregated model for all clubs, leads me to conclude that I my results are not significant enough to make such statement. In conclusion, I have to **reject my Generic Hypothesis** based on my current research data.

In the next section, I will analyze what limitations I found during my work and what would be my next steps, if I continued my research, in order to improve the robustness of my model and in order to be achieve results that allowed me to conclude that I can accept my Generic Hypothesis.

8.3. Limitations and Next steps

One of the main limitations of this study, which I already anticipated in my first chapters, was the impossibility to assess the existence of a causal relationship between our two main variables: stock prices and the independent variables related to professional football matches and its results. Initially, this impossibility rose from the time and resources constraint of a master thesis. However, now that my research work has finished, I can conclude that this limitation also extends to the fact that the results of my research are inconclusive and hold low statistically significance in order to make conclusive statements.

This research process has been a wonderful learning opportunity that has made me reflect deeply on the steps I have taken and my approach to give an answer to my research question. This includes what I believe has been a good approach and rightful steps, as well as what I believe I could either have done differently or the next steps I would take in order to refine and perfect my work. If this research process were to be continued, my line of work would be as follows:

1. Firstly, I believe that **including an independent quantitative variable** that can be representative of **average market returns** would be helpful to improve the accuracy of the model. We would be able to discern better what is the percentage of the effect on the club stocks that we have been analyzing that comes from the market itself, and what part of the effect is explained by the variables I suggested. I already started some initial research for these next steps and I was able to identify a stock that matches my criteria: ***iShares Core MSCI ETF***[1], a fund traded in the NYSE Arca market that follows the MSCI Index for the European market, where all the clubs are geographically located and where eight out of nine stocks are traded. I believe that including this variable would highly increase the R^2 value from a multiple variable regression.

2. Secondly, I believe that **including an additional independent variable** that represents the **betting odds** for each of the games would improve the accuracy of the model. Betting odds have been proven to be a reliable proxy of expectations on match outcomes (Palomino, Renneboog & Zhang, 2009). Results that are opposed to these expectations could potentially lead to overreaction by investors. I deliberately avoided betting odds in order to avoid overlapping with previous authors research, but I believe now that doing some extra research on the topic could potentially lead to more extensive empirical knowledge on betting odds and raise interesting questions or insights.

3. Thirdly, I believe that **another interesting independent variable to include** into the analysis could be the **goal difference** of the result. The results, although they held low statistical significance, showed that wins and losses had a positive and negative effect respectively on stock price returns. I believe it could be interesting to dive deeper on this finding and analyze whether matches that have been won or lost by a significant margin, for example, a goal difference of three, have a stronger effect on stock returns than matches won or lost by a lower goal difference.

4. Lastly, I believe that another interesting point to develop are the **geographic differences between national leagues**. During my regressions I observed substantial differences in the $\boldsymbol{R^2}$ values and p-value tests among variables for each individual regression. Initially, I hypothesized that the difference between my first two analyses (Manchester United and Borussia Dortmund) could be explained by the fact that Manchester United is traded in the NY Stock Market and investors might not experience the full effect of the variables in the model. I expected the rest of regressions to be aligned

[1] https://www.marketbeat.com/stocks/NYSEARCA/IEUR/

with the Borussia Dortmund results. However, only Lazio seemed to have similar results and all the other clubs showed statistically insignificant results. For those reasons, I believe diving deeper on these differences could bring more insides and help us understand better how each of these stocks reacts to football matches results.

9. REFERENCES

9.1. Bibliography and Literature review

Ariely, D. (2008*). Predictably Irrational: The hidden forces that shape our decisions*. Harper Collins.

Arkes, H. R., Herren, L. T., & Isen, A. M. (1988). *The role of potential loss in the influence of affect on risk-taking behavior*. Organizational Behavior and Human Decision Processes, 42(2), 181–193.

Bias, P., Smith, P., Jansson, H. (2012). *In defense of the rationality assumption.* Research in Business and Economics Journal - Special Edition. Florida Economic Symposium

Chan, W. (2003). *Stock price reaction to news and no-news: drift and reversal after headlines*. Journal of Financial Economics, 70 (2), 223-260.

Edmans, A., García, D., & Norli, Ø. (2007). *Sports Sentiment and Stock Returns*. The Journal of Finance, 62 (4), 1967-1998.

Hirshleifer, D. (2001). *Investor Psychology and Asset Pricing*. AFA 2001 New Orleans Meetings.

Kahneman, D., Tversky, A. (1974*). Judgment under Uncertainty: Heuristics and Biases*. Science, New Series. 185 (4157). 1124-1131.

Karim, M.K. (2021). *Stock Market Analysis using linear regression.* Chongqing University.

Mitchell, Mark L. & Mulherin, H. (1994). *The impact of public information on the stock market*. Journal of Finance 49, 923-950.

Palomino, F., Renneboog, L., Zhang, C. (2009). *Information salience, investor sentiment, and stock returns: The case of British soccer betting.* Journal of Corporate Finance. 15 (3). 368-387.

Panwar, B. (2021). *Stock Market prediction using linear regression and SVM.* IEEE.

Shiller, R. (2000). *Irrational Exuberance*. Princeton University Press. Princeton NJ.

Wann, D. L., Dolan, T. J., McGeorge, K. K., & Allison, J. A. (1994). *Relationships between spectator identification and spectators' perceptions of influence, spectators' emotions, and competition outcome*. Journal of Sport & Exercise Psychology, *16*(4), 347–364.

9.2. Database sources

https://finance.yahoo.com

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

9.3. Data analysis tools

Microsoft Corporation. (2018). *Microsoft Excel*. URL https://office.microsoft.com/excel

R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/