



**HAL**  
open science

# Pricing football transfers: determinants, inflation, sustainability, and market impact: finance, economics, and machine learning approaches

Moussa Ezzeddine

## ► To cite this version:

Moussa Ezzeddine. Pricing football transfers: determinants, inflation, sustainability, and market impact: finance, economics, and machine learning approaches. Economics and Finance. Université Panthéon-Sorbonne - Paris I, 2020. English. NNT: 2020PA01E064 . tel-03171642

**HAL Id: tel-03171642**

**<https://theses.hal.science/tel-03171642v1>**

Submitted on 17 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Doctoral Dissertation  
PhD in Economics**

**Moussa Ezzeddine**

17<sup>th</sup> December 2020

---

**Pricing Football Transfers: Determinants, Inflation,  
Sustainability, and Market Impact.**

**Finance, Economics, & Machine Learning Approaches**

---

**Thesis Director:**

**Pierre-Charles Pradier**

**Jury:**

- **Luc Arrondel**, directeur de recherches de CNRS, président,
- **Robert Simmons**, professor at Lancaster University, rapporteur,
- **Nicolas Scelles**, senior lecturer at Manchester Metropolitan University, rapporteur,
- **Emilio Garcia**, directeur juridique de FIFA, suffragant,
- **Wladimir Andreff**, président du conseil scientifique de l'observatoire de l'économie du sport, suffragant,
- **Pierre-Charles Pradier**, MCF at UP1, directeur des recherches.

## **Declaration**

I declare that this thesis is my research carried out in France at the University of Paris 1 – Pantheon Sorbonne under the supervision of Professor Pierre-Charles Pradier of the same university. This thesis has not been shared or submitted in any other institution for any other award or degree. Cited References have been fully acknowledged in this research.

**Moussa Kazem Ezzeddine**

## **Abstract**

Each year new transfer market news tops headlines due to the astronomical prices paid to recruit a superstar by top football clubs. The money paid by the buying club is assumed to be an estimate of the market value of the transferred player. Thus, the challenge is to determine the significant factors that affect the pricing function of a football player. In this research, a large data set has been extracted containing more than 87,000 transfers and more than 200,000 wage observation alongside two sets of variables; one contains real statistics of each player from the previous two seasons, while the other contains synthetic scores given by experts. This work has made use of one hedonic pricing function and three machine learning algorithms to estimate the most important factors affecting the financial value of the player. Albeit imperfect, but the models can predict the pricing functions of the transfer fees and wages with different promising precisions. Finally, a market model has been carried out to determine the effect of transfers, surprising match results, and COVID-19 on the market value of a football club. The overall findings were promising as they have provided interesting explanations about the different segmentations in the transfer market and the effectivity of transfers on the fluctuations of the share values of certain clubs.

## Dedication

I would like to thank my supervisor Professor Pierre-Charles Pradier for his outstanding contributions throughout the stages of the PhD research. I thank him for trusting my capabilities and for enrolling me for the PhD program at the *Universite Paris 1 – Pantheon Sorbonne*. His continuous support, time dedication, immense knowledge and valuable advices have kept me motivated, and I cannot appreciate his support enough neither I could imagine a better advisor.

Besides my supervisor, I would like to thank my parents for the enormous support they provided. My mother's prayers and my father's physical and fiscal contributions have insured my full-time safe stay in France. Without their contributions I couldn't have made it to this stage. In fact, no words can serve me to describe their unlimited support. Moreover, I'd like to thank my brother and two sisters for their spiritual support throughout the journey.

I can't forget to pay a tribute to the late lawyer Paul Kudjokoma for his unconditional and priceless support, alongside his colleague lawyer Kenneth Kodjorgie, during the most critical and challenging working incidents I faced between 2013 – 2015 in Ghana... An experience which has reshaped my career and probably contributed to the decision of moving forward towards higher education.

To Professor Pradier, my family, lawyers, colleagues and friends, I dedicate this humble work.

## Acknowledgement

My sincere thanks go to emeritus professor Wladimir Andreff for his great and continuous support, insightful thoughts, and his participation in my thesis committee for two consecutive years, covering long distances to be present. Equally, I would like to thank emeritus professor Francois Gardes for his great technical support, helpful comments, dedication, and participation in the thesis committee. I would like to thank the jury members for accepting to be part of the jury. To professors Luc Arrondel, Robert Simmons and Nicolas Scelles, I express my sincere gratitude for the added value triggered by their participation due to their outstanding academic records and achievements. Special appreciation to Dr. Emilio Garcia, chief legal officer at FIFA, for joining this jury since his presence creates a link between research and industry due to his wide knowledge and experience that is highly respected by the stakeholders of the industry. Overall, I feel proud to present my thesis before such a great competent jury...

I would like to thank professor Mathilde Maurel for the advices she gave me during the admission period and her subsequent support... Besides her, I would like to thank professor Marc-Arthur Diaye who has always been a good friend and a mentor. My French language professor, Mdm. Laurance Rico, deserves praise for improving my French speaking skills. Moreover, I am thankful for the technical unlimited support I have received from Dr. Peter Martey Addo and Maxime Nicolas. Additionally, I can't forget to thank all my fellow labmates especially Daniel Bastides, Davide Fortin, Mona Barake, Gaspard Tissandier, Yann Lecorps, Adham Jaber, Bin Wang, Zeineb Charif, Yasmine ElKhatib and Aref Arkadani... and other school friends like Lyes Zemouche for the good memories we have shared. Finally, I would like to thank the doctoral school of Economics, Centre d'Economie de la Sorbonne (CES), and the university of Paris 1 – Pantheon Sorbonne for providing me the opportunity to join a great academic institution giving me access to its research facilities and laboratories.

## Table of Contents

Jury:.....	i
Declaration.....	ii
Abstract.....	iii
Dedication.....	iv
Acknowledgement.....	v
<b>Table of Contents</b> .....	<b>vi</b>
List of Figures.....	ix
List of Tables.....	x
CHAPTER ONE.....	1
1 INTRODUCTION.....	1
1.1 Background.....	1
CHAPTER TWO.....	9
2 LITERATURE REVIEW.....	9
2.1 Introduction.....	9
2.2 Determination of Football Transfer Fees.....	9
2.2.1 Methodological questions.....	9
2.2.2 Valuation techniques and results.....	11
2.2.3 An Enduring Problem – the selection bias.....	14
2.3 Valuing Football Transfers.....	19
2.4 Machine Learning Techniques and Football Pricing.....	22
2.5 Stock Market.....	24
CHAPTER THREE.....	27
3 PRICING FOOTBALL TRANSFERS: FINANCIAL ECONOMETRIC APPROACH.....	27
3.1 Introduction.....	27
3.2 An Enduring Problem – the selection bias.....	29
3.3 The hypothesis to Solve Selection Bias.....	31
3.4 Data.....	34
3.5 Estimation Strategy and Findings.....	38
3.5.1 Micro-Level Results.....	39
3.5.2 Price Index.....	53
3.6 Conclusion.....	56
CHAPTER FOUR.....	57
4 PRICING FOOTBALL TRANSFERS: AN EXPLORATORY STUDY INTO THE GLOBAL MARKET USING SYNTHETIC DATA.....	57

4.1	Introduction .....	57
4.2	Data .....	59
4.3	Getting around selection issues .....	61
4.4	Estimation and results .....	66
4.5	Conclusion .....	75
CHAPTER FIVE .....		76
5	PRICING FOOTBALL TRANSFERS 2007-2018: FROM ECONOMETRICS TO MACHINE LEARNING ..	76
5.1	Introduction .....	76
5.2	The Data .....	77
5.2.1	Data structure .....	77
5.2.2	Descriptive Statistics .....	79
5.3	Methodology.....	80
5.3.1	Generalised Linear Models (GLM) .....	81
5.3.2	Random Forest (RF).....	82
5.3.3	Gradient Boosting Machines (GBM) .....	82
5.4	Results.....	83
5.4.1	Transfer Fees.....	83
5.4.2	Wages.....	88
5.5	Conclusion.....	92
CHAPTER SIX.....		94
6	FOOTBALL IN STOCK MARKET: TRANSFERS AND EUROPEAN COMPETITIONS EVENTS.....	94
6.1	Introduction .....	94
6.2	Data .....	97
6.3	Methodology.....	99
6.3.1	Hypothesis Testing .....	100
6.3.2	Surprise on match results .....	100
6.3.3	Surprise on transfer prices.....	102
6.3.4	COVID-19 Effect .....	103
6.3.5	Market Model .....	105
6.3.6	Robustness Check .....	107
6.4	Results.....	107
6.4.1	Expected Match Result .....	107
6.4.2	Unexpected Match Events .....	108
6.4.3	Transfers Events.....	113
6.4.4	COVID 19 Events .....	119
6.5	Conclusion.....	126



CHAPTER SEVEN .....	128
7 MAIN FINDINGS AND CONCLUSION.....	128
7.1 Introduction .....	128
7.2 Challenges .....	128
7.2.1 Data Issues .....	128
7.2.2 Econometric Issues .....	129
7.2.3 New Pricing Functions.....	129
7.3 Main Findings.....	129
7.3.1 Variables.....	129
7.3.2 Market Segmentation .....	131
7.3.3 From Econometrics to Machine Learning .....	132
7.3.4 Football events' weight in Stock Market .....	133
7.4 Recommendations .....	133
7.5 Future Scopes.....	134
8 REFERENCES .....	136
9 Appendices.....	144
9.1 Appendix 1: Tables and Figures for Chapter 3 .....	144
9.2 Appendix 2: Tables and Figure for Chapter 4.....	156
9.3 Appendix 3: Tables and Figures for Chapter 5 .....	163

## List of Figures

Figure 1.1: Spending growth (FIFA TMS report, 2019) .....	1
Figure 1.2: Maximum transfer fees in £2019 PPP.....	2
Figure 3.1: Price Index.....	54
Figure 4.1: Indices average salaries per age per season.....	64
Figure 5.1: Discrete vs Continuous data.....	79
Figure 5.2: GLM model – Most important variables affecting Transfer fee .....	86
Figure 5.3(a): RF model – Most important variables affecting Transfer fee.....	87
Figure 5.3 (b): GBM model – Most important variables affecting Transfer fee .....	87
Figure 5.4: GLM model – Most important variables affecting Wages.....	91
Figure 5.5: RF model – Most important variables affecting Wages.....	91
Figure 5.6: GBM model – Most important variables affecting Wages.....	92
Figure 6.1: JUVENTUS & MANCHESTER UNITED’s stocks percentage change. Source: Google Finance .....	96
Figure 6.2: JUVENTUS’s stock behaviour. Source: Google Finance.....	96
Figure 6.3: Manchester United’s stock behaviour. Source: Google Finance .....	96
Figure 6.4: Positive (blue) & Negative (orange) Surprises.....	108
Figure 6.5: Positive (blue) & Negative (orange) Surprises with prediction ratio > 2.....	109
Figure 6.6: All Incoming Transfers with different event windows .....	115
Figure 6.7: Incoming Transfers effect on different clubs’ share prices.....	115

## List of Tables

Table 1.1: AAGR of Max transfer fees in £2019 PPP .....	2
Table 1.2: Transfers’ spending in top 5 Leagues. Source: FIFA TMS seasonal reports .....	3
Table 3.1: Top transfers in the last two decades. Source: Transfermarkt.....	29
Table 3.2: Total transfers per year .....	34
Table 3.3: Frequency of Selling and Buying Contracts per Country respectively .....	35
Table 3.4: Percentage of Import/Export contracts out of Selling and Buying Contracts per Country respectively .....	36
Table 3.5: Basic Tests using “lfee” (transfer fee). Samples: Universal/Continental .....	44
Table 3.6: “Continental Model”: Tests across continents/regions.....	46
Table 3.7: “Countries Model”: Tests across countries.....	48
Table 3.8: “Universal Model”: Tests across Positions.....	50
Table 3.9: “European Model”: Tests across Positions.....	52
Table 3.10: “Latin American Model”: Tests across Positions.....	53
Table 3.11: Price Indices Computation.....	55
Table 4.1: Balance of Trade (2007 – 2020). Source: Transfermarkt.....	58
Table 4.2: Unique observations in our database per country 2007-2019 .....	60
Table 4.3: Regression with the transfer fee being the dependent variable .....	69
Table 4.4 : Global model + breakdown by continent for objective package / subjective complete package / subjective salary package.....	71
Table 4.5: Breakdown by position and continent for the subjective complete package – forward players .....	72
Table 4.6: Breakdown by position and continent for the subjective complete package - midfielders .....	73
Table 4.7: breakdown by position and continent for the subjective complete package - defenders .....	74
Table 5.1: Players’ attributes and salaries. Source: SOFIFA.....	76
Table 5.2: Players per country in the database.....	78
Table 5.3: Observations’ and variables’ breakdown .....	79
Table 5.4: Descriptive statistics for some effective variables.....	80
Table 5.5: Three Models test results – Transfer Fees. ....	85
Table 5.6: Variables’ Importance (Descending Order) per three Models – Transfer Fees.....	86
Table 5.7: Three Models test results - Wages.....	89
Table 5.8: Variables’ Importance per three Models – Wages .....	90
Table 6.1: Data coverage period .....	97
Table 6.2: Unexpected Match Events .....	98
Table 6.3: Transfers Events above £15m 2007-2018 .....	98
Table 6.4: Examples of Over-priced Transfers. Source: Transfermarkt.....	99
Table 6.5: Examples of Under-priced Transfers. Source: Transfermarkt.....	99
Table 6.6: Surprise according to pre-match odds .....	101
Table 6.7: Examples of Over-priced Transfers with different Prediction Error Categories..	103
Table 6.8: Examples of Under-priced Transfers with different Prediction Error Categories. .....	103
Table 6.9: List of Historical Financial Market Crashes. Source: Spindices .....	104
Table 6.10: Examples of some league-related events and their occurrence dates .....	104
Table 6.11: Events covered by this study/analysis .....	105
Table 6.12: Abnormal Returns-National League Matches on the first trading day.....	109
Table 6.13: Negative Surprise (Loss instead of Win).....	111
Table 6.14: Negative Surprise (Draw instead of Win) .....	112

Table 6.15: Positive Surprise (Win instead of Loss) .....	113
Table 6.16: Transfers Abnormal Returns at t=0 .....	116
Table 6.17: Transfers events .....	117
Table 6.18: Transfers [-5,+5] .....	117
Table 6.19: Over-priced Transfers .....	118
Table 6.20: Under-priced Transfers .....	119
Table 6.21: Effect of Italian PM Decree on the Italian Clubs .....	122
Table 6.22: Effect of Italian PM Decree on the Top Five Leagues .....	123
Table 6.23: Effect of Italian PM Decree on the All Clubs.....	124
Table 6.24: Effect of Italian League suspension on Italian Clubs .....	125

# CHAPTER ONE

## 1 INTRODUCTION

### 1.1 Background

Every summer football fans, analysts, media, and researchers attend eagerly to the rumours and indicators of possible new signings that could surpass the preceding transfer window. The Mercato has become an amusing festival for the stakeholders simply because it engages big names and significant amounts of money, boosting trade wars for talents' acquisition among clubs. Such wars have fuelled the speculation and affected the valuation of transfer fees and salaries of top football players among top football clubs. Media coverage, including social media, has also played a crucial role in reshaping the financial football model which has been evolving over the years influencing top managements' decisions and fanbases' reactions on certain transfers. In this context, the French giant PSG spent €220 M (£200 million) to lure the services of FC Barcelona star Neymar da Silva Santos Júnior in summer 2017, the highest ever amount paid for gaining the services of a football player. An all-time high, historical record, sale forced by meeting the player's release clause has led to a panic-buying reaction by the Spanish giant FC Barcelona triggering a spending spree of €145 M and €144 M for the purchases of Ousmane Dembele and Philippe Coutinho respectively. Figure 1.1 reveals the growing expenditure of the top five European leagues' football clubs.

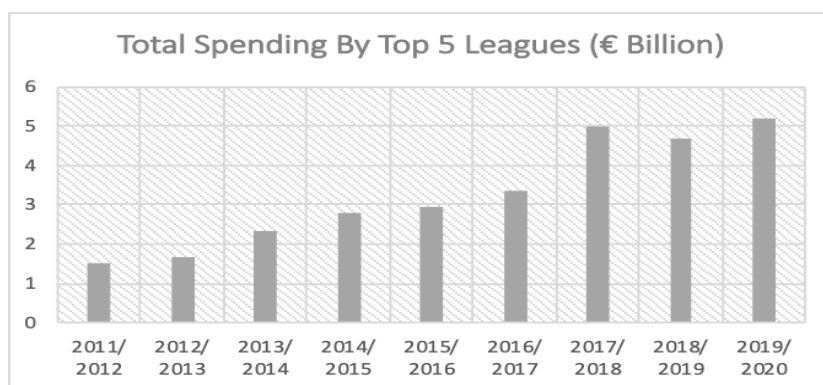


Figure 1.1: Spending growth (FIFA TMS report, 2019)

In general, the determination of transfer fees for professional footballers has attracted a lot of attention in recent times, not limited to Neymar’s move to PSG, since the acceleration of growth in transfer fees is spectacular, from tens of thousands of pounds in the 1960s to tens of millions in the 2000s. Since the late nineteenth century, the average annual growth rate (AAGR) of the most expensive transfer fee has been around 8%, with an apparent increase since the Bosman ruling (Figure 1.2 and Tables 1.1 and 1.2).



Figure 1.2: Maximum transfer fees in £2019 PPP

Period	Average Annual Growth Rate	Comment
1894-2017	8.1%	From the first paid transfer to the record transfer
1967-2017	9.1%	Last 50 years
1992-2017	11.5%	Last 25 years
<b>1995-2017</b>	11.8%	Since the Bosman ruling
2007-2017	11.7%	Last 10 years

Table 1.1: AAGR of Max transfer fees in £2019 PPP

Season	Summer Spending (€Billion)	Winter Spending (€Billion)	Spending (€Billion)	Seasonal Growth (%)	Total Paid Transfers (€M)	Avg. per Transfer (€M)	Seasonal Growth (%)	Total Free Transfers
2011/2012	1.3	0.24	1.54		434	3.55		971
2012/2013	1.24	0.44	1.68	9%	481	3.49	-2%	947
2013/2014	2.01	0.33	2.34	39%	509	4.6	32%	1056
2014/2015	2.34	0.47	2.81	20%	541	5.19	13%	1195
2015/2016	2.59	0.37	2.96	5%	613	4.83	-7%	1312
2016/2017	2.78	0.56	3.34	13%	702	4.76	-1%	1348
2017/2018	3.95	1.02	4.97	49%	766	6.49	36%	1389
2018/2019	4.05	0.65	4.7	-5%	754	6.23	-4%	1487
2019/2020	4.38	0.82	5.2	11%	777	6.69	7%	1605
Total	24.64	4.9	29.54	16% (CAGR)	5577	45.83	74% (CAGR)	11310
Overall Average	2.74	0.54	3.28	17.62%	619.67	5.09	9.25%	1256.67

Table 1.2: Transfers' spending in top 5 Leagues. Source: FIFA TMS seasonal reports

The Bosman ruling (1995) by the European Court of Justice (C-415/93) has liberated the players from being constrained by their clubs, after the expiry of their contracts, giving a boost to the competitive transfer market and rise to the transfer fees and salaries (Simmons, 1997; Antonioni and Cubbin, 2000; Ericson, 2000; Feess and Muehlheusser, 2003; Tervio, 2004). After that, transfer fees and salaries were scoring record highs gradually on an annual or biannual basis to feed the talent acquisition war fuelling the winning maximisation objectives of football clubs (Garcia Del Barrio, 2009; Kesenne, 2006). Whether we measure the inflation of transfer prices on the record transfer (Figure 1.2.), on the total volume or on the average value of transfer (Table 1.2), it seems that transfer prices have accelerated over the last century to a growth rate which is unsustainable because it exceeds by far the growth rate of the economy. Rising club revenues might cause rising transfer prices, while the characteristics of the players are unchanged, in which case it would be inflation, perhaps even a bubble if inflation is accelerating; on the other hand, the rise in transfer price might come from an increase in quality of players transferred, in which case there would be no *true* price inflation. Hence we shall ask whether we can explain the determinants of the transfer prices and can

compute more precisely a price index for transfers: this would, in turn, help us figure out more precisely whether the evolution of transfer prices and transfer money is sustainable.

As it is recognised for its uncertainty, sports economists tried to dig into the bank of secrets that could determine the factors affecting the movements of transfer fees, salaries and other allowances implicitly or explicitly. In principle, transfer fees do not correspond to the total market value of the player though, and they are only a feature of the contract between the player and his club. Additionally, the player is paid a salary plus possible signing-on and other contingent bonuses, but he cannot leave before the expiration of his contract without paying to his club an amount the club agrees on; generally, this amount is paid by a buying club as transfer fees. What most experts are interested in were the factors that determine such fees. Some economists built their theoretical model based on the monopoly economic theory to say that deciding the transfer fee is a bargain between the selling and buying club (Carmichael, 2006), in studying the characteristics of each club. Another approach was through the perfect competition economic theory by which transfer fees were determined by the skills and characteristics of the players (Carmichael, 2006). Some studies refer to it as hedonic pricing (Carmichael and Thomas, 1993; Dobson and Gerrard, 1999). From an economic point of view, the value generated by a player breaks down into sporting performance (including qualification for higher leagues and championships which bring higher income) and image (which attracts audience and sponsors). It is difficult to determine precisely the value of a single player since football is a game opposing entire squads (and their respective staves) and the production function is not likely to be separable or easily estimated. But it seems evident that teams cannot on average pay more than the value of their players without being bankrupt, even under “soft budget constraint” (Andreff, 2015). Hence the market value of players is broken down between the club and the players, and transfer fees are part of what clubs can get from player value. Our



main aim will be to show that a reasonably stable pricing function can be estimated: this will require both a proper procedure and a precise segmentation.

As the incremental spending to sign best players has become normal behaviour for clubs seeking winning/utility maximization objectives (Kesenne, 2006), the increasing data suppliers have offered bigger opportunities for economists seeking to understand such spending behaviour, giving bigger chances to reveal the dynamics of players' movements in the transfer market. Like previous studies, we try, in this study too, to follow the roadmap that many previous economists followed to determine the main factors that influence the transfer fees and salaries paid by the buying and selling clubs. A larger data set extending the limits, beyond the top five European leagues, to the far west (Latin American countries and the Major League Soccer (MLS)) and the far east (Chinese Super League) has been utilised to dig deeper into the understanding of the price function. A remarkable breakthrough was the volume of data containing information about both transfer fees and salaries. The data set helped not only to explore across a large number of countries but also across positions. A strategy we used to contribute further into tackling two econometric flows: selection bias and heteroscedasticity; which were hardly approached in the previous studies to broaden the limits of the pricing function into larger scopes by introducing more dependent and independent variables.

In the last decades, data science has emerged as modern science that can help many disciplines, including the ones that depended on econometrics and applied mathematics to analyse the past, where sports is no exception. The notable increase in open data sources has offered sports researchers more scopes for research with larger and more segmented data sets. In this context, some researchers tried to approach the uncertainty of football industry through machine learning techniques, applications of artificial intelligence, in estimation models (i.e. pricing

options and assets) replacing some econometric models still in its early stage though. Therefore, using data science in the search for the appropriate pricing methodology is still questionable for some stakeholders who are raising fears whether the data utilized can be tuned well to produce the best approximation for fees or wages. More recent machine learning algorithms like Generalized Linear Models (GLM), Random Forest (RM), Gradient Boosting Methods and many more, were used in the thesis in estimation models trying to estimate players' transfer fees and salaries. Such a prediction could help stakeholders in their club's or media/sponsorship contracts negotiations. The main question that could be asked is to what extent can such new techniques explain the dynamics of the pricing functions, especially for players superstars? The answer to such a question will be answered in the subsequent chapters.

To summarize, economists and mathematicians have introduced different estimation models. Some models differed by algorithms, while others differed by the type and volume of data sets. Most estimates relied on studying the factors affecting the transfer fees due to the shortage of open data sources about salaries. The extra mile that this study has achieved was the rich data sets aggregated: so far, empirical studies of transfer fees used rather small samples of *local* significance and rarely managed to prove representative. For instance, the study by CIES (2017) features "a sample of 1,648 paid transfers", which is both small (since it is around 2% of transfers disclosed on *Transfermarkt*) and likely to be affected by selection bias (since only paid transfers have been selected). To assess a correct pricing function for football players, we must build a sample which more or less represents the global market (a) and is not stained by selection problems (b). This is what we tried to achieve using a so-far untapped database and appending other data from public sources to form a 10,000+ observations base as a proof of concept. Using this base, we intend to show that pricing functions have a global significance and can satisfactorily be measured. In approaching football players' pricing models, the

hedonic pricing methodology was utilized first then machine learning techniques were involved later in a separate model to provide different exploratory evidence. Such approaches were supported by a price index that correctly measures inflation in the market.

Inflation in the market of football players is a controversial topic which has been discussed and expressed publicly in scientific papers and the sports medium. Such topic has been watched with great interest by experts who have been linking it to the augmentation of clubs' finances and sources of income, especially TV revenues. As sources of income evolved over the years, many clubs found the stock market as a new source of income which can help them in fundraising and in finding new financiers and investors. According to Dobson and Goddard (2001), the growth of revenues can directly have impacts on sports performance. More recently, some professional football clubs used listing in the stock market as a source of funding; and Tottenham Hotspurs was the first to be listed in 1983.

It is normal to see top clubs earning more revenues than other clubs since they receive higher prize money when achieving pole positions in local leagues or European competitions like UEFA Champions League or Europa League, but what's more interesting to know and analyse is investors behaviour (through the stock market), especially in the last two decades, where we have witnessed a large number of club acquisitions by business people or business parties, e.g. Manchester United's acquisition by the Glazers family and many more. The stock market behaviour differs from one football club to another. Investors usually look at the events that affect market behaviour such as match results, players' transfers (incoming or outgoing transfers, especially superstars) and many other events.

To understand how football events, impact the market success (failure) of football clubs, an event study using the market model was designed to serve that purpose. These types of debates are not new as previous studies have supported it (see, for instance, the works of Lehman and Weigand, 1997; Stadtmann, 2003, 2006). These studies have confirmed a positive effect of winning matches on the market returns, and the case of Borussia Dortmund is used as an example. This study depends on an extensive data set gathered from *bloomberg.com* for the duration between the years 2000-2019. The data contains a comprehensive number of matches and competitions for six football clubs: Manchester United, Juventus, AS Roma, Ajax Amsterdam, Borussia Dortmund, and Galatasaray. Moreover, the betting odds that were extracted for national leagues matches helped to study the effect of surprises caused by wins, draws, and losses on the abnormal returns.

There was an opportunity to study the effect of mega transfers (Paul Pogba from Juventus to Manchester United and Cristiano Ronaldo from Real Madrid to Juventus), among other transfers made by their purchasing clubs, on the market shares and revenues of the relevant clubs. Although the existing literature about transfers' events is relatively poor, such events from the booming transfer market that surprises us every season with controversial superstars' transfers and rumours are an essential topic to study. For instance, Cristiano Ronaldo's move from Real Madrid to Juventus has helped the share price of Juventus to surge remarkably before and after the signing day. As a result of the study, a positive effect for incoming transfers (purchases) on the abnormal returns of the listed football clubs was found knowing that such impact has been denied by previous research (Fotaki et al. 2009; Bell et al. 2012).

## CHAPTER TWO

### 2 LITERATURE REVIEW

#### 2.1 Introduction

This thesis was based on different, but conceptually linked, bodies of literature attempting to understand football pricing. While the first set of literature emphasized the determination of football transfer fees, the second set of literature focused on valuing football transfers, the third set of literature focused on the use of machine learning techniques to understand football pricing. The last collection of literature approached the stock market effects of match surprises and transfers of superstars on the market value of football clubs. This chapter reviews the theoretical and empirical literature underpinning this literature to provide the relevant yardsticks to be used in discussing the main results. The chapter is organised into five sections, with each section discussing the literature for the subsequent chapters, respectively.

#### 2.2 Determination of Football Transfer Fees

Numerous studies have been trying to understand what factors underpin the determination of football transfer fees. We provide an overview of this literature under the following sub-sections: methodological questions, the techniques and their results before we point out open questions and enduring problems.

##### 2.2.1 Methodological questions

The existing studies mostly use the degree of competition in the football transfer market as an essential starting point (Carmichael, 2006). Given the typical nature of competition in economic markets, there are diverse views about competition in the football transfer market and thus, the factors that determine transfer fees.

On one extreme, there is the view that there is a monopoly in the football transfer market that is driven by clubs and players (Carmichael, 2006). An example of the monopoly market is when a single club seeks to acquire a particular player or even when it involves few clubs. With this view, transfer fees are negotiated based on a bargain between clubs and players (Carmichael and Thomas, 1993; Riley and Witt, 1995; Dobson and Gerrard, 1999 and Gerrard and Dobson, 2000). This bargain is deemed appropriate because the transfer market is characterised by uncertainty because of the asymmetric information aspects of a player's quality and commitment and there is a risk because it will be unclear, before a transfer, how well a player will perform in the new team (Carmichael, 2006). As with the determination of wages in a typical labour market, a bargain (in this case) involves the setting of upper (expected value of playing services to the buying club) and lower (value of the player to the selling club) limits of the reservation fees of the buyer and seller (Carmichael, 2006). The actual outcome of the bargain depends on each club's bargaining power which is further dependent on the alternative players available, player's skills and ability, club's marginal valuation of player's talent and the degree of monopsony power exercised by the club (Carmichael, 2006).

On the other extreme, there is another view that the transfer market is competitive, and transfer fees are determined through a competitive process (Carmichael, 2006). Unlike the case of the monopoly market, the competitive market exists because the transfer market can be characterised by freedom of contracts, considerable potential for movement, many buyers and sellers and the availability of comprehensive information about the performance of players (Carmichael, 1999). For instance, Carmichael (2006) cites the incoming of the Bosman ruling in 1995 on transfer markets to have made the contemporary football transfer market very competitive. Before the Bosman ruling (1995), the club retained a right on player transfer even after the expiration of contracts; since the Bosman ruling, though, expiration of contract

releases a player from his duty to his club. Szymanski and Smith (1997) describe the process of football transfer fee determination in a competitive market to depend on the innate ability and human capital of players, which are reflected mainly in their marginal revenue product at the selling club. In this model, players do not only wish to play for more successful teams, but the latter also want to be associated with the former, and this becomes a significant factor in the determination of transfer fees.

Empirically, the specification of econometric models determining transfer fees follows the particular market in question. For instance, in monopoly markets, since a bargain is always reached between the buying and the selling clubs, the typical independent variables relied on are the innate characteristics of the buying and the selling clubs such as team performance and market size (Vrooman, 1996). This feature makes the monopoly market unique. For the competitive model, the typical independent variables used include the ability and human capital of players: age, experience (number of league appearances), goal record, the position being played, international appearances, selling club's status and performances, divisional standing and number of goals.

### **2.2.2 Valuation techniques and results**

The transfer prices of football players have become a significant concern for the industry, and consequently, there is a demand for explaining the determinants of transfer prices. Economists relied on econometric estimations of "hedonic" equations to provide a model of transfer price formation, which brought interesting results as it has done in many other fields. Kanyinda et al. (2012) recall the problems with previously used methods, mostly derived from naïve accounting and thus inconsistent with the time value of money or advanced (fair value)

accounting methodologies. They introduce themselves a real options method which, albeit appealing, is not tested on a proper sample but a handful of examples.

The canonical example of hedonic pricing is provided by Dobson and Gerrard (1999). They developed a model of the player-transfer market in soccer, in which observed transfer fees are determined by player characteristics, selling-club characteristics, buying-club characteristics and time effects. Similar models were estimated by Dobson, Gerrard and Howe (2000) and Szymanski (2015). Dobson and Gerrard (1999) model can also be used to investigate the rate of inflation in transfer fees. Despite this achievement, the review of the literature suggests that there is scarcity in the availability of data on the returns made by football players, their salaries, and their transfer fees once transferred from one club to another. Currently, the popularly cited sources are those for England and Germany; probably because of easy access to data.

The results showed that the factors that determined transfer fees varied markedly among market segments, which can be distinguished along several dimensions: geographical (regions or countries), sporting level (leagues), or through in-game positions. Poli, Ravenel, and Besson (2017) used the most advanced data to date to approach the transfer fees evaluation by two steps using multiple linear regression in each step. The first step estimates the status of buying clubs based on different variables characteristics; age, position, league, activity in clubs' and national teams' competitions. The second step evaluates footballers' characteristics: age, duration of the contract, level of buying club (estimated). They confirmed the different evaluation model between positions but did not conclude on geographical segmentation.

Typically, the variables found to be statistically significant in determining football transfer prices include goal-scoring record, league appearances, international appearances, age (and



sometimes age squared) (Carmichael, 1999; Dobson, Gerrard and Howe, 2000; Ruijg & van Ophem, 2014). Although transfer fees evaluation may differ from salaries evaluation by shape, they may intersect when the evaluation is based on individual characteristics. Mincer (1974) created a linear regression model to formulate the logarithmic function of the salaries in terms of individual (players) characteristics and enterprise (club) characteristics. Lucifora and Simmons (2003) did a regression model of the logarithmic salaries function in terms of four variables: experience determined by age and number of games played; performance determined by the number of goals and passes; reputation determined by the number of international games played with the national team; and determined by club's financial status and club's manager (coach) rating and performance. Age and reputation coefficients were proved significant. Age has been proved vital in many studies.

Ruijg and Ophem (2014) proved that age (up to 26 years old) and the number of games played have a significant effect on the transfer fees of footballers. Dobson, Gerrard and Howe (2000) found player characteristics, time effects, selling-club characteristics and buying-club characteristics to determine transfer prices in English Professional Football. Frick (2007), like Speight and Thomas (1997) and Carmichael et al. (1999) stated that many common significant variables are affecting both players' salaries and players' transfer fees like players' age, the number of games played with a club, international caps (games with the national team).

More general results have been found since Szymanski and Kuypers (1999) extended the discussion on transfer fees by proving two important propositions, which have been well known to spectators and followers of English football. First, they showed that teams which spend more on players would enjoy more on-field success and second, league position determines club income. Szymanski (2015) also showed that the transfer system sustains the

dominance of the elite clubs by ensuring that they are the only ones with the financial muscle to afford the transfer fees payable for the very best players. As it currently operates, the system is not only unfair to players but promotes the opposite of what was intended.

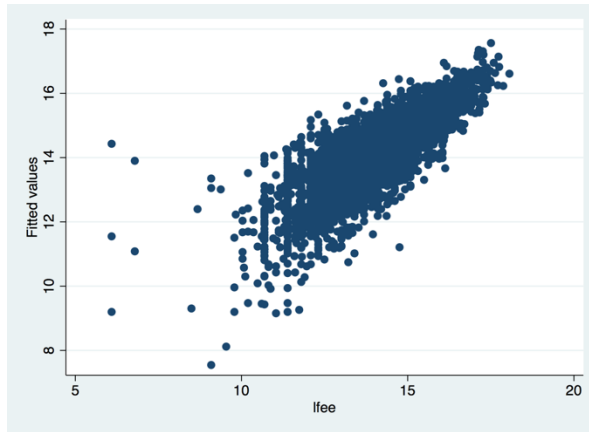
Many questions remain open, as they have not been addressed in a completely satisfactory way by the literature to date. For instance, Pedace (2007) noticed a premium for Latin American players in the European championship, the idea being that hiring Latin American players raise interest in their home continent. More general questions include the precise nature of the superstar effect in football or the distinction of consistent market segments for transfer pricing. These can only be addressed with consistent data *and* methods. Poli et al. (2017) have made the most sophisticated example to date of hedonic pricing, using expensive data provided by specialized firms as input. Unfortunately, they are affected by a selection bias they cannot measure, albeit, it has been diagnosed by Carmichael et al. as soon as 1999.

### **2.2.3 An Enduring Problem – the selection bias**

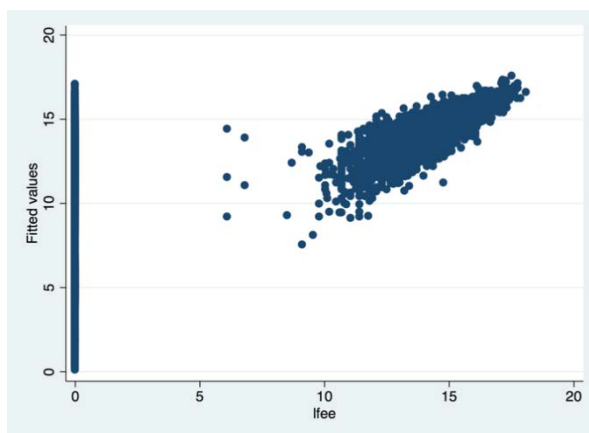
Selection bias is likely to be a problem undermining the currently used transfer fee valuation techniques. Let us start from the output of a standard hedonic regression. For instance, the figure below shows the correlation between the (logarithm of) actual transfer fee and the (logarithm of) the transfer fee computed by one such model<sup>1</sup>:

---

<sup>1</sup> The figure displays the predicted vs. actual price of defenders and defensive midfielders between 2007 and 2018 for the whole dataset referred to as stage 1 below.



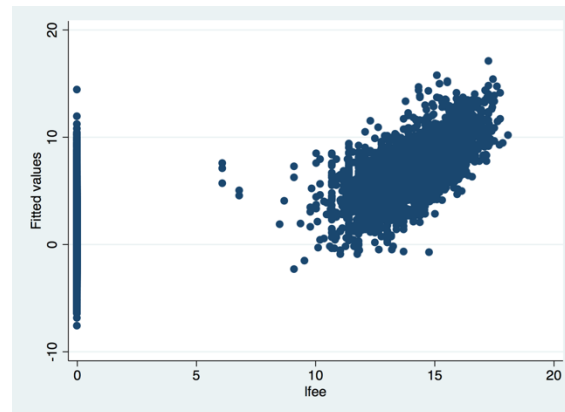
The fit is quite good here, maybe not as good as Poli, Ravenel, and Besson (2017), who claimed that their models gave  $R^2$  above 80%. This is to say, their model provides an account for more than 80% of the observed differences in the observed transfer fees (or, more precisely, the squares of the logarithms of those fees). And this is a very good result for a statistical estimation procedure. Unfortunately, this direct approach to the question suffers a so-called “sampling bias”, since only players transferring for a strictly positive transfer price are taken into account in the econometric regression. The consequence of this bias is that, if we try to value a player transferring for free with the pricing function derived from the model, this player will be given a strictly positive price. The model then overvalues most players, as the following figure shows:



The line along the left Y-axis is the set of players who transferred for free and are given a positive price by the model. The cluster on the right side of the picture is the same as on picture 1, although the slope has been changed to accommodate all the new observations which were not taken into account by the first procedure. One may object that this apparent problem can be solved by including a 'free transfer' dummy: when their contract is over, players transfer for free. Although this is a bit adhoc since it does not provide for a consistent econometric procedure, this approach would not be enough to tackle player loans during their contract: some players are lent by their club to another, sometimes for a fee and sometimes for free. The classical model cannot cope with these subtleties.

Apart from a line on a figure, how embarrassing is the selection bias? One can think from a quantitative point of view: in our samples, around 85% of players transfer for free and are thus assigned a false price by the model. *In more than 10% of the cases, the false price is in excess of 1 million euros*, which is far from being negligible. But more generally, the selection bias leads to a logical issue: while in real life some players transfer for a fee and some transfer for free, the model can only say for what fee a player *that transfer for a fee* should transfer. But, when the model assigns the same value to two players, one cannot say whether both will transfer for that fee, or only one of them, or none. Since 85% of players are assigned a wrong positive price, one can say there is a 72% chance on average that none will transfer (although this probability is conditional to the transfer fee computed by the model and could thus be refined) and 25% that only one will. In other words, what the model can do is compute the values of *past* transfers but not predict future values, nor tell who will transfer, nor even tell who transferred for a fee. The selection bias is thus very annoying since it entirely voids the predictive value of a model.

How shall we get rid of it? The naïve approach is to take into account all the players, including those who transfer for free, into the analysis. This is what has been done below:



The figure rightly shows most of the true picture. First, the shape on the right side of the figure seems less focussed, more dispersed than earlier diagrams; this is to say that the model is less accurate at predicting actual transfer prices. Conversely, the line along the Y-axis now crosses the X-axis, i.e. extends below zero. The mean error is now zero, but the number of players (transferring for free) wrongly priced stayed the same, some with a negative transfer value, which makes no sense at all, some with a positive one. Hence the naïve correction of the selection bias resulted in a less accurate prediction of actual prices and in absurd negative prices. This is, of course, not satisfactory either.

Should we do any better? A famous econometrician and Nobel Prize winner, James Heckman, has designed a procedure to correct selection bias. Heckman's theory is frequently introduced to account for wage determination: "Suppose that a researcher wants to estimate the determinants of wage offers but has access to wage observations for *only those who work*" (from Wikipedia, emphasis added). The idea behind Heckman's theory is that agents have an

implicit reservation price; if the market price offered for their work is below their reservation price, they will abstain from contracting. In order to correct the selection bias, one has to estimate the probability that an agent transacts on the market, and that probability distribution function may be used to compute a correction factor<sup>2</sup>.

Carmichael et al. (1999) offered an implementation of the Heckman method. This path has since only been followed by Ruijg and Ophem (2014), which offered a minor variation on a small sample (only 55 effective observations). Carmichael et al. (1999) suggest they were not able to determine the sign of the correction factor since it “depends on whether unmeasured factors that raise the transfer fee raise or lower the probability of transfer. We were unable to sign the sample selection effect by *a priori* reasoning. On the one hand, players who are, for example, disruptive to dressing room spirit may be particularly prone to getting transferred, but if their behavioural problems are well-known in the game, this will lower fee offers for them. In this case, uncorrected regression models will, on average, understate the value of measured characteristics.

---

<sup>2</sup> To be more specific, the transfer fee equation should be  $F_i = X_i\beta + Y_i\gamma + Z_i\delta + e_i$ , where  $X_i$  is a vector of directly measurable human capital attributes and productivity indicators (including experience, age and goals scored);  $Y_i$  is a vector of other player characteristics (e.g. position);  $Z_i$  is a vector of selling club characteristics that indirectly reflect a player's ability (e.g. divisional status); and  $e_i$  is a random variable of mean zero, reflecting unobserved characteristics which affect the fee offered to the selling club.  $\beta$ ,  $\gamma$  and  $\delta$  are vectors of parameters to be estimated. Since transfer only occur if this offered price is above the owning club's reservation price, this probability can be modeled by a function  $I_i^* = V_i\alpha + u_i$  where  $V_i$  is a vector of characteristics thought to affect transfer movements,  $u_i$  is a standard normal variable, and  $\alpha$  is a vector of parameters to be estimated. The player is supposed to transfer if  $I_i^* > 0$ . The complete pricing equation then comes as:

$$F_i = X_i\beta + Y_i\gamma + Z_i\delta + \lambda_i\sigma_{FP}$$

where  $\lambda_i$  is the inverse Mills ratio evaluated at  $V_i\hat{\alpha}$  and  $\sigma_{FP} = COV(e_i, u_i)$ .

On the other hand, quality of play on the football field is not well captured by published statistics, and players transferring may be known as, for example, unusually sparkling or creative performers. Alternatively, their style may give them the capability to raise the play of the rest of the team as a unit, a species of externality. In this case, uncorrected regression models will tend to overstate the value of measured characteristics”.

This long citation illustrates the idea that the phenomenon at hand may be more complex than simple reservation price since it involves an element of uncertainty, *i. e.* whether the transferred player will fit with his new team. One consequence of this uncertainty is that it is impossible to determine *ex-ante* the productivity of adding a new player to a team. Players prices are thus not like options but rather like contingent contracts with uncertain outcomes. Consequently, specific asset management options are provided to the parties, as some players move during their contracts *on loan* either for free or for a fee. The Heckman approach of Carmichael et al. (1999) cannot take these subtleties into account, but we see no reason not to take these loans into account as well since they provide a way to ensure optimal distribution of player talent across the market. How shall we proceed next?

### **2.3 Valuing Football Transfers**

According to Carmichael (2006), there are two ways of valuing football transfers. The first theoretical model follows the economic theory of monopoly, where transfer fees are determined based on a bargain between clubs. In this model, the characteristics of the buying and the selling clubs are the most important variables. The second theoretical model follows the economic theory of perfect competition: the clubs demand playing skills, individual players are like bundles of those skills, and transfer fees are determined based on the skill mix of the player (Carmichael, 2006). Most empirical papers on the determination of transfer fees refer to this

latter theory as *hedonic pricing*. The existing literature generally points to a mix of the two theories in determining transfer fees for both theoretical and empirical reasons. From a theoretical point of view, teams need skills to score goals, win matches, championships and cups: thus they earn money directly (through prizes granted to winners) and indirectly (via ticket sales, fan material and sponsorship contracts, all being increasing functions of the sporting achievements): as a consequence, more efficient clubs may extract more value-added from the same set of skills, hence the value of a player depends both on his personal qualities and on the club's efficiency. From an empirical point of view, both series of factors have been found to be statistically significant in determining transfer fees: the characteristics of the football player, the timing of the transfer, the size and status of the selling and buying clubs have all been found to be significantly linked (Dobson & Gerrard, 1997; Szymanski & Smith, 1997; Carmichael & Thomas, 1993; Dobson & Gerrard, 2000). This is typically the case for the major football leagues in the world, especially the top five European leagues (EPL, La Liga, Bundesliga, Serie A, and Ligue 1). Eventually, both series of factors can be combined in a hedonic pricing function if club performance is properly measured.

The canonical example of hedonic pricing is provided by Dobson and Gerrard (1999), who developed a model of the player-transfer market in soccer, in which observed transfer fees are determined by player characteristics, selling-club characteristics, buying-club characteristics and time effects. Similar models were estimated by Dobson, Gerrard and Howe (2000) and Szymanski (2015), etc. The results showed that the factors that determined transfer fees varied markedly among market segments, which can be distinguished along several dimensions: geographical (regions or countries), sporting level (leagues), or through in-game positions, etc. Poli et al. (2017) used the most advanced data to date to approach the transfer fees evaluation by two steps using multiple linear regression in each step. The first step estimates the status of



buying clubs based on different variables characteristics while the second step evaluates footballers' characteristics: age, duration of the contract, level of buying club. They confirmed the segmentation of pricing function between positions but did not conclude on geographical segmentation since they only consider Big-5 transfers. It is worth noting that these authors rely on synthetic data about match facts provided by OPTA sport.

By contrast, most other studies rely on hand-picked match facts. The variables found to be statistically significant in determining football transfer prices include goal-scoring record, league appearances, international appearances, age (and sometimes age squared) (Carmichael, 1999; Dobson, Gerrard and Howe, 2000; Ruijg & van Ophem, 2015). Although transfer fees evaluation may differ from salaries evaluation by shape, they may intersect when the evaluation is based on individual characteristics. Mincer (1974) created a linear regression model to formulate the logarithmic function of the salaries in terms of individual (players) characteristics and enterprise (club) characteristics. Lucifora and Simmons (2003) did a regression model of the logarithmic salaries function in terms of four variables: experience determined by age and number of games played; performance determined by the number of goals and passes; reputation determined by the number of international games played with the national team; and determined by club's financial status and club's manager (coach) rating and performance. Age and reputation coefficients were proved significant. Age has been proved vital in many studies. Ruijg and Ophem (2015) proved that age (up to 26 years old) and the number of games played have a significant effect on the transfer fees of footballers. Dobson, Gerrard and Howe (2000) found player characteristics, time effects, selling-club characteristics and buying-club characteristics to determine transfer prices in English Professional Football. Frick (2007), like Speight and Thomas (1997) and Carmichael et al. (1999) stated that many common significant

variables are affecting both players' salaries and players' transfer fees like players' age, the number of games played with a club, international caps (games with the national team).

Overall, we found 28 papers published on this issue totalling 24,000 observations, that is, on average, 800 observations per paper: not even 3% of the 30,000 professional footballers in Europe alone. Hence the results need further confirmation. Moreover, since many studies value the transfer prices from a set of transferring players, there might be a selection bias, which is rarely tested for, and more rarely even corrected. Carmichael et al. (1999) offered a rare implementation of the Heckman method and conclude: "That some players are more likely to be transferred than others implies that estimation of hedonic price equations requires correction for selection bias." Only one study (Ruijg and Ophem, 2014) attempted to use the Heckman approach and found a notable selection effect, albeit on a very small sample (only 55 effective observations). It seems then absolutely crucial to design our research to tackle this selection problem, both at the data and methodologic levels.

#### **2.4 Machine Learning Techniques and Football Pricing**

Determining the compensation of a football player has become a complex task affected by numerous sets of factors (Yaldo & Shamir, 2017). For instance, in the pre-information era, where football player statistics and data were difficult to obtain, football player compensations were based on more qualitative analysis (Frick, 2006). This made the comparison of football player performance and skills very difficult. More recently, and particularly since the 1990s, where football player statistics and data were easily available, the compensations could be easily calculated and compared. Indeed, Frick (2006) emphasized that information about player statistics and data were published and the comprehensiveness of the data allowed comparison at different levels, including multiple performance figures and salaries. There are several other

pieces of information about football players that have been published and have been established to be useful in calculating player compensations: player innate characteristics, ability and human capital (Szymanski and Smith, 1997); extra income generated for the club through tickets, sale of merchandise and broadcasting agreements (Garcia-del Barrio and Pujol, 2007); players performance in the previous season, games played internationally, and the number of goals scored by the player (Frick, 2011); passing skills, free-kick accuracy, speed, and tackles (Yaldo & Shamir, 2017).

It has been established that superstar football players typically earn more compared to other players and the effect is magnified by the limited supply of such players. For instance, Garcia-del Barrio and Pujol (2007) showed that superstar football players attract monopsony rents because multiple employers compete for their services, forcing clubs to increase their wages in order to compete with the other clubs and hire these players. There have been some other arguments to suggest that even though the performance of a player improves with absolute income, salary inequality can have adverse effects on other players (Torgler, Schmidt & Frey, 2006) and affects coaching decisions (Garcia-del Barrio & Pujol, 2007). For the latter, it has been observed that players with higher salaries tend to be used by the coaching staff in a fashion that is not proportional to their performance on the field compared to other players who are compensated less generously.

The post-information era has, however, witnessed the emergence of more sophisticated and multi-dimensional methods in determining football player compensations. Lames, McGarry, Nebel & Roemer (2011) described these methods as requiring computational statistics and pattern recognition, or simply machine learning approaches. The method is based on the application of pattern recognition algorithms to performance (e.g., scoring), behaviour (e.g.,

aggression), and abilities (e.g., acceleration) data of football players (Yaldo & Shamir, 2017). The methods mainly use the intrinsic features of the individual players. Giving its elaborate form, these systems have been successfully used in determining player and ball movement on fields (Feess, Gerfin & Muehlheusser, 2010), automatic detection of player's position (Siegle, Stevens & Lames, 2013), interactive football training environment (Jensen et al., 2014), and game tracking (O'Donoghue and Robinson, 2009; Castellano, Alvarez-Pastor & Bradley, 2014). More recently, Matesanz et al. (2018) have used the machine learning approach to confirm European competitions, such as the UEFA Champions League or UEFA Europa League, as indeed a "money game" where the clubs with the highest transfer spending achieve better sportive performance. Nsolo, Lambrix and Carlsson (2018) also used the technique to distinguish top-tier players and show that prediction performance is higher for forwards than for other positions, in five European top football leagues.

## **2.5 Stock Market**

The reliance on the stock market as an important source of funding to support football development dates back to 1983 when Tottenham Hotspurs first listed on the stock market (Scholtens, 2009). Since then, many other football clubs have joined the stock market. Cooper and McHattie (1997) and Mitchell and Stewart (2007) attempted to provide some reasons for this seeming interest in the stock market: intense competition and the need to improve the financial position of a club. Scholtens (2009) emphasized that in efficient markets (like the stock market) market participants respond to new information or news that in some way or another might regard the firms they invest in. It is further emphasized that stock market participants can interpret the result as information and integrate it into the revaluation of their firm (Stadtman, 2006). A market reaction can be induced by the expected imminent cash flow associated with new information.

There is already a growing body of empirical literature on the association between sporting results and financial performance in the stock market, and the existing evidence so far is favourable. For instance, Renneboog and Vanbrabant (2000) studied 17 British teams listed on the stock market during 1995-1998. They find positive reactions for a team's win and a negative impact for a draw or loss. These responses were similar even for national or European games. Stadtmann (2003) analysed 97 games of Borussia Dortmund, a German football club, during 2000-2002 and found national as well as international games impact on the club's share returns. Similar results were found for the difference between the results on national or European games. Ashton et al. (2003) used the event study methodology to analyse the economic impact of national sports events on the stock market. Their focus was not only on the impact of a match outcome on clubs' stock market prices but also on the impact of England's national football team results on the FTSE100 index. The results showed that stock market return is indeed positive after wins and negative after losses.

Contrarian views appeared with Zuber et al. (2005), who considered the game-related performance of 10 listed football clubs in the English Premier League. They find that there is no abnormal return from neither expected nor unexpected results and conclude that stocks are owned by fans who are insensitive to the financial performance of clubs' stocks. More recent studies, using larger samples (Scholtens & Peenstra, 2009; Berkowitz & Depken, 2018) or meta-analysis (Geyer-Klingeberg et al., 2018) have tended to show that match results have an asymmetric impact on stock returns since losses lead to a negative effect while wins have almost no effect. From these recent studies, we can conclude that there is no authoritative rejection of the impact of match results on clubs' stock prices. Before we study the impact of transfers then, we must check whether Zuber's results can be extended beyond the English Premier League, *i. e.* whether match results do not impact clubs' stock prices (Dobson &

Goddard, 2001; Ashton et al., 2003; Palomino et al., 2009; Bell et al., 2012; Bernile & Lyandres, 2011).

Concerning players transfers, it should first be noted that efficient market should price them out, so only unexpected transfers or transfer prices should impact clubs' stocks valuation. Stadtmann (2006) found no effect of players sales or loans on the share price of a football club while Fotaki et al. (2009) were able to detect a positive effect for sales (outgoing) transfers and loans while players acquisitions (incoming) have a negative effect, but they do only take into account raw events, i. e. they do not correct for expectations, and this may be a significant flaw if the market is somewhat efficient.

From this review, it shall be concluded that no serious study of the impact of an unexpected part of transfer events on football clubs' market valuation has been done. This is what we intend to do. If the efficient market hypothesis holds, only the unexpected component of transfers should impact the club's stock prices. Let us take an example with the Pogba case; he was transferred for £95m to Manchester United in 2016 while his market value was £65m on *transfermarkt*: if the club paid £30m too much, then the market value of the club should be significantly negatively impacted by this overpriced. This is the standard economic point of view. If, on the contrary, the stock price positively reacts to an overpriced transfer, it may mean that the markets favour the sports results over the financial results, and that may be consistent with Zuber et al. (2005) and Andreff's findings that the open-league clubs are sporting results maximisers.

## CHAPTER THREE

### 3 PRICING FOOTBALL TRANSFERS: FINANCIAL ECONOMETRIC

#### APPROACH

##### 3.1 Introduction

Football has been consistently rated as the most popular sport globally. Its booming transfer market keeps impressing stakeholders, almost every year (table 3.1), with new record spending for the recruitment of best players. In such a market, one may wonder how the prices of the transfers' contracts or wages are estimated. In this chapter, we carried out the estimation using the hedonic pricing methodology. Hedonic pricing is a common pricing technique for many assets, from real estate to artworks. Pricing football players so far proved difficult to achieve satisfactorily. While previous studies relied on the same arguments for the pricing function, such as personal characteristics of the player (age, height), his competitive record and fame, and some characteristics of the contract between him and the club, all such studies so far have used small samples to price either transfer fees or salaries, possibly exhibiting selection bias. The present study collected a huge dataset of almost 100,000 wages and more than 87,000 transfer contracts (paid, free, and loans) covering 12 seasons (2007 - 2018) of the top five European leagues, North and South American leagues (Argentina, Brazil, Chile, Uruguay, Mexico, USA), and also China.

This chapter of the thesis sheds light on different factors of the pricing function of a football player; it is very difficult to account for all the determinants though. So, for the independent variables used in this part of the study, we have relied on 'real performance statistics' data which represent players' contributions during local and international games in the last two seasons before the date of the contract. The data set is one of the largest among recent and old studies in terms of number of players, countries, and the number of independent variables;

some were used for the first time like remaining duration of the contract. The reason behind using such diversified type of data is that the evaluation can be performed across geographical or positional segmentation. The reason behind this strategy was to contribute a solution to the selection bias problem, which has not been appropriately addressed in previous studies.

The results show that valuation models can be consistent across time or across space, the difficulty of tackling selection bias and heteroscedasticity in a global model using the *transfer fee* alone was solved through aggregating various elements of players' cost in one overall *package*. Such aggregation generated promising results and findings. Thus, such a global model can explain satisfactorily the expenditure by buying clubs from 2007 to 2018 on players' transfers and wages. Eventually, the scope of our database enables us to compute yearly price indices, which could help build a reliable valuation model useful for club management, insurance pricing or financial planning in the football universe.

Player	Age	Season	From	To	(MV)	Fee
Neymar	25	17/18	FC Barcelona	Paris SG	€100.00m	€222.00m
Kylian Mbappé	19	18/19	Monaco	Paris SG	€120.00m	€185.00m
Philippe Coutinho	25	17/18	Liverpool	FC Barcelona	€90.00m	€145.00m
Ousmane Dembélé	20	17/18	Bor. Dortmund	FC Barcelona	€33.00m	€145.00m
João Félix	19	19/20	Benfica	Atlético Madrid	€70.00m	€126.00m
Antoine Griezmann	28	19/20	Atlético Madrid	FC Barcelona	€130.00m	€120.00m
Cristiano Ronaldo	33	18/19	Real Madrid	Juventus	€100.00m	€117.00m
Paul Pogba	23	16/17	Juventus	Man Utd	€70.00m	€105.00m
Gareth Bale	24	13/14	Spurs	Real Madrid	€65.00m	€101.00m
Eden Hazard	28	19/20	Chelsea	Real Madrid	€150.00m	€100.00m
Cristiano Ronaldo	24	2009/10	Man Utd	Real Madrid	€60.00m	€94.00m
Gonzalo Higuaín	28	16/17	SSC Napoli	Juventus	€65.00m	€90.00m
Neymar	21	13/14	Santos FC	FC Barcelona	€50.00m	€88.20m
Harry Maguire	26	19/20	Leicester	Man Utd	€50.00m	€87.00m
Matthijs de Ligt	19	19/20	Ajax	Juventus	€75.00m	€85.50m
Romelu Lukaku	24	17/18	Everton	Man Utd	€50.00m	€84.70m
Virgil van Dijk	26	17/18	Southampton	Liverpool	€30.00m	€84.65m
Luis Suárez	27	14/15	Liverpool	FC Barcelona	€2.00m	€81.72m
Lucas Hernández	23	19/20	Atlético Madrid	Bayern Munich	€70.00m	€80.00m
Nicolas Pépé	24	19/20	LOSC Lille	Arsenal	€65.00m	€80.00m
Kepa	23	18/19	Athletic	Chelsea	€20.00m	€80.00m
Zinedine Zidane	29	2001/02	Juventus	Real Madrid	-	€77.50m
Kevin De Bruyne	24	15/16	VfL Wolfsburg	Man City	€45.00m	€76.00m
Frenkie de Jong	22	19/20	Ajax	FC Barcelona	€85.00m	€75.00m



James Rodríguez	23	14/15	Monaco	Real Madrid	€60.00m	€75.00m
-----------------	----	-------	--------	-------------	---------	---------

Table 3.1: Top transfers in the last two decades. Source: Transfermarkt.

The rest of the chapter is organized as follows. Sections 3.2 and 3.3 discuss some econometric issues and how they are dealt with in the chapter. Section 3.4 discusses the data, while Section 3.5 discusses the results. Section 3.6 presents some concluding remarks.

### 3.2 An Enduring Problem – the selection bias

Selection bias is likely to be a problem undermining the currently used transfer fee valuation techniques. In the standard estimation of the hedonic model, only players that have positive transfer prices are captured. This inherently leaves out players on free transfers, generating a “sampling bias”. Evidence from the estimations by Poli, Ravenel, and Besson (2017) shows that the typical hedonic model explains just about 80 percent of variations in prices. One may object that this apparent problem can be solved by including a ‘free transfer’ dummy: when their contract is over, players transfer for free. Although this is a bit adhoc since it does not provide for a consistent econometric procedure, this approach would not be enough to tackle player loans during their contract. Some players are lent by their club to another, some time for a fee and sometimes for free. The classical model cannot cope with these subtleties.

A famous econometrician and Nobel Prize winner, James Heckman, has designed a procedure to correct this bias. Heckman’s theory is frequently introduced to account for wage determination: “Suppose that a researcher wants to estimate the determinants of wage offers but has access to wage observations for *only those who work*”. The idea behind Heckman’s theory is that agents have an implicit reservation price; if the market price offered for their work is below their reservation price, they will abstain from contracting. To correct the

selection bias, one has to estimate the probability that an agent transacts on the market, and that probability distribution function may be used to compute a correction factor<sup>3</sup>.

Carmichael et al. (1999) offered an implementation of the Heckman method. This path has since only been followed by Ruijg and Ophem (2014), which offered a minor variation on a small sample (only 55 effective observations). Carmichael et al. (1999) suggest they were not able to determine the sign of the correction factor since it “depends on whether unmeasured factors that raise the transfer fee raise or lower the probability of transfer. We were unable to sign the sample selection effect by *a priori* reasoning. On the one hand, players who are, for example, disruptive to dressing room spirit may particularly be prone to get transferred, but if their behavioural problems are well-known in the game, this will lower fee offers for them. In this case, uncorrected regression models will, on average, understate the value of measured characteristics. On the other hand, the quality of play on the football field is not well captured by published statistics, and players transferring may be known as unusually sparkling or creative performers. Alternatively, their style may give them the capability to raise the play of the rest of the team as a unit, a species of externality. In this case, uncorrected regression models will tend to overstate the value of measured characteristics”.

---

<sup>3</sup> To be more specific, the transfer fee equation should be  $F_i = X_i\beta + Y_i\gamma + Z_i\delta + \epsilon_i$ , where  $X_i$  is a vector of directly measurable human capital attributes and productivity indicators (including experience, age and goals scored);  $Y_i$  is a vector of other player characteristics (e.g. position);  $Z_i$  is a vector of selling club characteristics that indirectly reflect a player's ability (e.g. divisional status); and  $\epsilon_i$  is a random variable of mean zero, reflecting unobserved characteristics which affect the fee offered to the selling club.  $\beta$ ,  $\gamma$  and  $\delta$  are vectors of parameters to be estimated. Since transfer only occur if this offered price is above the owning club's reservation price, this probability can be modeled by a function  $I_i^* = V_i\alpha + u_i$  where  $V_i$  is a vector of characteristics thought to affect transfer movements,  $u_i$  is a standard normal variable, and  $\alpha$  is a vector of parameters to be estimated. The player is supposed to transfer if  $I_i^* > 0$ . The complete pricing equation then comes as:

$$F_i = X_i\beta + Y_i\gamma + Z_i\delta + \lambda_i\sigma_{FFP}$$

where  $\lambda_i$  is the inverse Mills ratio evaluated at  $V_i\hat{\alpha}$  and  $\sigma_{FFP} = COV(\epsilon_i, u_i)$ .

This long citation illustrates the idea that the phenomenon at hand may be more complex than simple reservation price since it involves an element of uncertainty, *i.e.* whether the transferred player will fit with his new team. One consequence of this uncertainty is that it is impossible to determine *ex-ante* the productivity of adding a new player to a team. Player prices are thus not likely options but rather like contingent contracts with uncertain outcomes. Consequently, specific asset management options are provided to the parties, as some players move during their contracts *on loan* either for free or for a fee. The Heckman approach of Carmichael et al. (1999) cannot take these subtleties into account, but we see no reason not to take these loans into account as well since they provide a way to ensure optimal distribution of player talent across the market.

### **3.3 The hypothesis to Solve Selection Bias**

Heckman tried to price the shadow salaries of those not participating in the labour market. Our problem is a bit different. To carry on the labour market analogy, we are trying to price the penalties paid by workers leaving their job without notice, since this is what transfer prices are actually. Should we analyse the determinants of the values of those penalties without looking at the determinants of workers' overall compensation? This is exactly what has been done with transfer prices so far. We argue on the contrary that, since the buying clubs pay the penalties as part of a larger bill (featuring salaries and certain as well as contingent bonuses), this whole bill should be taken into account, and this is the only way to get rid of the selection bias. It has not been so far since it would have required data collection processes which were probably too expensive for the sake of the preliminary economic studies. But if we want to go beyond the selection bias, we must build up far more comprehensive databases to tackle the problem. This is what we do in the subsequent section. We recall different packages for pricing players from (Ezzeddine & Pradier, 2020), different from transfer fees, as follows:

- **The objective package (OP):** this is what is objectively defined by the transfer contract.

It features the transfer fee plus the salary received throughout the contract. For an  $n$ -year contract, the OP will look like:

$$\text{Objective Package} = TF + \sum_{i=1}^n \frac{w_i}{(1+r)^i}$$

In particular, if  $r = 0$ , which might be the case in the time of QE, then

$$\text{Objective Package} \approx TF + nw$$

- **The Subjective Complete Package (SCP):** this package extends the computation beyond the time horizon of contract, including future player salaries. It takes into account the contingent payments (such as bonuses for scoring goals or playing additional tournament matches). This package is *complete* since it features all elements of costs; it is *subjective* as well since there is no objective assessment of it except the market price:

$$\text{Total cost} = TF + PV \text{ of expected future income in } t = TF + \sum_{i=1}^{+\infty} \frac{E(\tilde{w}_{t+i}|w_t)}{(1+r)^i}$$

It seems difficult to measure this package, which features undisclosed elements (such as the contingent payment scheme) and uncertainty on their realization (since they are contingent). As the interest rate during the period is low, we neglected  $r$ , and we approximate  $E(\tilde{w}_{t+i}|w_t)$  by a player current relative salary at age  $t$  (*i. e.* his salary divided by the average salary at the same time) multiplied by the average salary at  $t+i$ .

Therefore,

$$SCP \approx TF + \sum_{i=\text{age}}^{42} \frac{w_{\text{age}}}{E(w_{\text{age}})} E(w_{i+1}) \times \frac{\text{number of players of age } i+1}{\text{number of players of age } i}$$

Where  $E(w_{age})$  and *number of paid players of age i* are taken from the whole database. Eventually, SCP can be written as:

$$SCP \approx TF + w \times multiplier (age)$$

Where multiplier has to be estimated from the wage distribution in the sample. It should be mentioned here that the income multiplier was commonly used by UEFA in the 1990s to determine the transfer prices of players between European football clubs. This price had to be at least equal to the gross salary of the player multiplied by a coefficient depending on the age of the player (UEFA, art. 3, 1992). Accountants such as Morrow (1999), and Scarpello and Theeke (1989) criticized the inconsistency of the method with standard economic theory. The main difference between our approach and the UEFA-1990's own is that our multiplier is estimated from the data.

- **The Subjective Salary Package (SSP):** here we only include the future payments, i. e. salaries and contingent bonuses, hence:

$$PV \text{ of expected future income in } t = \sum_{i=1}^{+\infty} \frac{E(\tilde{w}_{t+i}|w_t)}{(1+r)^i}$$

Which can be approximated as

$$SSP \approx \sum_{i=age}^{42} \frac{w_{age}}{E(w_{age})} E(w_{i+1}) \times \frac{\text{number of players of age } i+1}{\text{number of players of age } i}$$

Or

$$SSP \approx w \times multiplier (age)$$

This approach is then close to the UEFA 1990's practice, with the only exception that the multiplier is fitted to the actual data. Of course, regression will be performed on

logarithms of those quantities since the hedonic pricing function is of the exponential form, *i.e.*  $package = \prod X_i^{\beta_i}$  hence

$$\ln(package) = \sum \beta_i \cdot \ln(X_i)$$

### 3.4 Data

One important characteristic of our study is the volume. Approximately 87,000 transfers have been harvested (web-scraped) to form a cross-sectional dataset containing hundreds of characteristics. These are basically ‘transfers’ from the top five leagues in the world (English Premier League, Spanish La Liga, German Bundesliga, Italian Serie A, and French Ligue 1), South and North American soccer leagues (Argentina, Brazil, Chile, Uruguay, Mexico, and the USA), and also Chinese Superleague. The data covered seasons between 2007/2008 and 2018/2019. The enormous size of the database in hand does not deny the fact that there still exist a few problems like incomplete dates not recorded at the source.

Contract Year	Mean	Std. Dev.	Freq.	Mean current salary	Std. dev. of current salary	Mean transfer fee	Std. dev. of the transfer fee
2004	297	136.4203	5				
2005	241.56	202.4475	25				
2006	241.2469	165.3259	324				
2007	175.0902	151.11	4334			633210.28	2309900.7
2008	176.2673	153.6337	7191			431481.64	1860224.8
2009	177.7497	155.1205	7392			410665.59	2404531.9
2010	175.3912	156.839	7819			278246.98	1500263.8
2011	177.9628	156.6341	8187			384035.12	2023285
2012	180.6538	153.1095	6813			126981.22	816017.22
2013	173.357	154.9533	8213	5.272	96.800	390248.36	2783166.8
2014	181.8622	168.0684	7705	41.154	896.846	344193.66	2260420.1
2015	176.0767	156.0042	7880	430.787	976.860	507264.85	2732396.7
2016	175.7828	152.53	8857	307.259	765.166	567657.03	3046370.4
2017	164.4996	150.9517	7884	368.764	850.294	804484.29	4262611.1
2018	137.8263	145.1862	4572	32.264	1073.760	1087477.5	5244653
<b>Total</b>			<b>87201</b>				

Table 3.2: Total transfers per year

Moreover, more than 100,000 salaries have been extracted for the period between 2013 and 2018 from *sofifa.com*. The salaries extracted served in creating different dependent variables in the form of packages, new forms of pricing function for football players (as discussed in the previous section). Out of the total extracted salaries, around 18,000 have been successfully merged with their relevant transfer fees. After successfully merging the fees and the salaries, we then computed the three types of packages (that were used as dependent variables): Objective Package (OP), Subjective Complete Package (SCP) and Subjective Salary Package (SSP). Table 3.2 represents a breakdown of average and total transfers per year. In contrast, the average transfer fees paid per country and summary of players' movement across buying/selling countries are shown in Tables 3.3 and 3.4, respectively.

Purchasing contracts per country				Selling contracts per country			
Country of current/Purchasing Club	Mean	Std. Deviation	number of contracts	Country of Previous/Selling Club	Mean	Std. Deviation	number of contracts
Argentina	0.9277	3.3907	8021	Argentina	1.4410	4.2404	8468
Bolivia	2.8547	5.5410	335	Belgium	4.7423	6.8687	356
Brazil	0.6799	2.9614	13363	Brazil	0.9448	3.5329	13411
Chile	0.2914	1.9111	4367	Chile	0.6569	2.9221	4207
China	2.6078	5.4745	2984	China	1.5219	4.2837	2549
Columbia	0.3018	2.0154	360	Columbia	1.8861	4.8168	426
Ecuador	0.4454	2.3309	280	Ecuador	1.5106	4.3674	297
England	3.5569	6.4188	7350	England	2.7557	5.7832	7184
France	3.3235	6.0800	4447	France	3.3190	6.1101	4584
Germany	5.7184	6.8392	3608	Germany	4.8366	6.5708	3605
Greece	2.9789	5.5818	411	Greece	3.4415	5.9312	314
Italy	2.4396	5.3907	13461	Italy	2.0274	4.9903	13243
Mexico	0.7717	3.2283	6925	Mexico	0.5004	2.6473	6607
Netherlands	3.0081	5.7551	262	Netherlands	6.5775	7.4757	359
Paraguay	0.3129	1.9811	326	Paraguay	1.5416	4.3709	341
Portugal	2.9132	5.8068	764	Portugal	4.5606	6.8010	831
Russia	6.4466	7.4861	212	Russia	6.5466	7.2299	230
Spain	3.1820	6.0504	5146	Spain	2.7959	5.7892	5195
Switzerland	2.1448	4.9224	260	Switzerland	4.5134	6.6621	296
Turkey	5.6393	6.8917	425	Turkey	3.8384	6.3714	290
USA	0.3167	2.0592	5692	USA	0.1909	1.6103	5011
Uruguay	0.0826	1.0423	4929	Uruguay	0.6631	2.9634	4980

Table 3.3: Frequency of Selling and Buying Contracts per Country respectively

Country	Total Sales contracts	Exports	local sales contracts	Exports (% of sales)	Total Purchases	Imports	local Purchases contracts	Imports (% of Purchases)
England	7184	1873	5311	26%	7350	2093	5257	28.47%
France	4584	1557	3027	34%	4447	1420	3027	32%
Germany	3605	1283	2322	35.50%	3608	1286	2322	35.64%
Italy	13243	2279	10964	17%	13461	2497	10964	18.55%
Spain	5195	2145	3050	41%	5146	2096	3050	40.73%
Argentina	8468	3207	5261	37.87%	8021	2760	5261	34.41%
Brazil	13411	2255	11156	16.80%	13363	2207	11156	16.51%
Mexico	6607	1255	5352	19%	6925	1573	5352	22.71%
Chile	4207	1165	3042	27.70%	4367	1325	3042	30.34%
Uruguay	4980	1817	3163	36.48%	4929	1766	3163	35.83%
China	2549	619	1930	24.28%	2984	1054	1930	35.32%
USA	5011	965	4046	19.26%	5692	1646	4046	28.92%

Table 3.4: Percentage of Import/Export contracts out of Selling and Buying Contracts per Country respectively

The transfers were further disaggregated into three different categories, as follows:

- Paid transfers (*PaidTransfer*): transfers that include payment of compensation fees from the buying club to the selling club. It is a mandatory fee paid when players move from one club to another before ending their contracts with their previous/selling clubs. The data set includes 11,300 contracts out of the total number of transfers.
- Free Transfers (*FreeTransfer*): represented by a dummy variable which differentiates free transfers (i.e. players moving freely after successfully finishing their contract's period) from other transfers or loans. The volume of free transfers in the data set is around 30,000. Players who have successfully ended their loan contracts and are back to their original/loaning clubs have also been treated as free transfers since the return to their initial clubs incurs no obligations or charges.
- Loans (*TT\_Loan*): These are temporary contracts that allow players to play for another club for a limited period (most often one year) before returning to their original clubs. Loans usually incur a little or no fee, unlike the cases of permanent paid transfer. The data set contains 20,000 loan contracts. Besides, it's necessary to state that the



remaining contracts lies under the category of *end of loans*. Such category is for the players returning from their loans back to their clubs.

Based on the structure of the model, a basic dataset was created consisting of personal players' information at the time of contract like players' age, contract characteristics (duration of the contract and remaining duration of the previous contract), player's height, preferred feet in use, position, country of origin (and current nationality; some players carry dual citizenship), and country of buying/selling club. Additionally, the dataset was supported with the relevant player's performance history, players' statistics per season, for the last two consecutive seasons preceding the date of transfer. *Previous* refers to the data from last season before transfer while *Previous2* to the data one before the previous season preceding the date of transfer. These statistics included the number of goals, assists, different competition appearances (including Champions League), international appearance, number of yellow/red cards. Most studies considered only this kind of data. We extracted every information available on the web. We added google statistics (as a proxy for players' marketability and popularity) such as google scores and trends at the moment of the contract to capture player's fame indicator. While the database contains no less than 76 independent variables that describe the performance of football players, their characteristics in multiple competitions have been grouped and tested based on each set of characteristics. Regression analysis for each subgroup was carried out on a vector of eight variables (ZGOALS1, ZGOALS2, ZASSISTS1, ZASSISTS2, ZEXP1, ZEXP2, ZBAD1, ZBAD2) as follows:

Where  $\ln fee$  is the log of transfer fees,  $X$ 's are the sub-groupings,  $\varepsilon$  is the error term, and  $\alpha$  is the associated coefficients of  $X$ . All variables with  $\alpha_i$  significantly different from zero are then aggregated into a group index computed as:

$$GroupIndex_{j,t} = \sum_{i=1}^n \exp(\hat{\alpha}_i) \cdot x_{i,t}$$

These variables are then centred, normalized and logged. Eventually, the log of transfer fees is regressed on all *GroupIndex* variables (ZGOALS1, ZGOALS2, ZASSISTS1, ZASSISTS2, ZEXP1, ZEXP2, ZBAD1, ZBAD2).

Moreover, other factors that media have always been pointing on like player agents have been quantified by generating a dummy variable (equals to 1 in case of presence of an agent and 0 otherwise). Other dummy variables were created to control for specific well-known agents (“Mino Raiola”, “Gestifute”). The previous salaries were attached and used as variables that may affect future contracts. Some buying and selling clubs’ characteristics were quantified into independent variables (generated using the available data set) such as: “*co\_cuclu*” (total number of transfers made by buying/current club per period of study), “*co\_preclu*” (total number of transfers made by selling/previous club per period of study), “*ltot\_cuclu*” (total amount spent by buying/current club during the period of study) “*ltot\_preclu*” (total amount spent by selling/buying club during the period of study). The data has been obtained from many sources like: “[www.google.com](http://www.google.com)”, “[www.bigsoccer.com](http://www.bigsoccer.com)”, “[www.tifosobilanciato.it](http://www.tifosobilanciato.it)”, “[www.footyrate.com](http://www.footyrate.com)”, “[www.sportune.fr](http://www.sportune.fr)”, “[www.calciofinanza.com](http://www.calciofinanza.com)”, “[www.totalsportek.com](http://www.totalsportek.com)”, “[www.sportlens.com](http://www.sportlens.com)”, “@swissramble”, [www.deloitte.com](http://www.deloitte.com) (Deloitte Football Money League), Sky Sports, and Sports Intelligence Reports.

### 3.5 Estimation Strategy and Findings

To price football players, we run a micro-model using performance data (players' statistics from previous years). In each model, we test four dependent variables: the log of transfer fees (lfee), the objective package (OP), the subjective complete package (SCP), and the subjective salary package (SSP).

### 3.5.1 Micro-Level Results

To start with, using multiple linear regression, we estimate the different dependent variables (lfee, OP, SCP and SSP) using different types of independent variables from the data set. This approach for calculating players' transfer fees or packages is based solely on the players' performance characteristics in the previous seasons as per the following function:

$$\ln Fee = \sum_1^i a_i \ln X_i + \sum_1^j a_j \ln Y_j + \sum_1^k a_k \ln Z_k + u_i \dots\dots\dots (1)$$

Where X represents the control variables vector (including country, position, and yearly dummies), Y is the personal characteristics vector, and Z represents the players' performance vector. In England, for instance, the Z-vector performs very well where most of the values are significant. The players' performance vector includes goals, assists, experience (aggregate appearances in local and international leagues), bad (negative characteristics, i.e. aggressiveness indicator based on the number of cards).

Firstly, it is noteworthy to mention that we have run thousands of tests on the abundant data in hand hence what we are presenting in this paper is just a summarised subset of the detailed results that are available upon request. In deciding on which estimations to present, we start our testing from the general overall sample before branching into smaller subsets. According to the results, we had to choose between minimising the problems of heteroscedasticity or a

higher coefficient of determination  $R^2$  as results showed a trade-off between the two. Lower heteroscedasticity comes at the cost of lower  $R^2$ . Promising results came from the use of the subjective complete package (SCP), besides the salary package (SSP), but the complete package offers more information about both transfer fees and salaries. First, we tried to break down the overall sample of data in an organized, gradual way into segments. We consider the geographical segmentation across continents then countries. After geographical segmentation, which could not help overcoming the heteroscedasticity issues, we try to segment the market across players' positions. Slight improvements were achieved in each segmentation, but the best came after segmenting the regions based on positions. At such level, European and Latin American segmentations across the three positions generated a heteroscedasticity free model adapting the relevant independent variables accordingly.

The beginning was from the generalised model represented in Table AP.1.1 (in Appendix 1), which represents evidence to support the assumption of packages as a better financial valuation for football players. The results show lower heteroscedasticity and  $R^2$  than previous test results when using the packages than transfer fees. Restricting the results to only transfer fees (Table 3.5) also generates some selection issues, as discussed in Sections 3.2 and 3.3. This finding gives a better price for Europe at the aggregate level, but disappointing ones for other markets despite the number of significant factors. Therefore, taking into account the preliminary findings in the first two general models which favoured the estimation using packages rather (less heteroscedasticity), we decided to test if market segments apply in the model. We restricted testing to specific samples across regions/continents. Although heteroscedasticity didn't disappear, it has decreased when using two packages: SCP and SSP as the dependent variable (Table 3.6). Moving deeper into the country level doesn't produce full, satisfying results as the problem still persist (Table 3.7). In Europe, the top five leagues (England, Spain,

Germany, Italy, and France), the model does not work much better, despite the apparent improvement. Using the chow test, we prove further that France and Spain adapt to the model as a unique market using SCP as the dependent variable. Some indicators can prevail, showing that English market pays a premium for players scoring more goals while Germany's market is affected by players with higher playing experience. Remarkably, the loan dummy for England is extremely lower than other top five countries, and such finding can be justified by the interest of other clubs to put their players on display in the Premier League in order to expose the player for better purchase alternatives in future.

Likewise, we performed regression across positions (Table 3.8). The analysis was carried on using samples of forward players, midfielders, and defenders independently. Like previous segmentation, the results favoured the estimation using packages but could not solve the heteroscedasticity problem. The tests carried on forward players revealed the factors that affect the different pricing functions (transfer fee or the packages). The number of goals, assists, experience quantified by the number of games played, represented by ZGOALS, ZASSISTS, and ZEXP respectively, have positively significant effects on the pricing functions of forward players (strikers), while in midfielders sample, non-surprisingly, assists indicator characterised by ZASSISTS is positively significant in addition to ZGOALS and ZBAD that are also significant. This can be justified by the presence of offensive and defensive midfielders in the sample. At the defenders' level, the variables referring to experience, aggressiveness, and even goal-scoring habits are positively significant. The height of defenders also has a strong positive significant effect. We find it difficult to explain the positive effects of the negative characteristics variable ("ZBAD") of the players influencing the logarithmic pricing function, except by a kind of selection bias: the database contains all players mentioned on the internet site, hence all "significant" players, even if they made some mistakes. It is thus likely that the

more mistakes they have done, the more value they must have to be still counted as “significant”. This is a remarkable result which can be further investigated and tested. However, history provides little evidence characterised by relatively aggressive players like Eric Cantona, Sergio Ramos (approaching a world record), Roy Keane, Paul Gascoigne, and Nemanja Vidic (one of the best defenders at the time). Maybe these players are identified as being more likely to be respected by others and / or prevent chances for the opponents (even if this requires a foul and a card)

To move forward to find the best market segments that could adapt the model without econometric challenges, we retrieved the continental segments. We divided it into smaller segments based on players positions hence double segmentation. Accordingly, regressing across positions using the European (Table 3.9) and Latin American (Table 3.10) samples showed that the newly designed packages work even better than previous segments. The European sample which contains transfers from the top-five leagues responds significantly and adapts collectively better as a single market to this model. Finally, we obtained a heteroscedasticity-free model with  $R^2$  equals 0.6, 0.5, and 0.6 across forward players, midfielders, and defenders’ samples, respectively, with significant independent variables. It confirms the findings concerning the significant independent variables affecting the pricing of forward players, midfielders, and defenders. Similarly, Latin American market adapts collectively better as a single market to this model with  $R^2$  equals 0.4, 0.52, and 0.36 across Forward players, midfielders, and defenders’ samples respectively. Noteworthy to mention that across the countries model, we could form a smaller sub-segments formed of two or three countries in among the European sample, while we needed to remove Brazil from the Latin American sample (Argentina, Brazil, Chile, Uruguay, and Mexico) in order to form a unique market adapting to the model. Only the last segmentation, across continents and positions together, was able to generate unique markets through our model.

Overall, it is revealed that the google variables were significant in most of the tests performed across forward positions. This could be justified in light of the media coverage that lucky strikers receive in general. Thus, the popularity and marketability could be an important factor affecting the price of some players, especially superstars like Neymar, Mbappe, Cristiano Ronaldo and many more. The performance alongside popularity may add value to the player and support the stardom controversy that both fame and performance create superstars.

Moreover, it is important to mention that this study made use of new independent variables (remaining duration of the contract, previous salary) that were used in Ezzeddine and Pradier (2020) in addition to new dependent variables (SCP) using a different type of data set composed of real data extracted from player and clubs' characteristics and match statistics. Heteroscedasticity disappeared entirely in the regression models across all positions in both Europe and Latin America except across forwards market in Latin America where it was still apparent but very low. Finally, the gradual segmentation helped in tackling heteroscedasticity by moving into segmented markets across continents first then countries and, lastly, positions using different dependent variables until we arrived at a significant model per region.

The main objective of this study was to find if there is one unique universal model or different models for estimating players' financial values. For that reason, there is a need to investigate further, using a bigger sample of European countries like Portugal, Netherlands, and Belgium that are feeding top clubs with good talents, if there are more segments to be considered.

2007 – 2018				
	World	Europe	Americas	USA+China
D. Variables	lfee	lfee	lfee	lfee
I. Variables				
freetransfer	-4.625***	-12.08***	-1.143***	-1.716***
tt_loan	-2.943***	-9.707***	-0.560***	-0.607***
ZGOALS1	0.758***	0.205***	1.173***	1.554***

ZGOALS2	0.337***	0.343***	0.399***	0.686***
ZASSISTS1	0.963***	0.391***	1.088***	1.404***
ZASSISTS2	0.398***	0.265***	0.277**	-0.0065
ZEXP1	1.219***	0.405***	0.552***	0.952***
ZEXP2	0.597***	0.202***	0.344***	0.0659
ZBAD1	0.485***	0.371***	0.521	-1.700***
ZBAD2	0.251***	0.242***	-0.351**	0.487*
minoraiola	-0.575	0.136	-0.24	-7.419**
mediabase	-0.136	-0.477		
gestifute	0.253	0.569**	-2.068***	0.615
lage	36.99***	32.23***	11.63***	48.56***
lage2	-6.411***	-5.162***	-2.169***	-7.866***
ldur_hyp	0.732***	0.665***	0.479***	0.397***
lheight	1.167***	4.302***	0.187	5.063***
lgoo	-0.00764	0.00141	-0.0293***	-0.00267
lgt	0.0977***	0.0406***	0.0808***	-0.0253
lco_cuclu	0.256***	0.218***	0.139***	0.119***
lco_preclu	0.0731***	-0.139***	0.0385***	0.0830***
ltot_preclu	0.0653***	0.0450***	0.0246***	0.0427***
ltot_cuclu	0.107***	0.0880***	0.0513***	0.0681***
footer_both	-0.355***	0.368***	-0.126	-0.575***
footer_left	0.146**	-0.318***	-0.0514	0.376***
footer_right	0.214***	-0.274***	0.0202	0.352***
nat_cont_africa	0.140***	-0.0132	0.261	-0.135
nat_cont_asia	0.174**	0.0697	-0.259	1.251***
nat_cont_australia	-0.553***	-0.395*	-1.278*	0.303
nat_cont_europe	0.217***	-0.189***	-0.0722*	-0.311***
nat_cont_north_america	-1.181***	-0.597***	0.237**	-0.796***
nat_cont_south_america	-0.473***	0.0774	0.522***	0.711***
pos_forward	-0.122**	0.0128	-0.172***	-0.117
pos_def	-0.202***	0.057	-0.232***	-0.197
pos_mid	-0.229***	-0.0161	-0.208***	-0.0369
y2007x	0.801***	0.493***	0.168	-0.0543
y2008x	0.573***	0.323*	0.193	-0.0769
y2009x	0.325**	0.310*	0.159	-0.117
y2010x	0.0453	0.266	0.142	-0.447
y2011x	-0.0351	0.216	-0.117	-0.164
y2012x	-0.575***	-0.0897	-0.134	0.0537
y2013x	-0.21	-0.05	-0.221	0.00394
y2014x	-0.164	-0.147	0.087	0.333
y2015x	-0.106	-0.0198	-0.0815	0.849
y2016x	-0.00703	-0.0236	0.156	1.213
y2017x	-0.0881	-0.301*	0.404**	0.554
y2018x	-0.340**	-3.420***	0.248	-0.0157
Constant	-59.15***	-49.64***	-19.23***	-84.25***
<b>Chi2</b>	<b>31501.02</b>	<b>2984.64</b>	<b>46659.77</b>	<b>7003.79</b>
<b>Prob&gt;Chi2</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Observations</b>	<b>84,304</b>	<b>32,762</b>	<b>36,287</b>	<b>8,397</b>
<b>R-squared</b>	<b>0.428</b>	<b>0.785</b>	<b>0.177</b>	<b>0.274</b>
<b>Adj. R-Squared</b>	<b>0.4275</b>	<b>0.7844</b>	<b>0.1763</b>	<b>0.2699</b>

Table 3.5: Basic Tests using “lfee” (transfer fee). Samples: Universal/Continental



Segmentation – Continents/Regions																
D. Variable	World				Europe				Latin America				China & USA			
	lfee	OP	SCP	SSP	lfee	OP	SCP	SSP	lfee	OP	SCP	SSP	lfee	OP	SCP	SSP
<b>I. Variables</b>																
freetransfer	-8.429***	-0.896***	-0.521***	-0.196***	-12.46***	-1.186***	-0.724***	-0.320***	-3.580***	-0.707***	-0.524***	-0.354***	-4.740***	-0.221	0.142	0.437***
tt_loan	-6.562***	-0.708***	-0.414***	-0.110**	-9.933***	-0.906***	-0.550***	-0.170***	-3.218***	-0.708***	-0.573***	-0.400***	-3.355***	-0.647***	-0.317**	0.0615
ZGOALS1	0.391***	0.274***	0.265***	0.230***	0.193	0.192***	0.184***	0.181***	2.009***	0.790***	0.776***	0.688***	0.454	0.222	0.219	-0.155
ZGOALS2	0.335**	0.119**	0.142***	0.112**	0.356**	0.0541	0.0798	0.0637	0.401	0.285	0.227	0.266	1.351*	0.485**	0.591***	0.407**
ZASSISTS1	0.702***	0.186***	0.187***	0.197***	0.393***	0.142***	0.140***	0.145***	0.913**	0.384*	0.453**	0.372*	1.373*	0.656***	0.616***	0.423**
ZASSISTS2	0.156	0.112**	0.102**	0.103**	-0.00758	0.105**	0.108**	0.0997*	0.157	-0.00116	0.0988	0.108	0.497	-0.432*	-0.423**	-0.22
ZEXP1	0.786***	0.144***	0.127***	0.0867**	0.562***	0.167***	0.165***	0.137***	0.0146	-0.0744	-0.179	-0.159	1.783***	0.172	0.155	0.164
ZEXP2	0.276**	0.277***	0.292***	0.278***	0.274**	0.242***	0.263***	0.248***	0.412*	0.429***	0.415***	0.397***	-1.470***	0.213	0.276*	0.396***
ZBAD1	0.853***	0.132	0.163**	0.164**	0.457*	0.107	0.137	0.139	1.927*	-0.705*	-0.722*	-0.730*	-1.185	-0.387	-0.114	-0.121
ZBAD2	0.391**	0.103*	0.117**	0.0747	0.246	0.0990*	0.118**	0.109*	-0.57	-0.00563	-0.075	-0.319	0.999	0.667***	0.532***	0.0936
<b>lRemDurDays</b>	<b>0.227***</b>	<b>0.0774***</b>	<b>0.109***</b>	<b>0.0866***</b>	<b>0.0341</b>	<b>0.0344***</b>	<b>0.0771***</b>	<b>0.0631***</b>	<b>0.393***</b>	<b>0.264***</b>	<b>0.243***</b>	<b>0.214***</b>	<b>0.835***</b>	<b>0.168***</b>	<b>0.134***</b>	<b>0.0654*</b>
minoraiola	-0.383	0.025	0.0465	0.0991	-0.548	0.00592	0.0322	0.0856	-4.775							
mediabase	-0.723	0.441	0.449	0.497	-0.566	0.429	0.424	0.463								
gestifute	-0.121	0.148	-0.0338	-0.181	-0.0123	0.189	-0.00356	-0.209	-1.169	1.611	1.946	2.336*				
lagem	0.22	-0.216**	-0.00088	0.0403	-0.0146	-0.222**	-0.00109	0.0546	0.348	0.123	0.163	0.215	1.26	-0.409	-0.121	-0.144
lagep	1.042***	0.665***	0.197***	0.115*	0.921***	0.733***	0.283***	0.221***	0.792***	0.653***	0.143	0.0526	1.870**	0.840***	0.286	0.0753
lagem2	-0.0594	-0.08	0.551***	0.542***	0.0192	-0.0968*	0.527***	0.513***	-0.111	-0.291	0.513***	0.489**	-0.663	0.214	0.771***	0.796***
lagep2	-1.019***	-0.310***	-0.346***	-0.300***	-0.714***	-0.323***	-0.366***	-0.336***	-0.847***	-0.299***	-0.327***	-0.270***	-1.083***	-0.368***	-0.398***	-0.300***
ldur	1.317***	1.012***	0.131***	0.102***	0.849***	0.984***	0.109***	0.0922***	1.415***	0.872***	-0.0257	-0.0678	0.797***	0.886***	0.105	0.163**
lheight	6.665***	0.769	0.707	0.297	4.705**	-0.134	-0.25	-0.587	5.350*	2.436	2.963*	2.301	10.49	-0.547	-0.456	-1.472
lgoo	-0.00455	0.0259***	0.0292***	0.0210***	0.0307	0.0313***	0.0347***	0.0247***	-0.0582**	0.0165	0.0188	0.0147	0.0036	0.0173	0.0196	0.0127
lgt	0.0283	0.0108	0.0118	0.0121	0.0115	0.0118	0.00575	0.000995	0.0805*	0.036	0.0354	0.0404	0.0662	0.0889**	0.0885**	0.0842**
lco_cuclu	-0.00108	0.0845***	0.0813***	0.0892***	0.409***	0.145***	0.132***	0.131***	-0.152	-0.00045	0.0032	-0.014	-0.889***	-0.0695	-0.0551	0.0168
lco_preclu	-0.0443	0.0639***	0.0744***	0.0726***	-0.0977**	0.0700***	0.0878***	0.0783***	0.0569	0.028	0.0112	0.0206	-0.0965	0.203***	0.211***	0.176***
ltot_preclu	0.0564***	0.0472***	0.0402***	0.0395***	0.0355***	0.0555***	0.0474***	0.0496***	0.0216	0.0152	0.0146	0.0109	0.0618*	0.00591	0.00131	0.00562
ltot_cuclu	0.162***	0.0538***	0.0460***	0.0422***	0.0910***	0.0600***	0.0513***	0.0511***	0.0964***	0.0204**	0.0175**	0.0160*	0.241***	0.0178	0.0129	0.0051

footer_both	-0.0588	-0.0923	-0.0297	-0.0616	0.271	0.00817	0.0389	0.029	-0.83	0.423	0.559	0.566	0.99	-0.0259	0.0934	0.0876
footer_left	-0.0602	-0.0355	-0.058	-0.0265	-0.143	-0.0317	-0.0441	-0.0441	0.344	-0.940***	-0.941***	-0.892***	0.179	0.303*	0.238	0.258*
footer_right	-0.0208	-0.0201	-0.0522	-0.029	-0.114	-0.0336	-0.0479	-0.0505	0.332	-0.819***	-0.858***	-0.862***	-0.129	-0.0952	-0.0809	0.018
nat_cont_africa	-0.0742	0.253***	0.234***	0.222***	0.0732	0.245***	0.229***	0.207***	-1.788*	-0.594	-0.738	-0.688	0.337	0.306**	0.318**	0.249*
nat_cont_asia	0.179	0.226**	0.229**	0.223**	0.23	0.217*	0.224**	0.205*	-0.747	-1.119**	-1.151**	-1.085**	0.851	0.838***	0.876***	0.669***
nat_cont_australia	-1.399***	-0.276*	-0.259*	-0.255*	-0.0669	-0.502**	-0.439**	-0.512**					0.0213	-0.219	-0.168	-0.0849
nat_cont_europe	-0.214*	0.049	0.0850**	0.0787**	-0.235*	-0.103**	-0.0565	-0.0568	-0.320*	0.307***	0.320***	0.329***	-0.626	0.038	0.114	0.146
nat_cont_north_america	-1.315***	-0.233***	-0.183***	-0.147**	-0.434*	-0.019	0.0276	0.0218	-1.432**	0.408	0.297	0.169	-0.399	-0.536***	-0.397***	-0.358***
nat_cont_south_america	0.0315	0.230***	0.229***	0.201***	0.519***	0.327***	0.311***	0.268***	-1.524**	0.33	0.298	0.253	2.294***	0.374**	0.235	-0.0725
pos_forward	0.213	0.170***	0.144**	0.142**	0.118	0.223***	0.200***	0.196***	0.204	0.307*	0.295*	0.256	-0.0534	0.0998	0.0018	0.129
pos_def	0.0543	0.0734	0.0735	0.0685	0.09	0.0296	0.0278	0.0254	-0.177	0.393***	0.432***	0.419***	-0.0948	0.153	0.0813	0.0806
pos_mid	0.0608	0.043	0.033	0.019	0.112	0.0917	0.0905	0.0871	-0.182	0.144	0.123	0.0875	-0.121	-0.14	-0.186	-0.1
y2014x	0.176	-0.140**	-0.164**	-0.175***	0.989***	-0.0863	-0.0687	-0.0891	-1.103***	-0.232	-0.243	-0.215	1.224	0.299	0.205	0.155
y2015x	0.0792	-0.124*	-0.171***	-0.172***	1.077***	-0.029	-0.0701	-0.086	-1.276***	-0.277	-0.262	-0.212	1.052	0.496**	0.462**	0.382**
y2016x	0.0703	-0.753***	-0.809***	-0.880***	0.992***	-0.750***	-0.793***	-0.899***	-0.864**	-0.600***	-0.645***	-0.621***	1.475	0.104	0.0692	-0.062
y2017x	0.00393	0.00645	-0.0159	-0.0247	0.858***	0.258***	0.268***	0.240***	-0.381	-0.409*	-0.419**	-0.402*	0.454	-0.0012	-0.153	-0.222
y2018x	-0.018	-0.0447	-0.0824	-0.0787									-0.0523			
y2013x					1.121***	0.11	0.147*	0.113	-2.246***	-0.226	-0.186	-0.0586		1.013***	0.786**	0.627*
Constant	-11.71***	3.291***	10.10***	10.73***	-3.657*	4.320***	11.01***	11.41***	-10.63***	3.180*	9.977***	11.11***	-18.06**	4.903*	11.27***	12.18***
Chi2	1758.97	115.85	42.62	12.35	286.72	83.03	21.45	5.98	1875.59	35.45	8.39	2.73	142.07	6.08	0.42	3.32
Prob>Chi2	0	0	0	0.0004	0	0	0	0.0145	0	0	0.0038	0.0985	0	0.0136	0.5192	0.0686
Observations	11,690	6,504	6,504	6,504	6,581	4,560	4,560	4,560	2,961	994	994	994	960	377	377	377
R-squared	0.618	0.716	0.549	0.516	0.791	0.765	0.582	0.543	0.392	0.597	0.444	0.42	0.477	0.831	0.703	0.638
Adj. R-Squared	0.6161	0.7145	0.5459	0.5131	0.7894	0.7627	0.5779	0.5391	0.3833	0.5801	0.4212	0.3959	0.454	0.8104	0.6672	0.5954

Table 3.6: “Continental Model”: Tests across continents/regions.

Countries										
D. VARIABLES	En	Fr	Gr	It	Sp	Ar	Br	CL	Mx	Ur
	SCP	SCP	SCP	SCP	SCP	SCP	SCP	SCP	SCP	SCP
I. Variables										
freetransfer	-1.035***	-0.500***	-0.781***	-0.685***	-0.444***	-0.721***	-0.489**	0.00401	-0.426*	0.673**
tt_loan	-1.010***	-0.13	-0.385***	-0.335***	-0.481***	-0.345*	-0.594***	0.043	-0.0467	0.582**
ZGOALS1	0.145*	0.198*	0.11	0.131	0.267*	0.42	0.604	1.311***	0.567	1.145
ZGOALS2	0.320***	-0.153	0.0365	0.134	-0.0222	-0.183	0.0206	-0.943*	0.373	4.933***
ZASSISTS1	0.113	0.0041	0.0253	0.263***	0.0758	0.316	0.728	1.550***	-0.0249	0.768
ZASSISTS2	0.0585	0.304**	0.072	-0.0372	0.158	0.957***	-0.501	-0.865	-0.158	-4.180***
ZEXP1	0.117	0.154	0.222*	0.131	0.358***	0.158	-0.417	-0.0838	0.0206	-0.22
ZEXP2	0.152**	0.384***	0.227**	0.270***	0.213*	0.482**	0.361	0.970***	0.643**	-0.833
ZBAD1	0.0713	0.328	0.367**	0.0382	-0.0251	-1.249**	-1.202		0.504	
ZBAD2	-0.166	0.152	0.144	0.183	0.25	-0.268	0.1		0.00825	
<b>IRemDurDays</b>	0.114***	0.0293	0.123***	0.0662***	0.0233	0.201**	0.255***	0.314**	0.016	0.365***
minoraiola	0.137	0.561	-0.066	0.124						
mediabase			2.129**		-0.00072					
gestifute	0.223	0.729*	0.911	-0.105	-0.873		2.206			
lagem	0.152	-0.248	-0.232	0.146	0.102	-0.0339	0.564	0.0574	0.0475	-0.828
lagep	0.364**	0.318*	-0.0433	0.16	0.209	0.269	0.226	-0.168	0.0592	1.911***
lagem2	0.399***	0.812***	0.578***	0.435***	0.508***	0.642**	0.315	0.525	0.507	1.485***
lagep2	-0.419***	-0.437***	-0.118	-0.285***	-0.386***	-0.331***	-0.377**	-0.0273	-0.366**	-1.156***
ldur_hyp	0.0289	0.213***	0.0749	0.131***	0.149*	-0.0423	-0.0717	-0.0884	0.443***	0.699**
lheight	-1.653	-3.287**	3.112*	4.828***	-4.818**	0.157	14.05***	-2.77	-1.29	6.355
lgoo	0.0012	0.0578***	0.0309*	0.0330**	-0.0115	0.0143	-0.00402	-0.00216	-0.0297	-0.000622
lgt	-0.0103	-0.0384	-0.0165	-0.0147	0.0405	0.0099	-0.0838	-0.0169	0.100**	-0.0426
lco_cuclu	0.274***	0.478***	0.393***	0.440***	-0.00408	0.246**	0.16	0.277*	-0.709***	-0.381*
lco_preclu	0.143***	0.200***	0.221***	0.130***	0.210***	0.0458	-0.0647	0.266**	0.207**	0.378***
ltot_preclu	0.0586***	0.00394	0.0167	0.0383***	0.0345*	0.0255	0.0236	-0.00732	0.0254	-0.0243**
ltot_cuclu	0.0337***	-7.10E-05	0.0242	0.0127	0.0826***	0.0405**	-0.0201	0.0125	0.160***	0.00563
footer_both	-0.247	-0.332	0.608		0.494	0.476	0.143		1.261	1.916***
footer_left	0.294*	0.488**	-0.505	-0.111	-0.415	-0.708	-0.505	1.816***	0.199	-0.736**
footer_right	0.244*	0.318	-0.498	-0.0871	-0.381	-0.57	-0.713**	1.939***	0.199	

nat_cont_africa	0.105	0.0558	0.270**	0.151*	0.198		-0.872	-2.137**	0.525	
nat_cont_asia	0.208	0.228	-0.0616	0.853***	-0.0849	-0.855	-0.834		-1.898***	
nat_cont_australia	-0.482*	-1.222**	-0.401	-0.244						
nat_cont_europe	-0.0927	-0.112	0.0493	-0.0707	-0.0887	0.175	0.351*	0.882**	0.127	1.201***
nat_cont_north_america	-0.0927	-0.251	-0.271	0.186	0.0111	0.18		-2.042**	-0.0828	
nat_cont_south_america	0.203**	0.0775	0.038	0.444***	0.177	3.543***	0.917		-0.241	
pos_forward	-0.349**	-0.00701	0.300*	0.565***	0.111	0.0518	1.114***	0.573	-0.602	-0.0746
pos_def	-0.297***	-0.169	0.00983	0.274***	0.03	0.184	0.919***	0.535	-0.826*	0.370*
pos_mid	-0.308**	-0.0171	0.228	0.405***	-0.157	-0.00713	0.915***	0.366	-1.009**	1.245
y2013x			0.348*	0.246*	-0.132	0.176	0.657		0.0158	0.575
y2014x	-0.292**	-0.427**	0.134		-0.354**	-0.863*	0.133	-1.754**	0.323	0.119
y2015x	-0.201*	-0.488***	0.269**	-0.0533	-0.186	-0.647	0.0289	-0.285	0.103	-0.189
y2016x	-0.912***	-1.240***	-0.497***	-0.631***	-0.992***	-1.345***	-0.099	-0.0709	-0.0315	0.151
y2017x	0.415***	-0.235	0.479***	0.448***	-0.11	-0.691	-0.205	-0.68	0.325	0.728*
y2018x	0.0422	-0.437**		0.235**				-0.604		
<b>Constant</b>	<b>13.30***</b>	<b>13.24***</b>	<b>7.055***</b>	<b>4.065**</b>	<b>16.70***</b>	<b>8.050**</b>	<b>-2.077</b>	<b>10.11**</b>	<b>13.56***</b>	<b>1.157</b>
<b>Chi2</b>	<b>17.28</b>	<b>1.4</b>	<b>14.32</b>	<b>8.05</b>	<b>0.72</b>	<b>8.94</b>	<b>0.16</b>	<b>2.1</b>	<b>7.22</b>	<b>1.15</b>
<b>Prob&gt;Chi2</b>	<b>0</b>	<b>0.2365</b>	<b>0.0002</b>	<b>0.0045</b>	<b>0.3976</b>	<b>0.0028</b>	<b>0.6923</b>	<b>0.1478</b>	<b>0.0072</b>	<b>0.2839</b>
<b>Observations</b>	<b>1,154</b>	<b>737</b>	<b>677</b>	<b>1,571</b>	<b>602</b>	<b>342</b>	<b>367</b>	<b>94</b>	<b>182</b>	<b>46</b>
<b>R-squared</b>	<b>0.658</b>	<b>0.695</b>	<b>0.624</b>	<b>0.589</b>	<b>0.603</b>	<b>0.583</b>	<b>0.448</b>	<b>0.876</b>	<b>0.718</b>	<b>0.996</b>
<b>Adj. R-Squared</b>	<b>0.6448</b>	<b>0.6765</b>	<b>0.5981</b>	<b>0.5782</b>	<b>0.5741</b>	<b>0.5307</b>	<b>0.3825</b>	<b>0.805</b>	<b>0.6399</b>	<b>0.9856</b>
<b>SSE</b>		<b>374.8554</b>		<b>1114.316</b>	<b>492.5323</b>	<b>238.18095</b>	<b>429.7305</b>	<b>17.541756</b>	<b>87.484216</b>	<b>0.3717068</b>
<b>K</b>		<b>43</b>		<b>42</b>	<b>41</b>	<b>38</b>	<b>39</b>	<b>34</b>	<b>39</b>	<b>32</b>

Table 3.7: “Countries Model”: Tests across countries.

Segmentation – Positions												
D. VARIABLES	Forward				Midfielders				Defenders			
	lfee	OP	SCP	SSP	lfee	OP	SCP	SSP	lfee	OP	SCP	SSP
<b>I. Variables</b>												
freetransfer		-0.970***	-0.544***	-0.191***		-0.903***	-0.494***	-0.189**		-0.834***	-0.519***	-0.228***
tt_loan	-1.531***	-0.822***	-0.499***	-0.159**	-1.345***	-0.752***	-0.415***	-0.127	-1.253***	-0.593***	-0.337***	-0.0711
ZGOALS1	0.235***	0.304***	0.316***	0.268***	0.276***	0.273***	0.270***	0.232***	0.0703	0.409***	0.381***	0.366***
ZGOALS2	0.112	0.251***	0.291***	0.264***	0.0316	0.149	0.148	0.13	0.0822	0.044	0.0742	0.0414
ZASSISTS1	-0.048	0.249***	0.256***	0.284***	0.00763	0.119	0.115	0.118	0.163	0.245***	0.263***	0.266***
ZASSISTS2	0.0858	0.187***	0.182***	0.186***	0.0744	0.182*	0.170*	0.159*	-0.0049	0.0473	0.0277	0.0141
ZEXP1	0.268***	0.0875	0.0683	0.00925	0.267***	0.138*	0.13	0.108	0.409***	0.179**	0.103	0.0558
ZEXP2	0.0563	0.138**	0.138**	0.134**	0.159**	0.240***	0.277***	0.246***	0.088	0.247***	0.293***	0.289***
ZBAD1	-0.095	-0.119	-0.156	-0.143	-0.041	0.331**	0.389***	0.410***	-0.168	0.113	0.162	0.118
ZBAD2	0.0707	-0.00863	0.0204	-0.0416	-0.0507	0.108	0.0853	0.0931	0.175*	0.187*	0.206*	0.109
<b>lRemDurDays</b>	<b>0.0954***</b>	<b>0.0492***</b>	<b>0.0823***</b>	<b>0.0654***</b>	<b>0.116***</b>	<b>0.0377*</b>	<b>0.0681***</b>	<b>0.0473**</b>	<b>0.0875***</b>	<b>0.100***</b>	<b>0.120***</b>	<b>0.0948***</b>
minoraiola	-0.338	0.479*	0.490*	0.580**	-0.0742	-0.507*	-0.38	-0.35	-0.3	0.14	0.0517	0.132
mediabase	-0.416				0.0783	0.195	0.25	0.272	1.243	2.075**	1.952**	2.148**
gestifute	0.086	0.175	-0.0219	-0.122	-0.113	0.578	0.0655	0.082	0.273	-0.348	-0.334	-0.711*
lagem	-0.136	-0.167	0.0921	0.153	0.2	-0.215	-0.0804	-0.0298	-0.0285	0.00422	0.212	0.211
lagep	0.391***	0.837***	0.338***	0.239**	0.516***	0.495***	0.115	0.0208	0.378**	0.664***	0.221*	0.106
lagem2	0.152	-0.0902	0.498***	0.472***	-0.0788	-0.0852	0.599***	0.587***	0.0239	-0.288***	0.363***	0.387***
lagep2	-0.271***	-0.393***	-0.423***	-0.382***	-0.362***	-0.251***	-0.324***	-0.266***	-0.359***	-0.354***	-0.395***	-0.322***
ldur_hyp	0.155***	0.940***	0.0574*	0.0439	0.272***	1.030***	0.145***	0.103**	0.237***	1.054***	0.154***	0.120***
lheight	1.415	-2.296**	-2.485***	-2.958***	0.08	-0.0473	-0.443	-0.941	6.486***	6.242***	6.215***	5.671***
lgoo	0.0515***	0.0324***	0.0301***	0.0234**	0.0472***	0.00452	0.0105	0.0018	0.0208	0.0389***	0.0437***	0.0315***
lgt	-0.00655	0.0348*	0.0372**	0.0366*	0.0112	0.00895	0.00454	0.00993	0.0303	0.00548	0.00361	0.00169
lco_cuclu	-0.220***	0.0598**	0.0601**	0.0714***	-0.352***	0.119***	0.102***	0.115***	-0.259***	0.113***	0.112***	0.119***
lco_preclu	-0.0275	0.0994***	0.0993***	0.0934***	-0.0368	0.0604**	0.0658**	0.0558*	-0.0442	0.0902***	0.108***	0.113***
ltot_preclu	0.167***	0.0439***	0.0384***	0.0376***	0.190***	0.0505***	0.0421***	0.0415***	0.225***	0.0493***	0.0412***	0.0402***
ltot_cuclu	0.450***	0.0571***	0.0482***	0.0432***	0.509***	0.0456***	0.0377***	0.0325***	0.491***	0.0610***	0.0552***	0.0520***
footer_both	-0.412**	-0.0788	-0.0179	-0.0383	0.305	-0.303	-0.205	-0.26	0.317	-0.0825	-0.0908	-0.168
footer_left	0.344**	-0.1	-0.133	-0.0965	-0.251	0.292*	0.285*	0.314*	-0.215	-0.0175	-0.0628	-0.0351
footer_right	0.384**	-0.0099	-0.057	-0.029	-0.162	0.194	0.193	0.227	-0.229	0.00283	-0.0554	-0.041
nat_cont_africa	0.207***	0.195***	0.167***	0.175***	0.241***	0.366***	0.328***	0.276***	0.025	0.295***	0.266***	0.261***

nat_cont_asia	0.164	0.17	0.2	0.226*	0.0614	-0.18	-0.142	-0.145	0.229	0.604***	0.515**	0.480**
nat_cont_australia	-0.0811	-0.623***	-0.567***	-0.542**	0.409	-0.902***	-0.876***	-0.840***	-0.0813	0.541**	0.541**	0.602**
nat_cont_europe	-0.0389	0.137**	0.159***	0.168***	-0.121	0.0141	0.0625	0.0448	-0.111	0.0307	0.0737	0.0619
nat_cont_north_america	0.446***	-0.213**	-0.149	-0.104	-0.13	-0.333***	-0.300***	-0.259**	-0.209	-0.201**	-0.152	-0.107
nat_cont_south_america	0.425***	0.288***	0.304***	0.284***	0.400***	0.141*	0.111	0.0558	0.230**	0.228***	0.248***	0.246***
y2013x	-0.0773	-0.127	-0.0644	-0.0605		0.382***	0.367***	0.361***		0.0203	0.0802	0.0947
y2014x		-0.184*	-0.164*	-0.177*	-0.108	0.0825	0.0117	0.0198	-0.0231	-0.0176	0.0329	0.0355
y2015x	0.0223	-0.107	-0.105	-0.085	0.0338	0.0116	-0.0758	-0.0831	0.146	0.112	0.126	0.112
y2016x	0.284***	-0.683***	-0.670***	-0.711***	0.251	-0.636***	-0.736***	-0.814***	0.263	-0.571***	-0.576***	-0.675***
y2017x	0.347***	-0.0296	0.0187	0.0229	0.307	0.176*	0.115	0.0957	0.404	0.198**	0.231***	0.205**
y2018x	0.452***				0.386**				0.493*			
Constant	-0.332	6.970***	13.91***	14.46***	-0.307	4.275***	11.49***	12.32***	-7.147***	-3.526***	3.355**	4.247***
<b>Chi2</b>	<b>21.94</b>	<b>63.23</b>	<b>19.22</b>	<b>3.14</b>	<b>45.35</b>	<b>34.43</b>	<b>14.78</b>	<b>6.56</b>	<b>50.5</b>	<b>34.18</b>	<b>13.98</b>	<b>3.51</b>
<b>Prob&gt;Chi2</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0.0766</b>	<b>0</b>	<b>0</b>	<b>0.0001</b>	<b>0.0104</b>	<b>0</b>	<b>0</b>	<b>0.0002</b>	<b>0.0611</b>
<b>Observations</b>	<b>1,127</b>	<b>2,460</b>	<b>2,460</b>	<b>2,460</b>	<b>822</b>	<b>1,864</b>	<b>1,864</b>	<b>1,864</b>	<b>826</b>	<b>1,923</b>	<b>1,923</b>	<b>1,923</b>
<b>R-squared</b>	<b>0.751</b>	<b>0.726</b>	<b>0.575</b>	<b>0.544</b>	<b>0.747</b>	<b>0.707</b>	<b>0.517</b>	<b>0.487</b>	<b>0.697</b>	<b>0.729</b>	<b>0.556</b>	<b>0.519</b>
<b>Adj. R-Squared</b>	<b>0.742</b>	<b>0.7218</b>	<b>0.5685</b>	<b>0.5368</b>	<b>0.7344</b>	<b>0.7009</b>	<b>0.5059</b>	<b>0.4753</b>	<b>0.6815</b>	<b>0.7229</b>	<b>0.5471</b>	<b>0.5087</b>

Table 3.8: "Universal Model": Tests across Positions

Europe (Top Five) - Across Positions									
Variables	Forwards			Midfielders			Defenders		
	SCP	SCP	SCP	SCP	SCP	SCP	SCP	SCP	SCP
	(all)	1 add, Var	2 Add, Var	(all)	1 add, Var	2 Add, Var	(all)	1 add, Var	2 Add, Var
freetransfer	-0.738***	-0.806***	-0.682***	-0.587***	-0,731***	-0,586***	-0,705***	-0,721***	-0,574***
tt_loan	-0.512***	-0.658***	-0.566***	-0,393***	-0,637***	-0,505***	-0,470***	-0,467***	-0,358***
ZGOALS1	0.289***	0.263***	0.178***	0,207***	0,190**	0,162*	0,218***	0,250**	0,251***
ZGOALS2	0.249***	0.168**	0.0686	0,242***	0.158	0,155*	0.0292	0.069	0.000866
ZASSISTS1	0.219***	0.245***	0.182***	0.102	0.0689	0.00479	0,172**	0,208**	0.143
ZASSISTS2	0.201***	0.150**	0.105	0,144*	0,215**	0.131	0,207**	0.0808	0.0396
ZEXP1	0.0751	0.0336	0.0208	0,125*	0.11	0.11	0,268***	0,230***	0,234***
ZEXP2	0.074	0.171**	0.123*	0,245***	0,208***	0.1	0,181***	0,199***	0.0879
ZBAD1	-0.0925	-0.2	-0.210*	0,382***	0,366***	0.173	0.189	0.0662	0.0382
ZBAD2	-0.0463	-0.0203	-8.07E-05	0.101	0.0308	0.0433	0,214**	0,305**	0,254**
<b>lRemDurDays</b>		<b>0.0668***</b>	<b>0.0415***</b>		<b>0,0412*</b>	<b>0.00381</b>		<b>0,0756***</b>	<b>0,0552***</b>
<b>lsalarp</b>			<b>0.379***</b>			<b>0,410***</b>			<b>0,332***</b>
minoraiola	0.508**	0.452*	0.0588	0.373	0.313	0.375	0.165	0.0758	0.0575
mediabase	0.524			0.42	0.329	0.314	2,693***	1,806*	0.848
gestifute	0.089	-0.0241	0.00924	0.257	0.0909	0.0825	0,466*	0.226	0.00501
lagem	0.00805	0.207	0.286**	-0,418***	0.105	0.0788	-0,275**	0.165	0,271*
lagep	0.405***	0.426***	-0.0089	0.161	0,283*	0.00564	0.167	0,238*	0.00962
lagem2	0.487***	0.396***	0.475***	0,741***	0,578***	0,730***	0,669***	0,359***	0,424***
lagep2	-0.425***	-0.453***	-0.289***	-0,333***	-0,370***	-0,302***	-0,335***	-0,387***	-0,322***
ldur_hyp	0.0474	0.0128	0.0595*	0,162***	0,130***	0,139***	0,110***	0,116**	0,148***
lheight	-2.680***	-3.948***	-2.600***	-1.321	-1.091	-1	5,646***	6,640***	4,884***
lgoo	0.0325***	0.0557***	0.0423***	0.000949	0.00126	0.003	0,0177*	0,0411***	0,0318***
lgt	0.0466***	0.0203	-0.00562	0,0325*	0.0281	0,0401*	0.0126	0.0104	0.00397
lco_cuclu	0.0880***	0.149***	0.150***	0,0810***	0,147***	0,0872**	0,105***	0,106**	0,0893**
lco_preclu	0.0895***	0.119***	0.0987***	0,0485**	0,0848**	0.036	0,0688***	0,112***	0.0462
ltot_preclu	0.0359***	0.0441***	0.0204***	0,0387***	0,0361***	0,0220**	0,0569***	0,0544***	0,0391***
ltot_cuclu	0.0603***	0.0484***	0.0196***	0,0454***	0,0326***	0,0220***	0,0628***	0,0792***	0,0549***
footer_both	-0.145	0.0562	-0.128	0.147	0.00164	0.0694	-0,573**	0.205	0.203
footer_left	0.0969	-0.0977	0.0858	0.114	0.103	0.0447	0.31	0.0347	0.167
footer_right	-0.00603	-0.102	0.115	0.105	0.0468	0.00159	0.352	0.0368	0.161
nat_cont_africa	0.248***	0.147**	0.0852	0,346***	0,330***	0,196***	0,246***	0,278***	0,202***
nat_cont_asia	0.490***	0.375**	0.404***	0.0379	0.19	0.29	0,472**	0.359	0.212
nat_cont_australia	-0.721***	-0.774***	-0.655***	-0,570**	-0,535*	0.0408	0.117	0.196	0.0329
nat_cont_europe	0.00531	0.0121	0.0652	0.0577	0.131	0.0269	0.03	0.0506	0.0838
nat_cont_north_america	0.186*	0.252*	0.257**	0.0224	0.169	0.0943	0.0218	0.122	0.0396
nat_cont_south_america	0.424***	0.376***	0.243***	0,399***	0,302***	0,269***	0,380***	0,344***	0,301***
y2013x	-0.0246	-0.0155	-0.0624				0,191*	0.105	0.0831
y2014x	-0.127	-0.157	-0.191**	0.101	-0,358***	-0,542***	0.0192	0.066	-0,228**
y2015x	-0.0175	-0.0449	-0.00625	-0,210**	-0,514***	-0,464***	0,169**	0.0443	0.042
y2016x	-0.737***	-0.768***	-0.701***	-0,764***	-1,126***	-1,040***	-0,631***	-0,706***	-0,740***
y2017x	0.274***	0.250***	0.491***	0,200**	0.0469	0.146	0,384***	0,353***	0,483***
y2018x				0.0358	-0,410***	-0,368***			
Constant	14.47***	15.30***	12.54***	13,43***	13,30***	11,73***	4,752***	2,955**	3,783***

<b>Chi2</b>	<b>18.01</b>	<b>4.96</b>	<b>5.05</b>	<b>5.92</b>	<b>5.69</b>	<b>41.57</b>	<b>4.76</b>	<b>3.82</b>	<b>6.72</b>
<b>Prob&gt;Chi2</b>	<b>0</b>	<b>0.0259</b>	<b>0.0246</b>	<b>0.015</b>	<b>0.171</b>	<b>0</b>	<b>0.0291</b>	<b>0.0507</b>	<b>0.0095</b>
<b>Observations</b>	<b>2,820</b>	<b>1,651</b>	<b>1,651</b>	<b>2,334</b>	<b>1,329</b>	1.329	<b>2.506</b>	<b>1.36</b>	1.36
<b>R-squared</b>	<b>0.588</b>	<b>0.624</b>	<b>0.713</b>	<b>0.492</b>	<b>0.522</b>	0.627	<b>0.577</b>	<b>0.61</b>	0.676
<b>Adj. R-Squared</b>	<b>0.5824</b>	<b>0.6152</b>	<b>0.706</b>	<b>0.484</b>	<b>0.507</b>	<b>0.615</b>	<b>0.571</b>	<b>0.6</b>	<b>0.666</b>

Table 3.9: “European Model”: Tests across Positions

<b>Latin America - Across Positions</b>									
<b>Variables</b>	<b>Forwards</b>			<b>Midfielders</b>			<b>Defenders</b>		
	<b>SCP</b>	<b>SCP</b>	<b>SCP</b>	<b>SCP</b>	<b>SCP</b>	<b>SCP</b>	<b>SCP</b>	<b>SCP</b>	<b>SCP</b>
	<b>(all)</b>	<b>1 add, Var</b>	<b>2 Add, Var</b>	<b>(all)</b>	<b>1 add, Var</b>	<b>2 Add, Var</b>	<b>(all)</b>	<b>1 add, Var</b>	<b>2 Add, Var</b>
freetransfer	-0.204**	-0.415**	-0.180***	0.0199	-0.455**	-0.342*	-0.0368	-0.239	-0.139
tt_loan	-0.158*	-	-0.217***	0.00534	-0.387*	-0.323*	-0.115	-0.349	-0.365*
ZGOALS1	0.469***	0.887***	0.11	0.512**	-0.391	-0.152	0.491*	0.638	0.409
ZGOALS2	0.534***	0.2	-0.0447	0.430**	0.226	0.025	-0.00462	-0.289	-0.345
ZASSISTS1	0.318	0.041	0.0664	0.801***	0.621*	0.718**	0.255	0.532	0.142
ZASSISTS2	0.716***	0.352	0.0571	0.0625	0.32	0.186	-0.107	-0.477	-0.804*
ZEXP1	0.00888	-0.211	-0.0919	0.0408	0.199	0.108	0.0726	-0.413	-0.0776
ZEXP2	-0.197	0.152	0.00192	0.321**	0.546*	0.44	0.802***	0.976***	0.793***
ZBAD1	-0.631	-0.744	-0.0701	-0.531	0.183	-0.0343	-0.752	-1.305*	-0.806
ZBAD2	0.508			0.25	-0.852	-0.591	-0.0554	-0.16	-0.303
lRemDurDays		<b>0.0649</b>	<b>0.0510***</b>		<b>0.364***</b>	<b>0.283***</b>		<b>0.309***</b>	<b>0.252***</b>
lsalarp			<b>0.957***</b>			<b>0.264***</b>			<b>0.347***</b>
minoraiola				-1.075**			-	-	-
gestifute				-1.645			1.499	2.724*	1.469
lage	-14.71**	7.362	-63.31***	-57.28***	-61.06***	-62.03***	-43.83***	-	-
lage2	1.303	-1.853	8.574***	7.598***	8.322**	8.437***	5.526***	9.089***	10.92***
ldur_hyp	0.137***	0.0171	0.138***	0.210***	0.0565	0.0796	0.137**	0.00867	0.0279
lheight	-0.643	1.192	0.584	7.898***	4.22	2.909	4.806***	2.143	1.265
lgoo	0.00927	-0.0262	0.000201	0.0105	0.0498	0.04	0.0147	0.0720**	0.0537**
lgt	0.0710***	0.0892**	-0.011	0.0645***	0.00729	-0.0281	0.107***	0.0523	0.0444
lco_cuclu	-0.0197	0.138	0.0576*	0.0312	-0.0779	0.0597	0.0309	0.0175	0.0049
lco_preclu	0.0433	0.0636	-0.0191	0.0423	-0.12	-0.097	-0.011	-0.0444	-0.107
ltot_preclu	0.0124**	0.00808	0.00437	0.0200***	0.0338	0.0334*	0.0255***	0.0231	0.0267
ltot_cuclu	0.0387***	0.0198	0.00281	0.0211***	0.00532	-0.0124	0.0173***	0.0259	0.00871
footer_both	0.266	-1.331**	-0.153	0.135	0.19	0.198	0.42	0.819	0.960*
footer_left	-0.0968	0.154	0.112	0.227	-0.244	-0.415	-0.118	-0.00098	0.189
footer_right	-0.159	0.582	0.228**	0.0491	-0.56	-0.484	-0.114		
nat_cont_africa	-0.259	-0.43	-0.0805	-0.556	0.771	0.466	-0.212		
nat_cont_asia	-0.584*	-1.245**	-0.112						
nat_cont_australia	0.265***	0.236	-0.0301						
nat_cont_europe	-0.440***	0.348	0.0957	0.238***	0.321*	0.284*	0.14	0.551***	0.467***
nat_cont_north_america	-0.313*	0.43	0.0535	0.000148	-0.0218	-0.0534	0.155	0.263	0.377



nat_cont_south_america				0.0101	-0.126	-0.0972	0.252	0.213	0.273
y2013x	0.525**								
y2014x	0.116	0.265	0.164	-0.395***	-0.605	-0.273	0.0287	0.332	0.0311
y2015x	0.261	0.0527	0.038	-0.320***	-0.664	-0.38	-0.0886	0.147	-0.219
y2016x	-0.082	-0.276	0.0214	-0.539***	-1.346***	-1.078***	-0.332***	-0.208	-0.606**
y2017x	0.235	-0.235	0.082	-0.173	-1.036**	-0.706*	-0.0904	0.125	-0.295
y2018x		0.0229	0.17	0.0311	-0.506	-0.49	-0.184	0.0913	-0.332
constant	47.45***	6.693	122.1***	109.4***	119.1***	120.6***	90.97***	123.5***	145.2***
Chi2	9.2	11.75	14.81	0.08	0.14	6.28	2.59	7.75	10.97
Prob>Chi2	0.0024	0.0006	0.0001	0.7717	0.7086	0.0122	0.1075	0.0054	0.0009
Observations	1,562	391	391	1,353	259	259	1,358	307	307
R-squared	0.442	0.453	0.958	0.5	0.586	0.654	0.456	0.426	0.553
Adj R-Squared	0.4302	0.4022	0.9543	0.4871	0.5248	0.6012	0.4426	0.3594	0.4986

Table 3.10: “Latin American Model”: Tests across Positions

### 3.5.2 Price Index

The main objective of this section is to display the variation of prices in the football transfers’ market across years. In general, it may help football clubs evaluate the players based on current market prices and bargain with selling/buying clubs based on indicators provided by such price indices. Thus, building the apparent and hedonic price indices could be a breakthrough and an achievement falling in the same context. It may also help to uncover any inflated figures in the market. From a financial and economic point of view, the price movements detected by the price indices shed light on probable market trends or shocks. In this study, the normal apparent price indices generated were based on the average annual transfer fees, normalizing at a reference starting point in the year 2007. Those indices were then displayed together (Fig 3.1) with the hedonic price indices generated based on the coefficients (table 3.11) obtained from the test results.



Figure 3.1: Price Index

	transfer fee	the naive TF price index	package_curr	the naive package price index	from 9699 obs	EXP	hedonic TF price index from 9699 obs	LPACK from 19K obs	EXP	hedonic package PI from 20k obs
2007	673922	100	562661	100	-0.141867	0.867737	100	0.274714	1.316155	100
2008	482098	111.59	414766	115.31	-0.1495	0.861139	111.08	0.18984	1.209056	136.70
2009	468120	108.35	400019	111.21	-0.20213	0.81699	105.38	0	1	113.06
2010	315042	72.92	264720	73.60	-0.25366	0.775956	100.09	-0.14824	0.862221	97.49
2011	432039	100.00	359681	100.00	-0.25454	0.775272	100.00	-0.12278	0.884455	100.00
2012	147091	34.05	139259	38.72	-0.2782	0.757148	97.66	-0.26931	0.763906	86.37
2013	450737	104.33	493816	137.29	-0.22019	0.802363	103.49	0.273025	1.313933	148.56
2014	395632	91.57	1012767	281.57	-0.12945	0.878583	113.33	0.490921	1.63382	184.73
2015	581671	134.63	1453719	404.17	-0.12864	0.879288	113.42	0.527964	1.695477	191.70
2016	651653	150.83	1301605	361.88	-0.080296	0.922843	119.03	0.018604	1.018778	115.19
2017	914181	211.60	1876970	521.84	-0.06006	0.941705	121.47	0.651606	1.91862	216.93
2018	1574154	364.35	3585939	996.98	0	1	128.99	0.76738	2.154114	243.55

Table 3.11: Price Indices Computation

Price indices are good economic indicators about how healthy markets are. The different levels of inflation generated by the apparent price indices and the hedonic price indices in this study may confirm the presence of inflation in the football industry. Still, the concrete value of inflation is determined according to the values of coefficients generated. The stronger the model, the easier it reveals economic shocks in the market.

### **3.6 Conclusion**

More recently, a lot of money is circulating in the football industry. All components in the football market have increased in price. This section approached the football players' financial values from a new perspective tested using the hedonic pricing methodology. This paper's approach utilized real data to determine the correlations between players' characteristics, performance, and their complete financial earnings and transfer fees. The main findings of the section are that there could be different markets and different pricing evaluations across countries, position, and time. Moreover, the volume of data (more than 87,000 transfer contracts and additional 20,000 salaries) helped to run multiple analysis for transfer fees and packages across different positions and regions. The number of transfer contracts and salaries for 12 years dataset (2007/2008 ending in 2018/2019) helped to generate price indices using the transfer fees and the packages. There are differences between the apparent price index and the hedonic price index, but that does not deny that the indices helped us to contribute one step further towards understanding the dynamics of price movements. We believe that this paper can lead to further future research to uncover more secrets lying within the booming figures of some players, especially superstars. We expect it also to converge with the Financial Fair Play (FFP) regulations objectives and pricing investigations.

## CHAPTER FOUR

### 4 PRICING FOOTBALL TRANSFERS: AN EXPLORATORY STUDY INTO THE GLOBAL MARKET USING SYNTHETIC DATA

#### 4.1 Introduction

This chapter follows a similar approach to the previous chapter in pricing football players with a different type of data set. While most studies used similar independent variables like the players' characteristics (age, height, goals, assists and appearance), his competitive record and fame, and some characteristics of the contract between him and the club, this study relied on 'synthetic scores' data which represent players' skills given by experts. The reason behind using such type of data is that every player, transferred or not, has got an evaluation by experts and therefore we were able to include a more comprehensive subset of players in this study, the reason behind this strategy was to contribute a solution to the selection bias problem which has not been addressed properly in previous studies. The results show that valuation models can be consistent across time or across space, the difficulty of tackling selection bias and heteroscedasticity in a global model using the *transfer fee* alone was solved through aggregating various elements of players' cost in one overall *package*. Such aggregation generated promising results and findings.

Accordingly, we try to build a random *global* sample by starting from an existing player universe and appending data gathered over the internet. Second, we develop new measures of transfer prices which are not susceptible to being censored (i. e. without selectivity) and which reflect the expectations of the stakeholders. The overall design is demanding in terms of information quality and quantity, but the modelling of transfer cost is better than with transfer fees. Thus, such a global model can explain satisfactorily the expenditure by buying clubs (table 4.1) from 2007 to 2018 on players' transfers and wages. We achieve homoskedasticity

on segments of the global market proven to be consistent by a series of Chow tests. This might illustrate how the global market for transfers might be adequately studied using this methodology on a larger scale.

Competition	Competition	Arrivals	Expenditure (€)	Departures	Income (€)	Balance (€)
Premier League	England	5588	15.39bn	5990	8.11bn	-7,276.90m
Championship	England	8247	2.03bn	9113	2.69bn	660.18m
Serie A	Italy	10952	9.67bn	11136	8.15bn	-1,523.50m
Serie B	Italy	8288	472.51m	8644	1.00bn	529.37m
LaLiga	Spain	4058	7.90bn	4373	6.74bn	-1,165.87m
Bundesliga	Germany	3002	5.47bn	3315	4.38bn	-1,089.90m
Bundesliga 2	Germany	3059	364.13m	3360	744.64m	380.51m
Ligue 1	France	3512	5.01bn	4075	5.31bn	301.87m
Liga NOS	Portugal	5532	1.32bn	5854	3.10bn	1.78bn
Eredivisie	Netherlands	3111	937.92m	3692	2.08bn	1.15bn
Jupiler Pro League	Belgium	3824	824.77m	4157	1.32bn	499.32m
Super League	Switzerland	1943	220.96m	2143	579.39m	358.43m
Scottish Premiership	Scotland	2500	244.45m	2898	324.15m	79.70m
Premier Liga	Russia	3231	2.19bn	3324	1.48bn	-712.72m
Süper Lig	Turkey	5370	1.28bn	5567	912.50m	-363.89m
Camp. Brasileiro Série A	Brazil	9748	1.07bn	10897	2.59bn	1.52bn
Superliga	Argentina	2792	424.47m	3154	957.06m	532.60m
Chinese Super League	Chinese	2232	1.97bn	2169	558.36m	-1,409.38m
Major League Soccer	USA	4147	388.28m	4286	167.55m	-220.73m
Saudi Prof. League	KSA	2045	469.44m	2242	93.07m	-376.37m
Qatar Stars League	Qatar	1391	268.32m	1535	58.50m	-209.83m
Arabian Gulf League	UAE	1383	285.98m	1484	112.79m	-173.19m
Premier Liga	Ukraine	3077	710.01m	3422	760.37m	50.37m
Super League 1	Greece	4555	363.81m	5137	405.31m	41.50m
Liga 1	Romania	4385	212.03m	4582	294.08m	82.05m

Table 4.1: Balance of Trade (2007 – 2020). Source: Transfermarkt

The rest of the chapter is organized as follows. Section 4.2 discusses the data, while Section 4.3 discusses some econometric issues. This is followed by a discussion of the results (Section 4.4), and Section 4.5 presents some concluding remarks.

## 4.2 Data

To solve the selection puzzle, we need not to start with a set of transferring players but with a somewhat *representative* set of players, to which we will later append further data such as transfer prices (if any) and salaries. The FIFA games series by EA Sports or Pro Evolution Soccer by Konami on PlayStation and other platforms provide a satisfying set of players, in as much millions of videogames and football enthusiasts are satisfied with it and do not look further. Player information is compiled in a so-called “Futhead” database which is available on fan pages around the internet: it contains tens of thousands of players rated by experts for their skills. They provide data for all players, *not just players who transfer*. The data was then supplemented by transfers data between 2007/2008 and 2018/2019 and salaries between 2012/2013 and 2018/2019, all harvested on the internet, notably from *transfermarkt* and from *sofifa*<sup>4</sup>.

From these transfer data, it appears that the transfer fees are not normally distributed (Appendix 2, Figure AP 2.1, provides kernel density estimation as well as common statistical tests of normality). Even when zero transfer fees and loans are taken out of the sample, the distribution does not appear to be normal, nor are the logarithms or any simple transformation of the transfer fees. This has an important implication since it means the Heckman correction, which rests on a normal dependent variable, could not be applied to correct for possible selection.

Most variables from the database are usual (see the complete list on Appendix 2, Table AP 3.1) except the number of days remaining in contracts, which we could not obtain for every player,

---

• <sup>4</sup> As well as numerous known sites such as “[www.google.com](http://www.google.com)”, “[www.bigsoccer.com](http://www.bigsoccer.com)”, “[www.tifosobilanciato.it](http://www.tifosobilanciato.it)”, “[www.footyrate.com](http://www.footyrate.com)”, “[www.sportune.fr](http://www.sportune.fr)”, “[www.calcioefinanza.com](http://www.calcioefinanza.com)”, “[www.totalsportek.com](http://www.totalsportek.com)”, “[www.sportlens.com](http://www.sportlens.com)”, “@swissramble”, [www.deloitte.com](http://www.deloitte.com) (Deloitte Football Money League), Sky Sports, and Sports Intelligence Report.

but which has never been tested so far to value transfer contracts and was thus worth trying even if it reduced the number of observations greatly. Moreover, age is measured on both sides of the age of peak transfer price: *lagem* (and *lagem2*) measure the years below 26 years while *lagep* (and *lagep2*) count the years above to test for possible asymmetry of an age effect. Eventually, six skills are singled out, two for each group of positions (forward players = pace + shooting, midfielders = dribbling + passing, defenders = defending + physicality).

Then, the measurement of club performance has to be developed a bit. While several global rankings of clubs do exist, they did not contain all the clubs we had in the database. We thus had to develop an endogenous measure. The idea was to count clubs especially active on the transfer market by counting the number of transfer contracts in the database: *co\_cuclu* and *co\_preclu* provide such a count for the current (*i. e.* buying) and previous (*i. e.* selling) clubs. The magnitude of transactions is recorded by *tot\_cuclu* and *tot\_preclu* which sum up the total value of transfers (in the database) for the current (*i. e.* buying) and previous (*i. e.* selling) clubs. Since the database only spans on seven seasons, it is likely that the hierarchy of clubs does not move much; on a longer time span, it would have been appropriate to consider those measures in a moving time window.

Country	# obs.	Country	# obs.
Argentina	1010	Italy	2819
Brazil	155	Mexico	1344
Chile	334	Spain	948
China	124	USA	1202
England	2212	Uruguay	165
France	1355	Other	1133
Germany	1250	Total	14051

Table 4.2: Unique observations in our database per country 2007-2019



Eventually, we could not append the full set of information to every player in the Futhead database since only 14,000 observations of salary (table 4.2) and 13,500 observations of contract duration were gathered, the intersection being around 8,000. Nevertheless, the database may be considered to retain the same properties as the starting universe (*i. e.* the Futhead database). Not only that our sample is larger than all previous studies, but it is also internationally diversified, so it makes it possible to ask whether there is a global transfer fee pricing function or whether the pricing function is segmented. It is worth mentioning this global approach to the subject has never been tried so far in the literature.

### 4.3 Getting around selection issues

While most studies focus on transfer fee, it might be worth refining the analysis before we decide on a dependent variable to be explained by the hedonic analysis. Transfer is a bargain between three sides: a buying club paying to a selling club a transfer fee and to a player some future (certain) salaries and (uncertain) bonuses (one may add the agents, but we do not have figures about their earnings and assume they perceive a percentage of the other payments). From the player perspective, he may want to extract the maximum out of the various clubs he is going to play with; hence his program is to

$$Max \sum_{t=1}^{\infty} \frac{CF_t}{(1+r)^t}$$

where  $CF_t$  represent the sum of incoming cash flows per period, comprising fixed salary and contingent payments (*i. e.* various bonuses as well as sponsorship revenues). This latter quantity is an expectation since bonuses are contingent on objectives, and future salaries

beyond the contract are not known, hence a more developed expression of this quantity for a contract lasting  $n$  period should be:

$$Max \sum_{t=1}^n \frac{w_t}{(1+r)^t} + EU \left( \sum_{t=1}^{\infty} \frac{\widetilde{CF}_t}{(1+r)^t} \right)$$

Where the tilde ‘ $\widetilde{CF}$ ’ denotes a random quantity, whose expected utility might depend on the player’s risk preferences. From an objective point of view, now, the club is willing to minimize the cost incurred when hiring the player, and this cost breaks down into a transfer fee plus an agreed-on salary for the duration of the contract, and some additional *contingent* costs such as bonuses, which are not known on the day the contract is signed (but the list of events triggering bonuses may be in the contract):

$$Min TF + \sum_{t=1}^n \frac{w_t}{(1+r)^t} + \sum_{t=1}^n \frac{\widetilde{CF}_t}{(1+r)^t}$$

These quantities may be approximated by (in order of greater complexity):

- a. **The transfer fee (lfee in the variables list):** this has been done by the previous studies.
- b. **An “objective package” (OP in the variables list):** is the sum of transfer fee and annual salary during the duration of current/new contract. By objective we mean we can objectively measure it as long as we have the player salary and duration of the contract, as well as the transfer fee when applicable, we thus compute:

$$OP = TF + \sum_{t=1}^n \frac{w_t}{(1+r)^t}$$

which we believe to be an approximation for the whole

$$TF + \sum_{t=1}^n \frac{w_t}{(1+r)^t} + \sum_{t=1}^n \frac{\widetilde{CF}_t}{(1+r)^t}$$

The advantage over the raw transfer fee is not just to add some marginal information for transferring players: taking salaries into consideration guarantees that the dependent variable

is not censored. For better adequacy, the remaining contract duration should be taken into account. In order to ensure the model is meaningful, we also included in the database players on loan and players whose contract has ended, which should thus transfer for free. Those latter players help test the consistency of estimations provided by the model, since strictly speaking, a player with six months remaining in the contract should not have a package very different from a player with a contract that just ended, albeit the distribution between the club and the player may differ significantly. Our model is not suitable to analyse this effect, though.

c. A “**Subjective Complete Package**” (SCP in the variables list): This package is *complete* since it features all elements of costs, it is *subjective* as well since there is no objective assessment of it all:

$$Total\ cost = TF + PV\ of\ expected\ future\ income\ in\ t = TF + \sum_{i=1}^{+\infty} \frac{E(\tilde{w}_{t+i}|w_t)}{(1+r)^i}$$

This package features undisclosed elements (such as the contingent payment scheme) and a double uncertainty, both on the realization of the contingent events and on what will happen beyond the horizon of contract. We can think of all those elements to be conditional to the current salary, and it is not unreasonable to think the future salaries can be expected to vary according to the cross-sectional variation of salaries in the base. That is to say, when a player’s age grows by one unit, his salary is adjusted according to the average variation for players of his age and the probability that he remains a professional player is given by the average probability of players of his age. Those salaries and probabilities certainly do not evolve uniformly across the spectrum of all players (figure 4.1), but this coarse approximation of evolution patterns is a starting point to compute this subjective package as:

$$SCP \approx TF + \sum_{i=age}^{42} \frac{w_{age}}{E(w_{age})} E(w_{i+1}) \times \frac{number\ of\ players\ of\ age\ i + 1}{number\ of\ players\ of\ age\ i}$$

Where  $E(w_{age})$  and *number of paid players of age i* are taken from the whole database.

Eventually, SCP can be written as:

$$SCP \approx TF + w \times multiplier (age)$$

Where multiplier has to be estimated from the wage distribution in the sample (Figure 4). It should be mentioned here that the income multiplier was commonly used by UEFA in the 1990s to determine the transfer prices of players between European football clubs. This price had to be at least equal to the gross salary of the player multiplied by a coefficient depending on the age of the player (UEFA, art. 3, 1992). Accountants such as Morrow (1999) and Scarpello and Theeke (1989) criticized the inconsistency of the method with standard economic theory. The main difference between our approach and the UEFA-1990's own is that our multiplier is estimated from the data.

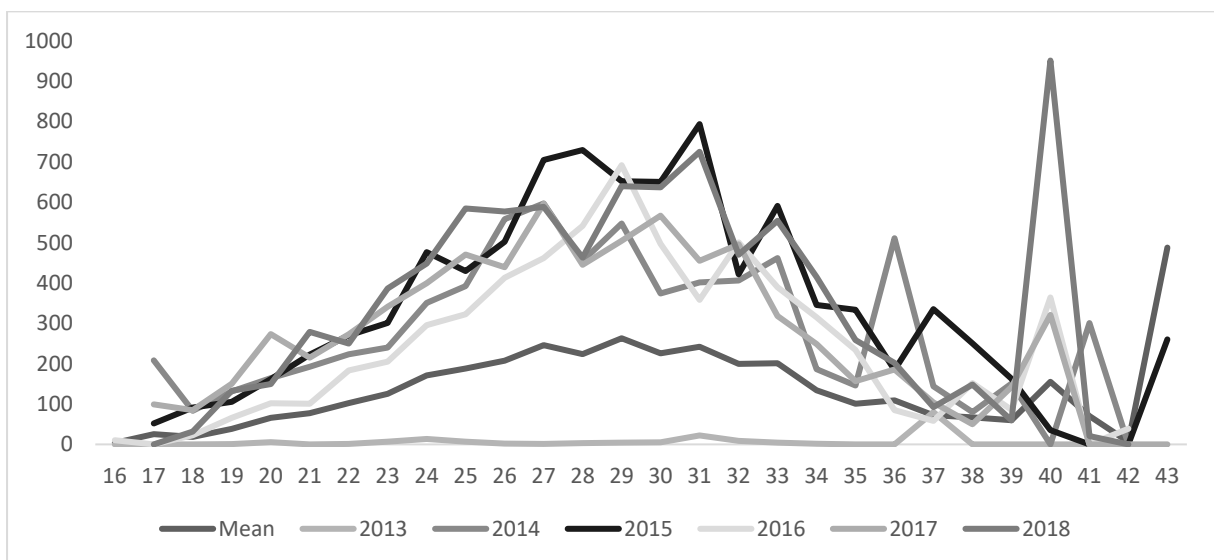


Figure 4.1: Indices average salaries per age per season.

Eventually, the player may be interested in the transfer fee as a signal for the willingness of the club to pay, but it is likely that he is concerned only by what he will take from his club; hence we can define:

**d. A Subjective Salary Package (SSP):** This package assumes clubs keep the player during his whole career. If markets and information were perfect, this should match the income generated by the player, hence:

$$\sum_{t=1}^n \frac{w_t}{(1+r)^t} + \sum_{t=1}^{\infty} \frac{\widetilde{CF}_t}{(1+r)^t}$$

We assumed as with the SCP that the future salaries are dependent on the current relative salary and the average evolution in the database hence:

$$SSP \approx w \times multiplier (age)$$

It seems pretty obvious that the complete specification of the subjective package is well beyond our current knowledge of the stakeholders: we do not have data on the risk preferences of football players, nor on the interest rate players and clubs consider to discount future opportunities. We had to do a series of assumptions to discuss the general idea that we want to capture a non-linear relationship between transfer fee, age and current salary. We thus assumed players to be somewhat risk-neutral and to adopt a zero-discount rate, which is consistent both with the current state of market interest rates and with the apparent preference for the present of football players.

It is quite obvious that the aforementioned packages do not provide a *direct* estimate for the transfer fee. The transfer fee can be computed very simply, though, from the packages predicted by the models, since:

$$\widehat{transfer\ fee} = \widehat{package} - \widehat{salaries}$$

Eventually, we have four dependent variables to try to value, and three of them are *not* censored. We can thus use these packages in the selection equation of a Heckman-inspired regression. Since the OP and SCP depend on the transfer fee, it might be better to look only at the SSP. This will be done in the next section.

#### 4.4 Estimation and results

Our hedonic model is entirely classic, as it rests on a log-linear equation where the value of the dependent variable is a function of the player's skills, personal characteristics, as well as control variables, *i. e.*:

$$\ln depvar = \sum_i^m \alpha_i \ln X_i + \sum_j^n \beta_j \ln Y_j + \sum_k^l \delta_k \ln Z_k + u_i$$

Where *depvar* (dependent variable) can be either the transfer fee or any of the packages explained in previous section and  $X_i$  is the players' skills vector,  $Y_j$  is the personal characteristics vector, and  $Z_k$  is the control variables vector (country, position, and yearly dummies).

We started estimating the dependent variables on the whole sample which includes all countries and positions aggregated, then we tried disaggregated regressions per continent, per position and performed a series of Chow tests to choose between the aggregate and the multi-level model. To achieve the proper segmentation, many estimations have been generated per country, per continent, per year, per position. While the whole process might look like pointless data mining, it appears *a posteriori* that the results of regressing the four dependent variables are mostly convergent; the main difference is the ability of a given specification to reduce heteroscedasticity in and across market segments. While providing all the Chow tests would be very fastidious, we only give an intuition of how segmentation works by providing some decisive examples of segmentation in Tables 4.3-4.7.

Regressing the transfer fee (as previous studies have done) brings fairly significant results (table 4.3): the transfer price is an increasing function of the duration of contract, of the internet visibility (google hits), of the player skills, of the buying club transfer activity and is negatively

affected by the end of contract ('free transfer') or the transfer being a loan. There are some consistency problems when the regression is broken down by continent or position: for instance, yearly dummies are significant at the world level but not at the continent level, while the Chow test shows that the disaggregated model is better. The Breusch-Pagan statistics ('Chi2') indicates heteroscedasticity, which cannot be reduced by disaggregation. While this does not make the model irrelevant, it means that the granularity of the data is not well rendered by the regression.

Using the packages to look at the breakdown by continent gives the same kind of results as with the transfer fee, while the Breusch-Pagan statistics is significantly lower to a point heteroskedasticity can disappear in some instances (Table 4.4). The yearly dummies are consistent between the global market and the European market but not with other continents: since these price dummies may be interpreted as a price index, it seems that the price of transfers is not evolving consistently across continents, hence breaking down the regression is required, as does the Chow test show.

Breaking down by positions (tables 4.3-4.7) is performed only with the subjective complete package: while the result with other packages is consistent, SCP almost kills heteroscedasticity. This indicates not only that the estimated coefficients are unbiased, but also that the segmentation may be relevant to the resolution of the data. Table 4.5 is devoted to the subset of forward players: while there is heteroscedasticity at 5% (but not at 1%), the Chow test indicates the aggregate model is better than a breakdown by continent. Unsurprisingly, the value of forward players' transfer is linked to their specific skills such as "*shooting*" and "*dribbling*" rather than other skills, which are more related to other positions. In addition to those skills variables, some other factors were significant like the age factor, the google scores

(“lgo0”), the involvement of the buying club in the transfer business was also noticeable through appropriate variables (“lco\_cuclu” and “ltot\_cuclu”), for selling clubs the strength of the causal relationship is less certain. In this section of regression, we notice that “lRemDurDays” which represents the remaining duration of contract before a transfer deal is reached is positively significant. This variable has not been tested in previous studies due to lack of data. The remarkable finding was about the loan’s variable (“tt\_loan”). While a Chow test does not favour disaggregation by continent, a separation between England and the rest of the world makes sense. In this case, the loans coefficient in England is highly significant and negative. That can be due to the intention of clubs to lend their players even at little or no cost to the Premier League clubs to put them on temporary display in order to sell them more easily.

The same feature is true for the defenders, and defensive midfielders (Table 4.6) the best segmentation (according to Chow tests) is England vs the rest of the world. Unsurprisingly, the defending skill (“ldefending”) is valued but while England values physicality (“lphysicality”), the rest of the world defenders’ markets rely on height. The English market is clearly more visible since google trends have a positive impact on player value. The other variables have the same impact as for the forward players. Defensive midfielders have been added to defenders as the result of another Chow test. For (non-defensive) midfielders, there is a higher heteroscedasticity than for other positions. Not only the English but the Italian market as well is singled out: there seem to be some singularities both in the appreciation of the players (height is preferred, youth is an asset) and of their situation (loans seem to be priced like regular transfers of shorter duration). The number of observations may be too small, and the significance of the regression is more in showing the difference with the rest of the world than in specifying a very precise model.



Dependent var = log (transfer fee)	World	Europe	L. America	USA+China
D. VARIABLES	lfee	Lfee	Lfee	lfee
<b>I. Variables</b>				
freetransfer	-7.364***	-12.43***	-2.377***	-1.362***
tt_loan	-4.977***	-9.497***	-1.728***	-0.841**
lage	81.73***	59.30***	78.34***	50.52*
lage2	-13.74***	-9.582***	-12.72***	-8.033**
ldur	1.412***	0.931***	1.623***	0.620***
lheight	3.439**	4.483***	1.142	0.979
lgt	0.0937***	0.0433	0.149***	0.129
lgo	0.0188	0.0438**	0.0222	-0.0301
footer_both	-0.396	0.748**	-0.424	-0.187
footer_left	0.302	-0.466	-0.22	-0.249
footer_right	0.503**	-0.402	0.133	0.0209
lpace	0.551*	0.733**	-0.0188	1.831**
lshooting	1.040***	0.720***	2.075***	0.551
ldribbling	0.467	0.466	-0.824	0.474
lpassing	2.085***	0.241	1.706**	0.76
ldefending	0.698***	0.523***	0.616	-0.199
lphysicality	3.749***	1.537***	1.238	4.167***
lco_preclu	-0.420***	-0.240***	0.0874	-0.292***
lco_cuclu	0.117***	0.411***	0.159*	0.00184
ltot_cuclu	0.172***	0.0957***	0.112***	0.0645***
ltot_preclu	0.167***	0.0651***	0.0723***	0.112***
pos_forward	1.181**	1.814*	1.014	-1.664
pos_mid	0.152	1.442	-0.0668	-2.209
pos_def	0.447	1.462	0.342	-2.706
nat_cont_asia	0.577**	0.214	-1.814	0.449
nat_cont_africa	0.273**	-0.0404	1.164	0.41
nat_cont_australia	0.338	-0.287		4.049***
nat_cont_europe	1.291***	0.0141	-0.106	0.841***
nat_cont_south_america	-0.0124	0.284**	0.564***	2.006***
y2008x	4.905			
y2009x	3.634**	0.889	-1.272	
y2010x	5.239***	1.871	0.267	
y2011x	5.874***	2.243	-0.08	0.0668
y2012x	5.318***	2.353	0.331	0.488
y2013x	5.622***	2.133	0.189	0.104
y2014x	5.555***	2.071	0.619	0.34
y2015x	5.637***	2.107	0.713	0.377
y2016x	5.783***	2.035	1.068	1.292
y2017x	6.143***	1.893	1.628	1.524
y2018x	5.728***	-0.0335	4.346	
Constant	-171.2***	-112.0***	-153.1***	-113.6***
<b>Chi2</b>	<b>1753.93</b>	<b>361.57</b>	<b>1521.37</b>	<b>848.45</b>
<b>Prob&gt;Chi2</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Observations</b>	<b>13,996</b>	<b>8,554</b>	<b>3,001</b>	<b>1,302</b>
<b>R-squared</b>	<b>0.542</b>	<b>0.77</b>	<b>0.291</b>	<b>0.258</b>
<b>Adj. R-Squared</b>	<b>0.541</b>	<b>0.7694</b>	<b>0.2822</b>	<b>0.2369</b>

Table 4.3: Regression with the transfer fee being the dependent variable

D. VARIABLES	World			Europe			L. Americas			USA+China		
	OP	SCP	SSP	OP	SCP	SSP	OP	SCP	SSP	OP	SCP	SSP
<b>I. Variables</b>												
freetransfer	-0.678***	-0.357***	-0.114***	-1.063***	-0.613***	-0.230***	-0.410***	-0.276***	-0.198**	0.256*	0.424***	0.509***
tt_loan	-0.442***	-0.172***	0.044	-0.717***	-0.332***	0.0215	-0.337***	-0.234***	-0.156*	0.235	0.372**	0.416***
lage	47.65***	-21.12***	-25.74***	42.17***	-23.17***	-27.69***	66.72***	-14.26*	-16.83**	18.43	-53.51***	-55.71***
lage2	-7.277***	2.356***	3.075***	-6.342***	2.760***	3.442***	-10.21***	1.233	1.638	-2.691	7.362***	7.692***
ldur	1.093***	0.214***	0.173***	1.054***	0.180***	0.157***	1.109***	0.264***	0.220***	1.187***	0.368***	0.331***
lheight	2.423***	2.577***	2.399***	1.576***	1.732***	1.527***	2.063*	2.398**	2.487**	-0.385	0.472	-0.0133
lgt	0.0549***	0.0583***	0.0551***	0.0151	0.0225*	0.0144	0.0697***	0.0640***	0.0639***	-0.00828	-0.0163	-0.0225
lgo	0.0198***	0.0207***	0.0180***	0.0310***	0.0326***	0.0255***	0.0084	0.00751	0.00882	-0.00724	-0.00942	-0.00273
footer_both	0.112	0.187**	0.188**	-0.051	0.0284	0.0165	1.409***	1.501***	1.536***	0.198	0.175	0.0549
footer_left	0.0133	-0.0415	-0.0461	0.0952	0.0299	0.0372	-0.428***	-0.513***	-0.515***	0.0325	0.0871	0.159
footer_right	0.0165	-0.0493	-0.0576	0.0795	0.00826	0.0166	-0.272*	-0.364**	-0.379***	0.0868	0.101	0.15
lpace	0.217**	0.240**	0.196**	0.0335	0.0403	0.0109	0.699***	0.779***	0.720***	0.187	0.27	0.0761
lshooting	0.492***	0.527***	0.459***	0.532***	0.559***	0.501***	0.434*	0.444*	0.356	-0.272	-0.217	-0.298
ldribbling	0.457***	0.400***	0.385***	0.261*	0.19	0.203	0.515*	0.564**	0.636**	0.883*	0.914**	0.592
lpassing	1.742***	1.709***	1.725***	1.568***	1.605***	1.657***	1.543***	1.504***	1.445***	1.865***	1.838***	1.904***
ldefending	0.366***	0.416***	0.387***	0.372***	0.432***	0.404***	0.355**	0.394***	0.379***	-0.11	-0.069	-0.119
lphysicality	1.287***	1.129***	1.013***	1.063***	0.932***	0.900***	1.401***	1.436***	1.393***	1.344***	1.201***	0.883**
lco_preclu	-0.0141	-0.00141	0.00425	0.0407**	0.0594***	0.0565***	-0.0151	-0.0228	-0.0264	0.034	0.0339	0.0441
lco_cuclu	0.0232*	0.0245*	0.0311**	0.0992***	0.0825***	0.0791***	0.0284	0.0444	0.0429	0.162***	0.158***	0.166***
ltot_cuclu	0.0701***	0.0630***	0.0573***	0.0732***	0.0676***	0.0654***	0.0471***	0.0430***	0.0407***	0.0223**	0.0200*	0.0113
ltot_preclu	0.0636***	0.0563***	0.0511***	0.0571***	0.0485***	0.0471***	0.0314***	0.0298***	0.0276***	0.0498***	0.0493***	0.0440***
pos_forward	0.503**	0.600***	0.580***	1.134***	1.223***	1.166***	0.591**	0.651**	0.644**	-0.615	-0.669	-0.288
pos_mid	-0.0252	0.0877	0.0989	0.625*	0.707**	0.670**	0.0711	0.154	0.179	-0.99	-1.001	-0.545
pos_def	0.251	0.354*	0.367*	0.861***	0.932***	0.903***	0.385	0.476	0.488*	-0.914	-0.906	-0.463
nat_cont_asia	0.240**	0.210**	0.216**	0.259**	0.241**	0.237**	-0.382	-0.366	-0.294	0.592**	0.519*	0.439*
nat_cont_africa	0.280***	0.256***	0.248***	0.244***	0.231***	0.218***	0.403	0.345	0.325	0.171	0.134	0.117
nat_cont_australia	-0.222	-0.231	-0.259*	-0.337*	-0.324*	-0.398**				0.132	0.0593	0.148
nat_cont_europe	0.262***	0.261***	0.234***	-0.118**	-0.0921**	-0.0865*	0.332***	0.365***	0.363***	0.358***	0.357***	0.372***
nat_cont_south_america	0.210***	0.227***	0.237***	0.255***	0.251***	0.241***	0.193**	0.197**	0.207***	0.225*	0.157	0.00125
y2013x				0.232***	0.211***	0.162**	0.118	0.271	0.376			
y2014x	-0.210***	-0.227***	-0.228***	0.064	0.0335	-0.00924	-0.583*	-0.519*	-0.422	0.137	0.161	0.17
y2015x	-0.0705	-0.113**	-0.110**	0.201***	0.140***	0.105*	-0.27	-0.22	-0.126	0.135	0.137	0.15
y2016x	-0.687***	-0.760***	-0.832***	-0.602***	-0.694***	-0.822***	-0.425	-0.408	-0.327	-0.26	-0.25	-0.358**
y2017x	0.137**	0.102*	0.0769	0.428***	0.391***	0.337***	-0.0658	-0.00481	0.0586	-0.268	-0.340*	-0.366**
y2018x	-0.032	-0.0704	-0.0753							-0.168	-0.128	0.0363
Constant	-93.88***	34.98***	43.79***	-82.85***	40.14***	48.11***	-125.4***	22.39*	27.33**	-42.91**	90.44***	97.89***

<b>Chi2</b>	<b>113.01</b>	<b>148.47</b>	<b>97.18</b>	<b>131.57</b>	<b>118.66</b>	<b>73.87</b>	<b>5.65</b>	<b>10.58</b>	<b>8.73</b>	<b>15.66</b>	<b>6.83</b>	<b>3.9</b>
<b>Prob&gt;Chi2</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0.0175</b>	<b>0.0011</b>	<b>0.0031</b>	<b>0.0001</b>	<b>0.009</b>	<b>0.0482</b>
<b>Observations</b>	<b>7,854</b>	<b>7,854</b>	<b>7,851</b>	<b>5,220</b>	<b>5,220</b>	<b>5,217</b>	<b>1,505</b>	<b>1,505</b>	<b>1,505</b>	<b>524</b>	<b>524</b>	<b>524</b>
<b>R-squared</b>	<b>0.711</b>	<b>0.57</b>	<b>0.549</b>	<b>0.771</b>	<b>0.592</b>	<b>0.561</b>	<b>0.605</b>	<b>0.519</b>	<b>0.509</b>	<b>0.699</b>	<b>0.484</b>	<b>0.511</b>
<b>Adj. R-Squared</b>	<b>0.7094</b>	<b>0.5677</b>	<b>0.5467</b>	<b>0.7698</b>	<b>0.5894</b>	<b>0.5581</b>	<b>0.5964</b>	<b>0.5078</b>	<b>0.4976</b>	<b>0.6785</b>	<b>0.4476</b>	<b>0.4771</b>

Table 4.4 : Global model + breakdown by continent for objective package / subjective complete package / subjective salary package

<b>Forward (Strikers)</b>			
	<b>World</b>	<b>World - England</b>	<b>England</b>
<b>D. VARIABLES</b>	<b>SCP</b>	<b>SCP</b>	<b>SCP</b>
<b>I. Variables</b>			
freetransfer	-0.594***	-0.487***	-0.925***
tt_loan	-0.476***	-0.379***	-0.747***
lagem	0.284**	0.24	0.465
lagep	0.352***	0.368***	0.271
lagem2	0.277***	0.318***	0.14
lagep2	-0.386***	-0.387***	-0.381***
ldur	0.0717**	0.116***	-0.048
lheight	0.292	1.578*	-1.671
lgt	0.0464**	0.0355	0.0636
lgo	0.0391***	0.0262*	0.0371*
footer_both	0.0994	0.319	-0.164
footer_left	-0.0822	-0.317	0.223
footer_right	-0.139	-0.323*	0.106
lpace	-0.215	-0.0272	-0.327
lshooting	1.321***	2.054***	0.124
ldribbling	2.824***	2.750***	2.743***
lpassing	-0.38	-0.465*	0.613
ldefending	0.215**	0.135	0.113
lphysicality	0.621***	0.105	1.849***
<b>IRemDurDays</b>	<b>0.0579***</b>	<b>0.0616***</b>	<b>0.0304</b>
hjlsl_follow	0.199***	0.194***	0.181***
lco_preclu	0.0547*	0.0664**	-0.0337
lco_cuclu	0.0504*	0.0594*	0.12
ltot_cuclu	0.0417***	0.0388***	0.0191
ltot_preclu	0.0389***	0.0307***	0.0559***
nat_cont_asia	0.0343	0.173	-0.411
nat_cont_africa	-0.0584	-0.117*	0.0298
nat_cont_australia	-0.13	-0.137	0.0294
nat_cont_europe	-0.00726	0.0383	-0.260*
nat_cont_south_america	-0.113*	-0.038	-0.325*
y2013x		-0.0258	0.22
y2014x	0.018	0.0837	0.00756
y2015x	0.0209	0.0748	-0.0511
y2016x	-0.754***	-0.701***	-0.799***
y2017x	0.0227	0.0429	0.153
y2018x	-0.0831		
Constant	-6.515***	-8.067***	-6.938*
<b>Chi2</b>	<b>6.61</b>	<b>3.77</b>	<b>0.89</b>
<b>Prob&gt;Chi2</b>	<b>0.0101</b>	<b>0.0521</b>	<b>0.3462</b>
<b>Observations</b>	<b>1,627</b>	<b>1,275</b>	<b>352</b>
<b>R-squared</b>	<b>0.675</b>	<b>0.673</b>	<b>0.723</b>
<b>Adj. R-Squared</b>	<b>0.6682</b>	<b>0.664</b>	<b>0.6928</b>
<b>SSE</b>	<b>1028.058</b>	<b>792.6396</b>	<b>189.1208</b>
<b>K</b>	<b>35</b>	<b>35</b>	<b>35</b>

Table 4.5: Breakdown by position and continent for the subjective complete package – forward players

Midfielders				
	W	W - (En+It)	En	It
D. VARIABLES	SCP	SCP	SCP	SCP
freetransfer	-0.513***	-0.366***	-0.959***	-0.658***
tt_loan	-0.327***	-0.209*	-1.058***	-0.275
lagem	0.132	0.216	0.595	-0.272
lagep	0.231*	0.266	0.0255	-0.418
lagem2	0.503***	0.415***	0.221	0.860***
lagep2	-0.347***	-0.380***	-0.164	-0.00698
ldur	0.138***	0.146**	-0.0215	0.103
lheight	0.212	0.107	-2.351	4.634**
lgt	-0.0101	0.0333	-0.00105	-0.0842
lgo	0.00811	0.019	0.000377	0.00118
footer_both	-0.116	0.157	-0.00443	-0.312
footer_left	0.287	0.154	0.48	-0.00998
footer_right	0.194	-0.00733	0.426	
lpace	-0.146	0.102	1.643**	-1.128*
lshooting	0.0739	-0.504	-0.199	1.261*
ldribbling	1.847***	1.890***	1.459	2.017**
lpassing	2.731***	2.666***	4.684***	2.556*
ldefending	0.00341	-0.201	0.133	0.246
lphysicality	1.098***	1.648***	0.915	0.488
lRemDurDays	0.0248	0.0401	-0.0332	0.0391
lsf_follow	0.191***	0.238***	0.120**	0.0194
lco_preclu	-0.0112	0.0206	-0.0593	0.132*
lco_cuclu	0.0557	0.0656	0.064	0.361***
ltot_cuclu	0.0203**	0.0113	0.0177	0.00152
ltot_preclu	0.0286***	0.0164	0.0536**	-0.00115
nat_cont_asia	-0.307	-0.229	-1.667*	-0.99
nat_cont_africa	0.112	0.146	0.125	-0.0744
nat_cont_australia	-0.703*	-0.446		-0.51
nat_cont_europe	-0.0726	0.0386	-0.35	-0.400**
nat_cont_south_america	0.0254	0.0494	-0.112	0.0372
y2013x				
y2014x	-0.360***	-0.397**	-0.149	-0.208
y2015x	-0.516***	-0.529***	-0.341	-0.442
y2016x	-1.268***	-1.405***	-0.883***	-1.016***
y2017x	-0.351***	-0.667***	0.457*	0.179
y2018x	-0.468***	-0.583***	-0.681*	-0.111
Constant	-10.57***	-10.87***	-20.35***	-12.16**
<b>Chi2</b>	<b>26.41</b>	<b>6.67</b>	<b>1.75</b>	<b>8.18</b>
<b>Prob&gt;Chi2</b>	<b>0</b>	<b>0.0098</b>	<b>0.1864</b>	<b>0.0042</b>
<b>Observations</b>	<b>1,227</b>	<b>674</b>	<b>227</b>	<b>326</b>
<b>R-squared</b>	<b>0.642</b>	<b>0.679</b>	<b>0.736</b>	<b>0.675</b>
<b>Adj. R-Squared</b>	<b>0.6318</b>	<b>0.6616</b>	<b>0.6891</b>	<b>0.6373</b>

Table 4.6: Breakdown by position and continent for the subjective complete package - midfielders

<b>Defenders + Defensive Midfielders</b>			
<b>VARIABLES</b>	<b>World SCP</b>	<b>W - En SCP</b>	<b>En SCP</b>
<b>I. Variables</b>			
freetransfer	-0.412***	-0.324***	-0.740***
tt_loan	-0.221***	-0.126	-0.690***
lagem	0.0531	0.103	-0.0902
lagep	0.191	0.147	0.423
lagem2	0.434***	0.419***	0.567***
lagep2	-0.317***	-0.284***	-0.455***
ldur	0.190***	0.221***	0.0282
lheight	2.474***	3.071***	-0.921
lgt	0.00105	-0.016	0.132**
lgo0	0.0288**	0.0308**	0.0127
footer_both	1.530***	1.559***	-0.531
footer_left	-1.369***	-1.260***	-0.0277
footer_right	-1.420***	-1.319***	
lpace	-0.192	-0.107	-0.178
lshooting	0.250*	0.304*	-0.039
ldribbling	0.096	0.198	-0.152
lpassing	0.601**	0.43	0.644
ldefending	1.983***	1.918***	3.361***
lphysicality	0.927***	0.553	2.040***
<b>IRemDurDays</b>	<b>0.0506**</b>	<b>0.0544**</b>	<b>0.00557</b>
lsf_follow	0.223***	0.228***	0.133***
lco_preclu	0.0523*	0.0694**	0.104
lco_cuclu	0.0729**	0.0730**	0.203*
ltot_cuclu	0.0404***	0.0416***	0.0136
ltot_preclu	0.0322***	0.0199*	0.0350**
pos_mid	0.0746	0.128*	-0.0744
pos_def			
nat_cont_asia	0.359*	0.390**	
nat_cont_africa	0.225***	0.232***	0.261*
nat_cont_australia	-0.121	-0.0992	0.0557
nat_cont_europe	0.0571	0.052	0.253
nat_cont_south_america	0.208***	0.201***	0.461**
y2013x	0.588***	0.320**	
y2014x	0.147		-0.690***
y2015x	0.0886	-0.159	-0.590***
y2016x	-0.668***	-0.902***	-1.295***
y2017x	0.139*	-0.173*	-0.0367
y2018x		-0.245**	-0.261
Constant	-4.406**	-3.219	-10.79***
<b>Chi2</b>	<b>10.27</b>	<b>4.18</b>	<b>3.03</b>
<b>Prob&gt;Chi2</b>	<b>0.0014</b>	<b>0.0409</b>	<b>0.082</b>
<b>Observations</b>	<b>1,525</b>	<b>1,209</b>	<b>316</b>
<b>R-squared</b>	<b>0.659</b>	<b>0.639</b>	<b>0.783</b>
<b>Adj R-Squared</b>	<b>0.6505</b>	<b>0.6281</b>	<b>0.7562</b>
<b>SSE</b>	<b>1048.553</b>	<b>836.641</b>	<b>147.7246</b>
<b>K</b>	<b>36</b>	<b>36</b>	<b>34</b>

Table 4.7: breakdown by position and continent for the subjective complete package - defenders

## 4.5 Conclusion

This chapter did not intend to provide the ultimate transfer price equation, but to prove the workability of a concept: it is possible to model transfer prices on the global market using (1) A pseudo-random sample and (2) Non-censored measures of transfer cost. The first property was achieved by starting from a large pre-existing set of players (the Futhead database) and appending data gathered across the internet. The second property draws upon variable of interest in the transfer price negotiation, including both transfer money and future salaries (the latter being always  $>0$  unless the player quits playing professional football). The main finding of the chapter is that the global transfer market is segmented across positions and geographic entities.

For both attackers and defenders, the English Premier League (EPL) plays a special role since clubs happen to lend players to EPL clubs for free as if they were putting those players on display to sell them at a price boosted by popularity. The case of midfielders is trickier, and the resolution of our data is probably insufficient to analyse in detail the complex interaction between semi-recessed leagues and the global market. Nevertheless, by considering the “total cost of ownership” of players rather than the straight transfer fee, we achieved the reduction of heteroscedasticity and have thus shown that market segments were consistent not only from the Chow test but also from the variance/residuals. These findings raise many questions but it seems obvious that to answer them, we must significantly expand the database to achieve better granularity about the details.

## CHAPTER FIVE

### 5 PRICING FOOTBALL TRANSFERS 2007-2018: FROM ECONOMETRICS TO MACHINE LEARNING

#### 5.1 Introduction

The financial evaluation of football players transfers, and salaries has changed over time as well as its methods of evaluation. It is enough to compare the transfer price records between the Brazilian star Neymar when moving from FC Barcelona to the French club Paris Saint Germain in summer 2017 for €220 million to that of the Argentine legend Diego Armando Maradona when moving from FC Barcelona to the Italian club SSC Napoli in July 1984 for \$10.48M, which is equivalent to \$24.57M (€22.7M) in 2017 value. To notice the extreme pricing differences between old and new financial valuations of superstars. Frick (2007) assumes that the increasing rates of transfer fees are affecting players' wages, which have been increasing consistently, thanks to the generous data resources that have exposed players skills and helped in formulating different pricing schemes in recent times.

Name	Club	Season	Wage	HITS	Age	Potential	Overall
L. Messi	FC Barcelona	2004-2021	€560K	904	32	94	94
E. Hazard	Real Madrid	2019-2024	€470K	380	28	91	91
C. Ronaldo	Juventus	2018-2022	€410K	753	34	93	93
K. De Bruyne	Manchester City	2015-2023	€370K	480	28	91	91
A. Griezmann	FC Barcelona	2019-2024	€370K	402	28	89	89
L. Suárez	FC Barcelona	2014-2021	€350K	315	32	89	89
T. Kroos	Real Madrid	2014-2023	€340K	303	29	89	89
K. Benzema	Real Madrid	2009-2022	€340K	248	31	88	88
L. Modrić	Real Madrid	2012-2020	€320K	363	33	89	89
S. Agüero	Manchester City	2011-2021	€310K	341	31	90	90
R.Lewandowski	FC Bayern München	2014-2023	€300K	373	30	91	91
Sergio Ramos	Real Madrid	2005-2021	€300K	329	33	89	89
Neymar Jr	PSG	2017-2022	€290K	803	27	92	92
Piqué	FC Barcelona	2008-2022	€280K	178	32	88	88
S. Busquets	FC Barcelona	2008-2023	€280K	182	30	88	88
Casemiro	Real Madrid	2013-2023	€280K	300	27	89	88
R. Sterling	Manchester City	2015-2023	€250K	475	24	90	88

Table 5.1: Players' attributes and salaries. Source: SOFIFA.



In this chapter, we maintain our search in the same direction of finding the best possible pricing model. To diversify our approaches, machine learning techniques were selected to analyze the large data set, which includes both transfer fees and salaries. Three different algorithms were applied on both transfer fees and salaries independently. By evaluating three different techniques for generating pricing models, we could distinguish which factors are more important. The data set contains a lot of players performance attributes in addition to clubs' features. Thus, it helps to compare better and contrast. The techniques offer the study the privilege to compare findings of different methodologies and select the best.

The rest of the chapter is organized as follows. Section 5.2 discusses the data, while Section 5.3 discusses the methodology used for the estimation. This is followed by a discussion of the results (Section 5.4), and Section 5.5 presents some concluding remarks.

## **5.2 The Data**

### **5.2.1 Data structure**

To achieve a selection-free sample, we tried to collect the largest possible set of players. The basis of our data collection was the *sofifa.com* website: it features more than 200,000 observations of 30,642 unique players across 35 leagues ranging from 2013 to 2019. The site records every significant change in the variables it stores, which includes player's basic data (name, age, height, weight, club, position) as well as expert opinion on his 33 skills<sup>5</sup>, and, very importantly, the player weekly salary and end of the contract date. While this information may be inexact in detail, it is still a precious instrument.

---

<sup>5</sup> Namely: crossing, finishing, headingaccuracy, shortpassing, volleys, dribbling, curve, fkaccuracy, longpassing, ballcontrol, acceleration, sprintspeed, agility, reactions, balance, shotpower, jumping, stamina, strength, longshots, aggression, interceptions, positioning, vision, penalties, marking, standingtackle, slidingtackle, gkdiving, gkhandling, gkkicking, gkpositioning, gkreflexes.

Since every change leads to a record, there is a high number of duplicates for players going on loan for a short time, or young emerging players with evolving skills, while steady players have gaps in their history. But before we could clean up the database and fill the panel, we had to append information on transfers taken mostly from *transfermarkt.de* and additional information on internet activity (such as followers or likes on fansites) and, most importantly, the end of the contract. This information has not been used systematically in the previous studies, leading to an obvious omitted variable bias. Only Ezzeddine and Pradier (2019) made limited use of it, i.e. with few observations because of the difficulty to gather consistent information. After cleaning up the duplicates and carrying forward the data to fill the gaps, we ended up with an (unbalanced) panel structure with no less than 176,248 unique observations of which 26,709 (15%) featured a transfer fee. Some descriptive statistics are presented table 5.2:

<b>country</b>	<b>Number of observations</b>	<b>country</b>	<b>Number of observations</b>
Argentina	8045	Korea	3318
Australia	2929	Mexico	6139
Austria	4143	Netherlands	5895
Belgium	5408	Norway	4817
Brazil	5105	Poland	5719
Chile	4833	Portugal	6080
China PR	3685	R. Ireland	2940
Colombia	7012	Russia	781
Croatia	301	S. Arabia	6211
Czech Rep.	938	Scotland	4725
Denmark	4152	S. Africa	184
England	36648	Spain	15413
Finland	370	Sweden	4874
France	14437	Switzerland	3677
Germany	20273	Turkey	6665
Greece	784	Ukraine	669
Italy	16359	USA	6511
Japan	5736		

Table 5.2: Players per country in the database

## 5.2.2 Descriptive Statistics

After analysing the data, the following statistics in Table 5.3 describe the nature of the data and the variables utilised in the study. Further explanation, some of the significant effective independent variables are displayed in Table 5.4 and Figure 5.1.

Independent Variables (Columns)	84
Discrete Variables (Columns)	24
Continuous Variables (Columns)	60
Original Dataset players (entries)	200,000+ entries
Dataset utilized	25,000+ entries

Table 5.3: Observations' and variables' breakdown

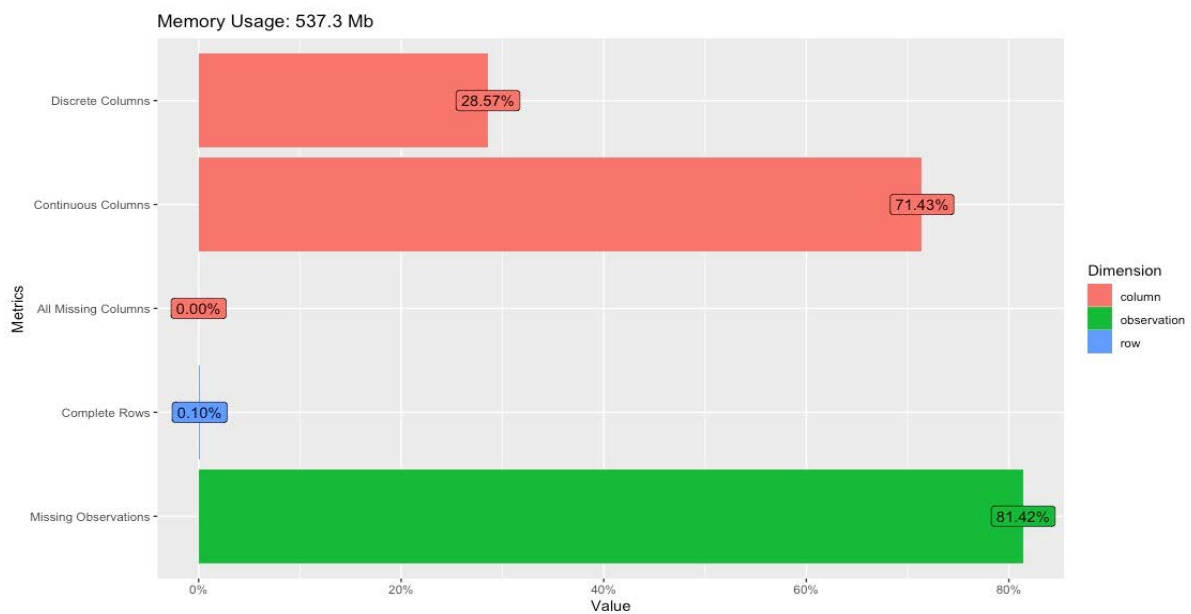


Figure 5.1: Discrete vs Continuous data

	Age	height (cm)	weight (kg)	co_cuclu	tot_preclu	tot_cuclu	Potential	Dribbling	Ballcontrol
Min.	16	152	52	50	0	0	50	5	5
1st Q	21	177	71	635	7.20E+05	5.94E+06	67	52	57
Median	24	182	75	724	1.55E+09	5.49E+09	71	63	65
Mean	24	181.2	75.65	928.6	3.20E+10	6.13E+10	71	58.58	61.77
3rd Q	28	186	80	878	1.32E+10	3.05E+10	75	71	72
Max.	41	203	107	4661	1.14E+12	1.54E+12	95	96	95

Table 5.4: Descriptive statistics for some effective variables

We noticed that most variables do contain a high percentage of missing information. As such, we filter only observations with transfer fees information. Thus 25,220 observations with no missing information persisted for our study. In preparing for modelling, we split an existing data set according to ratios (0.6, 0.2, 0.2) corresponding to training, validation and test datasets. In modelling wages, we omit the “transfer fees” variable. Likewise, we omit wages when modelling the transfer fee. We remove all duplicates of columns (like "playername", "tm\_playername", "familyname", "tm\_familyname"), dates (like "dateofbirth", "current\_date"), and other variables like “so\_fi\_fa\_player\_value\_in\_k”, “tm\_marketvalue”, “sofifa\_release\_clause\_in\_k”. Thus, we use 70 features as the number of predictor variables in modelling processing.

### 5.3 Methodology

Many machine learning techniques can be used in predicting an asset price. Knowing that football players are fundamental assets for football clubs, this paper presents three different models to approximate wages and transfer fees of football players. Generalised Linear Models (GLM), Random Forests (RF), and Gradient Boosting Machines (GBM) are the models that were used to estimate the predictors of the pricing function of football players.

### 5.3.1 Generalised Linear Models (GLM)

It is a flexible generalised linear regression model composed of a set of predictors aligned to estimate a continuous variable, i.e. the transfer fee or the wage. In general, the model aggregates predictors and error terms or residuals. The presence of the errors could be due to high variance (when independent variables are correlated or abundant), selection bias, or other unknown reasons. In order to avoid typical linear regression flaws like multicollinearity, endogeneity, and overfitting, a regularisation approach is used by this GLM model. Thus, an Elastic Net regression which is a mixture of Lasso and Ridge regressions is applied. Recalling Addo et al. (2018) and Friedman et al. (2010) we represent Elastic Net function as follows. If  $y_i$  represents the predicted values and  $x_i$  the observed ones for  $i = \{1, 2, \dots, n\}$ ,  $n$  being the number of variables, Lasso regression is then characterised by the following:

$$SSE_{lasso} = \sum_1^n (x_i - y_i)^2 + \lambda \sum |\beta| \quad (1)$$

Second, Ridge regression is characterised by the following:

$$SSE_{Ridge} = \sum_1^n (x_i - y_i)^2 + \lambda \sum_1^n |\beta|^2 \quad (2)$$

Finally, the Elastic Net function is represented by the following aggregation:

$$SSE_{EN} = \sum_1^n (x_i - y_i)^2 + \lambda [(1 - \alpha) \sum_1^n |\beta|^2 + \alpha \sum |\beta|] \quad (3)$$

Where Elastic Net penalty is simply determined by the value of  $\alpha$  such that:

$$P_\alpha(\beta) = (1 - \alpha) \sum |\beta|^2 + \alpha \sum |\beta| \quad (4)$$

Lasso regression is used to shrink coefficients of some variables to zero, while Ridge regression shrinks the coefficients to non-zero values in order to prevent overfitting. In this model, a perfect mixture of Lasso and Ridge ( $\alpha = 0.5$ ) was selected for evaluation. The results generated (Table 5.7) shows that the variables respond well to the GLM model (Fig. 5.4) when we are estimating the wages of the football players, but not for transfer fees (Table 5.5).

### 5.3.2 Random Forest (RF)

In general, RF is a forest composed of multiple decision trees. It is an ensemble algorithm that utilizes an averaging/bagging technique to curb the variance and reduce overfitting sometimes produced by individual decision trees. By aggregating an increasing number of decision trees, RF provides stable prediction outcomes of higher accuracy by averaging the outcomes of all decision trees. This procedure is usually referred to as Bootstrap Aggregation, also called Bagging. The main advantages of this algorithm are that generalization error can be minimised by avoiding overfitting. Such procedure was proposed by Breiman (2000, 2004) to form an ensemble model of decision trees, with controlled variance, from a randomly-selected subset of features. Therefore, predicting the value of football players' wages and transfer fees with low variance and high accuracy was the motive behind using the RF model.

In the training algorithm, the training dataset is  $X_i = x_1 + x_2 + \dots + x_n$  with outcomes  $Y_i = y_1 + y_2 + \dots + y_n$ . The number of samples / trees is  $t = 1 + 2 + \dots + T$

After training on  $t$  samples, predictions on untrained samples  $x'$  by generating the average of all individual regression trees is given by the following equation:

$$\hat{f} = \frac{1}{T} \sum_1^T f'_t(x')$$

This bagging technique is intended to reduce the variance in the model. Unlike the GLM methodology, it controls the variance at no bias cost.

### 5.3.3 Gradient Boosting Machines (GBM)

This is an ensemble machine learning technique that uses multiple models to improve results and reduce errors trying to generate better predictions. GBM regression uses an ensemble of weak prediction models, decision trees, by adding them sequentially to correct the errors in the overall prediction model. Adding all trees, weak and strong predictors, iteratively together in

one single strong unit aims at controlling the error level and improving the results. This methodology differs from the previous RF one by adding sequentially the learners, decision trees, instead of bagging them (Addo et al., 2018; Friedman, 2001).

Therefore, the aim is to train the model  $f$  to predict  $\hat{y} = f(x)$  by minimizing the mean square error (MSE). Recalling from the previous approach, the input variables represented by  $X$  and the output variables represented by  $Y$  from the training set we attempt to reduce the loss function  $L(Y, f(x))$  which can be squared error  $\frac{1}{N} \sum_1^T (\hat{y}_i - y_i)^2$  or an absolute error  $|\hat{y}_i - y_i|$  if  $Y \in \mathbb{R}$ .

## 5.4 Results

Three methodologies were used to estimate football players' transfer fees and salaries: GLM, RF, and GBM. Due to missing values in some variables used to estimate the transfer fees, the sample of players that were used in the transfer fees' estimation was reduced to a smaller sample than the one which was used in the wages' estimation. But in each case (fees and/or wages), samples were divided into three parts: training phase (using 60% of the dataset), validation phase (20%), and testing phase (20%). Therefore, two batches of tests and results were carried and generated as follows:

### 5.4.1 Transfer Fees

The three models generated different results. For instance, the GLM model did not respond well to the predictors in this model as the  $R^2$  value was extremely low (0.000000000686) (Table 5.5). We could confirm that it is not a good model for transfer fees through validation and test data as well were low levels of  $R^2$  persisted. Eventually, the levels of the rooted-mean-squared-

error (RMSE) and the mean absolute error (MAE) had high values. Looking at the level of importance of the variables, very few showed up to the level (only two variables) (Figure 5.2) and (Table 5.6). This leads us to the next machine learning methodology, RF.

Using the random forest methodology (RF), the model seems to respond well by selecting the significant variables according to their importance level. Looking at the most important variables selected by the model (Figure 5.3(a)), we notice that effective variables include the purchasing club's characteristics (*sofifa\_currentclub*, *tot\_cuclu*), which plays an important role in this model. In addition, players' skills (potential, reactions, ball control, and dribbling) in addition to players' nationality and position are among the top variables affecting the transfer fee pricing of this RF model. Remarkably, the social media variables (*sofifa\_likes*, *sofifa\_dislikes*, and *sofifa\_follows*) are highly important in this model. With a good  $R^2$  of 0.73 in the training phase and 0.82 in the validation and the testing phases, it is confirmed that the model performs very good and selects significantly and correctly its variables.

Furthermore, estimation using the GBM model generates equally convincing results as the RF model. As the results in Table 5.5 are showing, the training phase generates very high  $R^2$  of 0.94 which, for instance, looks to be high before it's adjusted to a value of 0.78 and 0.8 in the validation and testing phases respectively. Looking at the most effective variables (Figure 5.3 (b)), no big difference from the previous model was revealed. It is noteworthy to mention that the remaining duration of the contract (*tm\_remaining\_dur\_cont\_days*) lies among the important factors affecting the pricing function of this model in addition to the players' and clubs' characteristics mentioned in the previous model.



Finally, the three models compete among each other to generate the most accurate predictive model with the least variation possible. For that, the GLM model is excluded firstly from the competition. Among the remaining two models: RF and GBM, the lower the RMSE and the MAE, the better is the model. Looking at those figures (Table 5.5), the random forest approach provides a slightly lower error level and a better R<sup>2</sup>. Therefore, we can conclude that transfer fees are better modelled using the RF approach.

<b>Transfer Fees - Test Results - ML Models</b>									
	<b>GLM</b>			<b>RF</b>			<b>GBM</b>		
	Training	Validation	Test	Training	Validation	Test	Training	Validation	Test
<b>MSE</b>	3.34E+17	2.62E+17		8.91E+16	4.79E+16		2.00E+16	5.76E+16	
<b>RMSE</b>	5.78E+08	5.12E+08	6.32E+08	2.99E+08	2.19E+08	2.7E+08	1.41E+08	2.4E+08	2.77E+08
<b>MAE</b>	2.19E+08	2.1E+08	2.29E+08	68549202	61431603	67640627	33776526	70219606	71704650
<b>R<sup>2</sup></b>	<b>6.86E-10</b>	<b>-0.0002</b>	<b>-0.00033</b>	<b>0.733536</b>	<b>0.817407</b>	<b>0.816769</b>	<b>0.940213</b>	<b>0.780668</b>	<b>0.807629</b>

Table 5.5: Three Models test results – Transfer Fees.

<b>Top 30 Important Variables' per Model -Transfer Fees</b>		
<b>GLM</b>	<b>RF</b>	<b>GBM</b>
sofifa_likes	sofifa_dislikes	sofifa_dislikes
sofifa_dislikes	sofifa_currentclub	tm_nationalitycurrent
tm_previousclub.1.fc köln	tm_nationalitycurrent	sofifa_currentclub
tm_previousclub.1.fc heidenheim	Potential	sofifa_likes
tm_previousclub.1.fc k'lautern	sofifa_likes	tm_durationofcontract
tm_previousclub.1.fc köln	Reactions	Dribbling
tm_previousclub.1.fc magdeburg	tm_durationofcontract	Potential
tm_previousclub.1.fc nuremberg	tm_remaining_dur_cont_days	Reactions
tm_previousclub.1.fsv mainz	tot_cuclu	tm_remaining_dur_cont_days
tm_previousclub.12 de octubre	Ballcontrol	tot_cuclu
tm_previousclub.1860 munich	player_nationality	sofifa_player_position_2
tm_previousclub.NA	tot_preclu	sofifa_player_position
tm_previousclub.a.quilmes	sofifa_follows	player_nationality

tm_previousclub.a.bucaramanga	sofifa_player_position_2	Ballcontrol
tm_previousclub.aalborg bk	sofifa_player_position	Finishing
tm_previousclub.aalesund	Dribbling	Positioning
tm_previousclub.aarhus gf	co_cuclu	tm_player_position
tm_previousclub.abano	Fkaccuracy	sofifa_countryofcurrentclub
tm_previousclub.abc fc	tm_player_position	tm_countryof_previousclub
tm_previousclub.aberdeen fc	sofifa_countryofcurrentclub	tot_preclu
tm_previousclub.ac ajaccio	sofifa_player_position_3	sofifa_player_position_3
tm_previousclub.ac arles	Sprintspeed	Sprintspeed
tm_previousclub.ac barnechea	Composure	Vision
tm_previousclub.ac horsens	tm_countryof_previousclub	Headingaccuracy
tm_previousclub.ac le havre	virtual_rem_dur_days	Composure
tm_previousclub.ac milan	Positioning	instagram_follows
tm_previousclub.ac pisa	tm_endofcontract	Longshots
tm_previousclub.acassuso cf	Vision	Volleys
tm_previousclub.acrington	Finishing	Mercato
tm_previousclub.acrmessina	Jumping	Shortpassing

Table 5.6: Variables' Importance (Descending Order) per three Models – Transfer Fees.

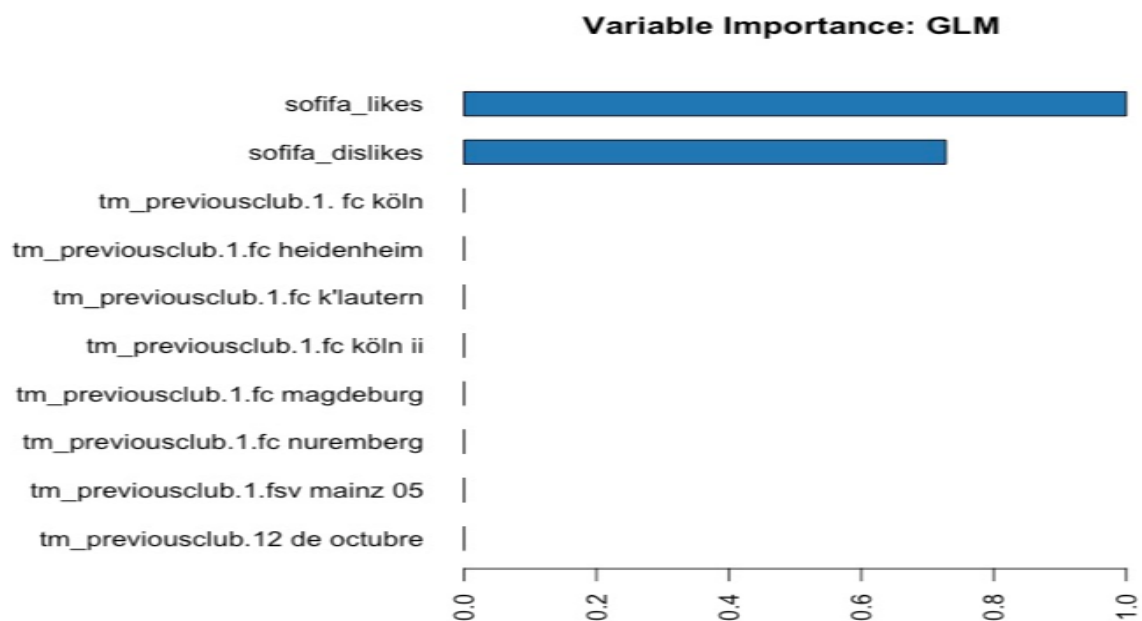
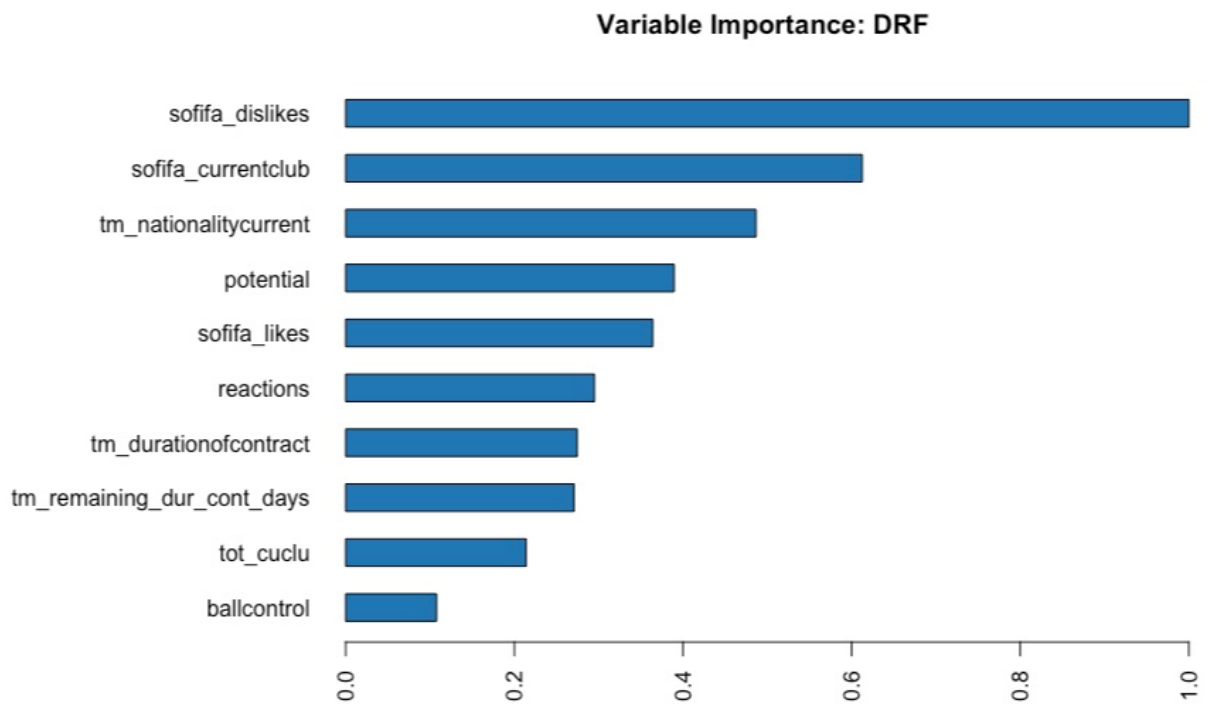
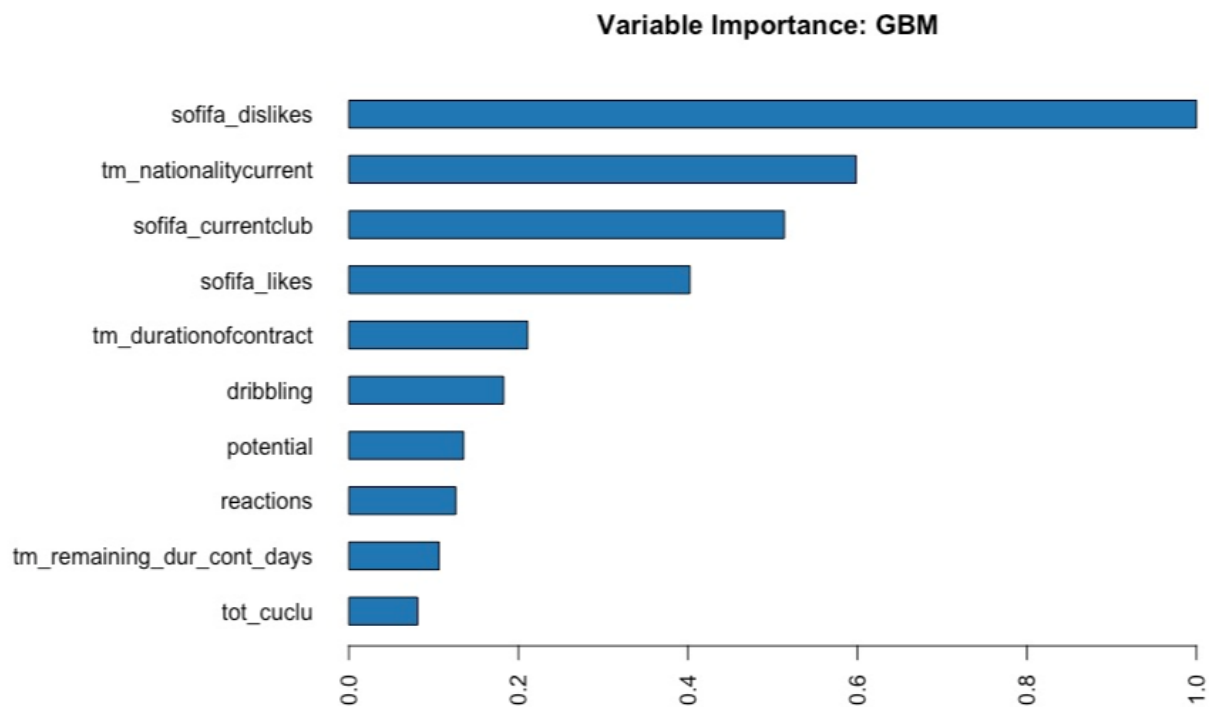


Figure 5.2: GLM model – Most important variables affecting Transfer fee



*Figure 5.3(a): RF model – Most important variables affecting Transfer fee*



*Figure 5.4 (b): GBM model – Most important variables affecting Transfer fee*

### 5.4.2 Wages

In this part, the same three approaches were used again to predict the wages of football players but using a larger data set than the one used to estimate transfer fees. A quick look at the results (Table 5.7), leads us to the following conclusions:

First, the GLM model, which is a linear model introduced with regularization additions to control variation and avoid overfitting to estimate the players' wages, has responded differently than when applied to estimate the transfer fees. In the training phase (Table 5.7), the  $R^2$  was 0.53, which shows that the variables selected by the model were good. This value remains stable in the validation and testing phases. The variables that had more importance in this model (Table 5.7) were players' performance skills (*reactions*, *potential*, *headingaccuracy*, and *volleys*), players' personal characteristics (*age* and *weight\_kg*), purchasing club's characteristics (*sofifa\_currentclub*, *co\_cuclu*, *tot\_cuclu*), selling club's characteristics (*tot\_preclu*), and the social media variables (*sofifa\_likes*, *sofifa\_dislikes*, and *sofifa\_follows*). (Figure 5.4).

Secondly, when we approach the wages through the RF methodology, the model selects its variables with its best possible accuracy degree. In the training phase, the  $R^2$  was almost 0.8 and improved further in the validation and the testing phases respectively arriving at 0.82 at the later phase (Table 5.7). Looking at the top important variables in this model, some similarities with the previous GLM model prevail especially in some players' skills (*reactions*, *potential*, *headingaccuracy*) with some additional skills like *dribbling* and *finishing*. Moreover, personal characteristics (*age* and *nationality*) have an important effect on the wages of the players, proven by this model. The top 30 variables are displayed in importance order from the most to the least important ones in (Table 5.7). Remarkably, the purchasing or the current club

of the player determined by *sofifa\_currentclub* variable is the most important factor suggested by this model (Figure 5.5).

Another model was experimented in the race to estimate the wages of football, the GBM model. Results were similar to a larger extent with the RF model in terms of characteristics importance and high  $R^2$  level (Table 5.7). For example, in the training phase, the  $R^2$  was 0.97, which seems to be very good on one side but may raise certain doubts about overfitting possibilities. The fears are relieved when validation and testing phases generate stable figures of 0.975 and 0.815, respectively. Players' *reactions* and *ballcontrol* are chosen by this model to have the greatest importance in the prediction process (Table 5.8). Players' nationality (*tm\_nationalitycurrent*) has the same importance as in the previous model with the players' current club (*sofifa\_currentclub*) being among the top three important factors (Figure 5.6).

Finally, choosing the best model depends on the lowest variance and error. For that, the RF model seems to provide the best  $R^2$  with the least RMSE and MAE (Table 5.7). Thus, it stands to be the best machine learning technique utilised in estimating the wages of football players.

	<i>Wages - Test Results - ML Models</i>								
	GLM			RF			GBM		
	Training	Validation	Test	Training	Validation	Test	Training	Validation	Test
<b>MSE</b>	280.4548	275.1813		129.7613	116.2489		19.27647	121.0588	
<b>RMSE</b>	16.74678	16.58859	16.5565	11.39128	10.78188	10.27117	4.390498	11.00267	10.65392
<b>MAE</b>	10.7473	10.66691	10.65374	5.987552	5.75794	5.53866	2.420433	5.883503	5.574294
<b>R<sup>2</sup></b>	<b>0.564612</b>	<b>0.53453</b>	<b>0.555494</b>	<b>0.798554</b>	<b>0.803365</b>	<b>0.828927</b>	<b>0.970075</b>	<b>0.795229</b>	<b>0.81594</b>

Table 5.7: Three Models test results - Wages.

<i>Top 30 Important Variables per Model - Wages</i>		
GLM	RF	GBM
reactions	sofifa_currentclub	reactions
sofifa_dislikes	Reactions	ballcontrol
tot_cuclu	Ballcontrol	sofifa_currentclub
co_cuclu	tm_nationalitycurrent	tm_nationalitycurrent
potential	sofifa_likes	tot_cuclu

sofifa_follows	tot_cuclu	mercato
sofifa_likes	tot_preclu	sofifa_player_position_2
tot_preclu	Mercato	age
tm_remaining_dur_cont_days	sofifa_dislikes	player_nationality
headingaccuracy	Dribbling	sofifa_likes
gkicking	Positioning	sofifa_dislikes
age	Age	dribbling
volleys	standingtackle	standingtackle
gkdiving	player_nationality	finishing
facebook_follows	sofifa_player_position_2	marking
facebook_likes	Slidingtackle	composure
weight_kg	Finishing	tm_countryof_previousclub
tm_player_position.defender.centre-back	co_cuclu	slidingtackle
twitter_follows	Shotpower	co_cuclu
gkreflexes	tm_countryof_previousclub	tm_player_position
sofifa_countryofcurrentclub.England	Composure	positioning
gkhandling	headingaccuracy	sofifa_player_position
tm_division2	Shortpassing	vision
co_preclu	Longshots	headingaccuracy
instagram_follows	sofifa_follows	tot_preclu
virtual_rem_dur_days	tm_player_position	sofifa_player_position_3
sofifa_player_position.CB	sofifa_player_position_3	gkreflexes
sofifa_endofcontract	Marking	interceptions
sprintspeed	sofifa_player_position	tm_endofcontract
vision	Interceptions	gkdiving

Table 5.8: Variables' Importance per three Models – Wages

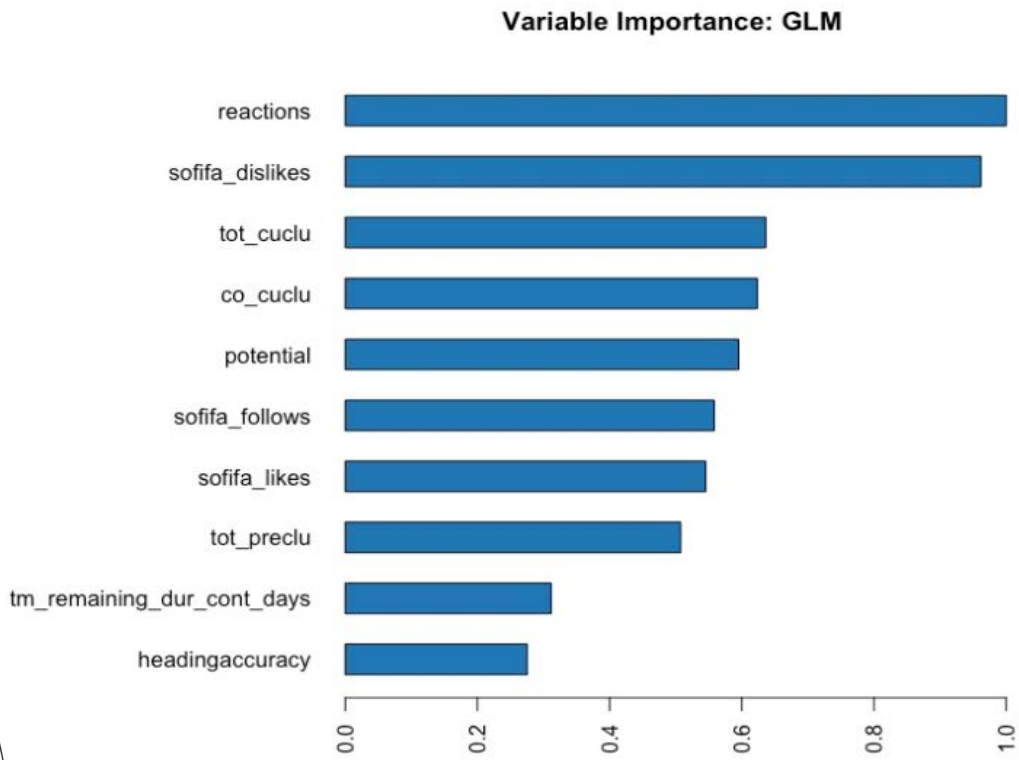


Figure 5.5: GLM model – Most important variables affecting Wages

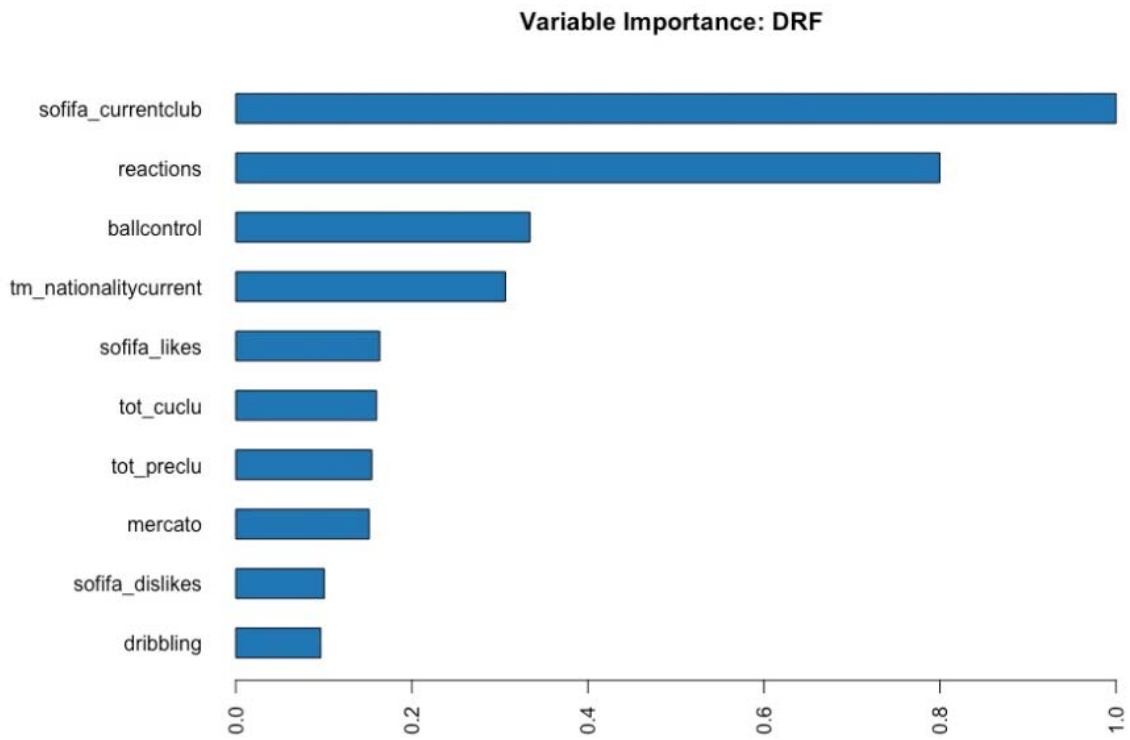


Figure 5.6: RF model – Most important variables affecting Wages

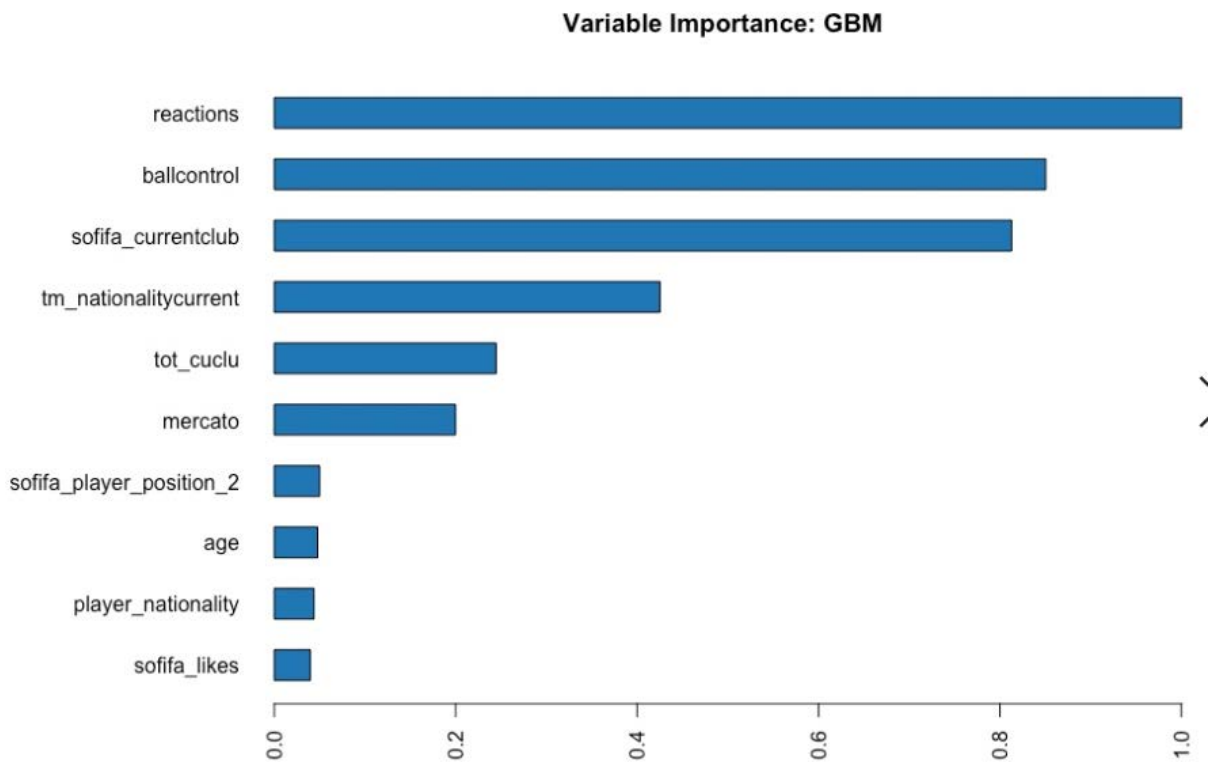


Figure 5.7: GBM model – Most important variables affecting Wages

## 5.5 Conclusion

Approaching football players' pricing estimations through non-traditional approaches like machine learning is a new challenge in recent times. The three models used in this chapter: GLM, RF, and GBM have generated promising results in this context. Previous papers challenges of heteroscedasticity and selection bias are automatically taken care of by the current new modelling schemes despite data challenges. This exercise can stand as a foundation for more complex subsequent estimations if better data sets are ensured.

Moreover, it is crucial to consider different models in the search for the best estimation. The results generated in the transfer fees estimation gave us a remarkable indication to drop the linear model approach, while approaching the matter through random forest technique generated the best estimation with a minimized number of residuals (RMSE and MAE), thus



lowest variance. By selecting the top important variables in each model utilised, we tried to expose similarities and differences between the three models. Such a strategy may help the reader and experts to build their conclusions based on the selected models and variables. It is a noteworthy approach to help clubs' executives and managers to reformulate their evaluations not only on preferences' basis but also on a more complex scientific modelling basis.

Finally, this chapter has explored new scopes in the domain of predicting the wages and transfer fees beyond the previous econometric approaches. It is a hopeful approach to improve further if a larger and more professional data set is available.

## CHAPTER SIX

### 6 FOOTBALL IN STOCK MARKET: TRANSFERS AND EUROPEAN COMPETITIONS EVENTS

#### 6.1 Introduction

For a long time, the valuation of football assets seemed a rather inexact science. When Roman Abramovich bought Chelsea, in 2003, a preferred joke was: “in order to become a millionaire, you only need to be a billionaire and buy a football club”, meaning pricing was not very accurate. But transactions on football clubs became more common, and football clubs became public as well. In the same year 2012, the Qatar Sports Investment acquired the Parisian club Paris Saint-Germain (PSG), and the Glazer’s family launched a public offering of Manchester United on the New York Stock Exchange. Does that mean that football valuation became more scientific? Now, the balance sheets of the most famous clubs are published and independent valuation can be accessed through internet. For instance, Manchester United had an asset value estimated to be 1.897bn USD at the end of June 2019<sup>6</sup>: by that time, the players were estimated at 0.906bn USD by transfermarkt<sup>7</sup> or 48% of the club’s assets. More generally, the player roster account for a large share of a club’s assets. A single player could account for some percent of the balance sheet: here, 6% in the case of Paul Pogba but Drut (2019) reports that Anelka was worth 59% of the income of his club in 2000. Hence one peculiar question worth answering is whether player transfers do have an impact on club valuation. It seems obvious, then, that substantial transfers may have an impact on the club’s overall value, especially (if markets are efficient) when the transfer fee paid is different from what was expected.

---

<sup>6</sup> <https://ycharts.com/companies/MANU/assets>

<sup>7</sup> 797.6m€, in fact: [https://www.transfermarkt.com/manchester-united/startseite/verein/985?saison\\_id=2018](https://www.transfermarkt.com/manchester-united/startseite/verein/985?saison_id=2018)

To correctly assess the impact of transfers on clubs' valuation, we must disentangle other possible causes of variation in prices of their stocks. Obviously, the outcome of matches may have an impact as well on clubs' stock prices and must then be controlled for, even if most transfers happen usually in a window when matches do not take place. Changes of coaches are also often said to potentially impact the clubs' returns. Systemic events should be taken into account by our methodology, it is thus not entirely out of line to look at the consequences of the coronavirus outbreak, which caused massive spillovers across sectors and countries, to calibrate and assess a methodology. The football industry first suffered the loss of its basic nerves, the football fans when clubs were forced to play behind closed doors. The following step was the temporary suspension of most renowned football leagues in the world. The fragility of the football industry was revealed by the severe loss in the market value of the likes of Juventus, Borussia Dortmund, Olympique Lyon, and many more (Figure 6.1, 6.2, 6.3).

We are thus about to study the impact of large transfers on football clubs' valuation and as an extension, the impact of recurring shocks such as football matches results and exceptional shocks such as the pandemic. Since we want to control for perturbation we will focus on short-term consequences and use the event study methodology. The rest of the chapter is organized as follows. Section 6.2 introduces the methodology, and Section 6.3 introduces the data. This is followed by a discussion of the results (Section 6.4), and Section 6.5 presents some concluding remarks.



Figure 6.1: JUVENTUS & MANCHESTER UNITED's stocks percentage change. Source: Google Finance



Figure 6.2: JUVENTUS's stock behaviour. Source: Google Finance



Figure 6.3: Manchester United's stock behaviour. Source: Google Finance

## 6.2 Data

The dataset used for the analysis was extracted from different sources. For instance, stocks (securities) and indices (market) data were extracted for the period covering January 2000 till August 2020 (twenty years), taking into consideration that some clubs went public after January 2000 like Manchester United (second listing in 2012) (Table 6.1). For the two tested hypothesis: unexpected match surprises and incoming transfers, data extraction covered six football clubs: Manchester United, Juventus, AS Roma, Ajax Amsterdam, Borussia Dortmund, and Galatasaray. The stock prices were extracted from *Bloomberg*, while historical match results and transfers data were extracted from *worldfootball.net* and *transfermarkt*, respectively. Market indices historical data were extracted from AEX, DAX, NYA, FTSEMIB, XU100.

Club	AS Roma	AFC Ajax	BVB	GSRAY	JUVE	MANU
Stock Ticker	ASR IM Equity	AJAX NA Equity	BVB GR Equity	GSRAY TI Equity	JUVE IM Equity	MANU US Equity
Index Ticker	FTSEMIB Index	AEX Index	DAX Index	XU100 Index	FTSEMIB Index	NYA Index
Start	5/22/2000	1/4/2000	10/30/2000	2/19/2002	12/19/2001	8/9/2012
End	12/16/2019	12/16/2019	12/16/2019	12/16/2019	12/16/2019	12/16/2019

Table 6.1: Data coverage period

After extraction of stocks and market data, betting odds were extracted from *football-data.co.uk* to distinguish between expected and unexpected (surprises) match results. The betting odds were released by the following bookmakers: *oddsportal.com*. So, through the extracted odds, the weights of the home win/loss, away win/loss, and a draw was determined. In total, 2731 matches had their expectations and actual match results (Table 6.2) covering both national leagues' and Champions League's matches. Historical match data were extracted from *worldfootball.net*.

Unexpected Match Events Breakdown						
Clubs	AJAX	ASR	BVB	GSRAY	JUVE	MANU
National League Matches per club	494	453	572	428	453	247
Champions League Matches per club	33	22	38	18	47	26
<b>Total number of matches per Club</b>	<b>527</b>	<b>475</b>	<b>510</b>	<b>446</b>	<b>500</b>	<b>273</b>
(-2) Negative Surprise (Loss instead of Win)	133	144	183	166	137	91
(-1) Negative Surprise (Draw instead of Win)	367	281	277	259	327	161
(+2) Positive Surprise (Win instead of Loss)	27	50	50	21	36	21
<b>Total number of Events</b>	<b>2731</b>					

Table 6.2: Unexpected Match Events

Although Bell et al. (2009) assumed that there is no association between transfers and the market value of football clubs, we attempted to verify the effect of incoming transfers of superstars on the stock price of the chosen clubs by extraction of transfer data from *transfermarkt* (Table 6.3). We attempted to breakdown the transfers into different categories to detect if different segments of transfers may have different effects on the share price, so we divided them into over-priced and under-priced transfers.

In order to build a sufficient sample, we took all inbound transfers into the chosen clubs costing more than 15m from the 1<sup>st</sup> of July 2007 onward.

Transfers Events							
Clubs	AJAX	ASR	BVB	GSRAY	JUVE	MANU	Total
Transfers	11	9	11	8	6	8	53
Under-Priced Transfers	4	4	3	7	1	0	19
Almost correctly priced transfers	2	1	1	1	1	1	7
Over-Priced Transfers	5	4	7	0	4	7	27

Table 6.3: Transfers Events above £15m 2007-2018

Among the available transfers, 19 transfers were classified under-priced and 27 over-priced transfers (Tables 6.4 and 6.5). The classification was based on the difference between the actual selling price (paid by the purchasing club) and the predicted market value by *transfermarkt* which is the prediction error which varies in magnitude from one transfer to the other.

<b>Over-priced Players (Prediction Error &gt; £ 15 M)</b>					
<b>Player's Name</b>	<b>Selling Club</b>	<b>Buying Club</b>	<b>Market Value (£)</b>	<b>Actual Trans. Fee (£)</b>	<b>Prediction Error</b>
Cristiano Ronaldo	Real Madrid	Juventus	90 M	110 M	£20 M
Paul Pogba	Juventus	Manchester United	63 M	95 M	£32 M
Anthony Martial	Monaco	Manchester United	7.2 M	54 M	£46.8 M
Angel Di Maria	Real Madrid	Manchester United	45 M	68 M	£23 M
Romelu Lukaku	Everton	Manchester United	45 M	76 M	£31 M
Paulo Dybala	Parma	Juventus	21 M	36 M	£15 M

Table 6.4: Examples of Over-priced Transfers. Source: *Transfermarkt*

<b>Under-priced Players (Prediction Error &lt; (£ 10 M)</b>					
<b>Player's Name</b>	<b>Selling Club</b>	<b>Buying Club</b>	<b>Market Value (£)</b>	<b>Actual Trans. Fee (£)</b>	<b>Prediction Error</b>
Mohamed Salah	Chelsea	AS Roma	16 M	4.5 M	(£11.5 M)
Eden Dzeko	Man City	AS Roma	18 M	3.6 M	(£ 14.4 M)
Bryan Cristante	Atalanta	AS Roma	23 M	4.5 M	(£18.5 M)
Mario Gotze	Bayern Munich	BVB	32 M	20 M	(£ 12 M)
Paco Alcacer	FC Barcelona	BVB	14 M	1.8 M	(£11.2 M)
Felipe Melo	Juventus	Galatasaray	15 M	1.4 M	(£13.6 M)

Table 6.5: Examples of Under-priced Transfers. Source: *Transfermarkt*

For testing the effect of the COVID-19, we sought the help of *finance.yahoo.com* to extract data of 10 football clubs: Manchester United, Juventus, AS Roma, SS Lazio, Olympique Lyon, Ajax Amsterdam, Borussia Dortmund, and Galatasaray, Beşiktaş J.K., Fenerbahçe S.K., in addition to their relevant market indices including CAC40, AEX, DAX, NYA, FTSEMIB, XU100.

### 6.3 Methodology

In this paper, we use the event study methodology, which was first developed by James Dolley (1933), to study the effect of different events independently (match results and incoming transfers) on the stock market performance of six reputable football clubs from different countries listed in various stock exchanges. In event studies, different hypotheses can be tested using specific expected return model, which is usually chosen to determine the theoretical returns to calculate the abnormal return after subtracting it from the actual stock return. Therefore, we use the Market Model in this paper. Tests carried out using the market model aims at testing whether publicly released information can trigger an immediate market response as the efficient market hypothesis specifies.

### **6.3.1 Hypothesis Testing**

We tested in this paper three hypotheses that we believed they could affect the share prices of football clubs. Thus, they could consequently affect revenues and profitability. In the football industry, the successful performance on the pitch and the transfer market activities are important factors to study. Therefore, we selected our hypothesis accordingly as follows:

**[H0: Expected results have no impact]**

**H1:** Unexpected victory (defeat/draw) triggers strong positive (negative) abnormal returns at match events.

**H2:** Players acquisitions (incoming transfers) triggers positive abnormal returns.

**H3:** Announcements associated with the pandemic Corona Virus (COVID-19) triggers negative abnormal returns.

### **6.3.2 Surprise on match results**

If the financial market is efficient, then expected victory or defeat as such should have no impact on stock prices: only unexpected results should have an impact since they are not



already included in prices. How can we define an unexpected match result? Simply when the result of the match is different from what was expected. The consensus expectation is given by betting odds, which are usually consistent from one operator to another; moreover, some data repositories add an average of market odds. Assume that, for a match between team 1 and team 2, 1X2 odds were  $O_1$  for team 1 winning the match,  $O_2$  for team 2 winning the match and  $O_X$  for draw. Team 1 is favourite, if  $O_1 < O_2$  and  $O_1 < O_X$ , which mean together that “team 1 winning the match” is more likely than both “team 2 winning the match” and “draw”. Table 6.6 gives a set of handy definitions for what we call surprise.

Pre-match odds	Favourite Team to Win	Expected result	Surprise	Greater surprise
$O_1 \leq \frac{1}{2}O_2$ $O_1 < O_X$	Team 1 - Strongly	Team 1 wins = Team 2 loses	Draw (negative for team 1)	Team 1 loses = Team 2 wins (negative for team 1)
$\frac{1}{2}O_2 < O_1 < O_2$ $O_1 < O_X$	Team 1	Team 1 wins = Team 2 loses	Draw (negative for team 1)	Team 1 loses = Team 2 wins (negative for team 1)
$O_X \leq O_1$ $O_X \leq O_2$	-	Draw	Team 1 wins (positive for team 1) OR Team 2 wins (negative for team 1)	X
$O_1 > O_2 > \frac{1}{2}O_1$ $O_X > O_2$	Team 2	Team 1 loses = Team 2 wins	Draw (positive for team 1)	Team 1 wins = Team 2 loses (positive for team 1)
$\frac{1}{2}O_1 \geq O_2$ $O_X > O_2$	Team 2 - Strongly	Team 1 loses = Team 2 wins	Draw (positive for team 1)	Team 1 wins = Team 2 loses (positive for team 1)

Table 6.6: Surprise according to pre-match odds

The table will help us introduce the concept of a *greater* surprise. First, surprise can be greater according to the number of “notches”: a draw game when a victory was expected is one notch, but a victory for a team which was expected to lose is a two-notch, hence greater surprise. Let us write (+1) when a team was expected to lose and eventually draw or (-2) when a team was expected to win and eventually loses.

There is another dimension of surprise, which is featured in the first and last lines: when odds are at least 2:1, *i. e.*  $O_1 \leq \frac{1}{2}O_2$ , team 1 is likely to win of course, and team 2 is more likely to lose than when odds are  $\frac{1}{2}O_2 < O_1 < O_2$ , hence it would be a “larger” surprise if team 2 draws or wins when odds are larger. Let us call such surprises a (x2) surprise since the odds ratio was greater than 2, but there can be (x3) or (x5) surprises as well, in the latter case when  $\frac{1}{5}O_2 < O_1 \leq \frac{1}{4}O_2$  and eventually, team 2 wins (+2x5) or draws (+1x5).

### 6.3.3 Surprise on transfer prices

When football transfer market shocks investors and stakeholders of the industry with unexpected transfers, it becomes crucial to test the reaction of the financial markets to such transfers and their effects on the listed clubs’ market value. The difference between the actual transfer fees paid for some transfers and their estimated market values given by *Transfermarkt* determines the nature of the transfer as an over-priced or under-priced one; a positive difference means the transfer is over-priced, while a negative one refers to under-priced one. It could be more reasonable to scrutinise the depth by which a transfer outperforms or underperforms the market value, thus we’ve tested transfers that exceed the market value by £5M, £10M, and £20M (Table 6.7). Moreover, we have tested those that lag behind the market value by £10M, £15M or more (Table 6.8).

<b>Over-priced Manchester United Incoming Transfers with different Prediction Error Categories</b>						
<b>Player</b>	<b>Selling Club</b>	<b>Date</b>	<b>Pricing error (£)</b>	<b>P. Error &gt; 5M</b>	<b>P. Error &gt; 10M</b>	<b>P. Error &gt; 20M</b>
Juan Mata	Chelsea	1/24/2014	6M	✓		
Angel Di Maria	Real Madrid	8/26/2014	23M	✓	✓	✓
Anthony Martial	Monaco	9/1/2015	46M	✓	✓	✓
Henrikh Mkhitaryan	Dortmund	7/6/2016	11M	✓	✓	

Paul Pogba	Juventus	8/9/2016	32M	✓	✓	✓
Romelu Lukaku	Everton	7/10/2017	31M	✓	✓	✓
Nemanja Matic	Chelsea	7/31/2017	8M	✓		

Table 6.7: Examples of Over-priced Transfers with different Prediction Error Categories.

Under-priced AS Roma Incoming Transfers with different Prediction Error Categories						
Player	Selling Club	Date	Pricing error (£)	P. Error < 0	P. Error < -10M	P. Error < -15M
Mohamed Salah	Chelsea	8/6/2015	(11M)	✓	✓	
Edin Dzeko	Man City	8/12/2015	(14M)	✓	✓	
Bryan Cristante	Atlanta	7/2/2018	(19M)	✓	✓	✓
Steven Nzonzi	Sevilla FC	8/14/2018	(3M)	✓		

Table 6.8: Examples of Under-priced Transfers with different Prediction Error Categories.

### 6.3.4 COVID-19 Effect

The outbreak of the pandemic coronavirus in late 2019 in Wuhan in China, then its quick diffusion towards Europe has left the whole continent in flounder and fear leading to the disruption of many entertaining industries, football being one of them, and to the acceleration of interconnected events and consequences. In general, financial markets responded aggressively by falling to unprecedented lows since the Black Monday of 1987, raising alerts of financial crisis which could lead to a severe recession across many sectors of the world economy. During this pandemic, the financial market scored remarkable crashes on three occasions leading them to be among the top 15 historical crashes (Table 6.9).

Historical Market Crashes				
Rank	Date	Close	Net Loss	% Loss
1	10/19/1987	1738.74	-508.00	-22.61
2	<b>3/16/2020</b>	<b>20188.52</b>	<b>-2,997.10</b>	<b>-12.93</b>
3	10/28/1929	260.64	-38.33	-12.82
4	10/29/1929	230.07	-30.57	-11.73
5	<b>3/12/2020</b>	<b>21200.62</b>	<b>-2,352.60</b>	<b>-9.99</b>
6	11/6/1929	232.13	-25.55	-9.92
7	12/18/1899	58.27	-5.57	-8.72
8	8/12/1932	63.11	-5.79	-8.40
9	3/14/1907	76.23	-6.89	-8.29

10	10/26/1987	1793.93	-156.83	-8.04
11	10/15/2008	8577.91	-733.08	-7.87
12	7/21/1933	88.71	-7.55	-7.84
13	<b>3/9/2020</b>	<b>23851.02</b>	<b>-2,013.76</b>	<b>-7.79</b>
14	10/18/1937	125.73	-10.57	-7.75
15	12/1/2008	8149.09	-679.95	-7.70

Table 6.9: List of Historical Financial Market Crashes. Source: *Spindices*

While market indices started falling from the 22<sup>nd</sup> of February, the deepest crashes occurred on March 9<sup>th</sup>, March 12<sup>th</sup>, and March 16<sup>th</sup>. The positive side of analysing the Covid-19 effects is that the designed events of study are not contaminated with other events, i.e. overlapping games or players transfers. Many events were attractive to study (Tables 6.10 and 6.11) during the trying period of the pandemic ranging from the temporary suspension of the top European Leagues until further notice, cancellation of the entire season 2019/2020 (France and Netherlands), and finally to positive announcements of resuming national leagues in the majority of the renowned European leagues.

League	Other Events	Closure Announcement	Reopening Announcement	Suspension Announcement	League Restart Date
Premier League	See Text Below	March 13	May 28	-	June 17
Bundesliga		March 13	May 6	-	May 16
French Ligue 1		March 13	-	April 30	-
Italy Serie A		March 10	May 28	-	June 20
Netherlands - Erdivisie		March 12	-	April 21	-
Turkish League		March 19	May 13	-	June 12
Spanish La Liga		March 12	May 23	-	June 8

Table 6.10: Examples of some league-related events and their occurrence dates

Of equal importance may have been the FIFA president Gianni Infantino's speech on the 28<sup>th</sup> of February or the official announcement by the Italian prime minister on the 4<sup>th</sup> of March (decreto ministeriale del Presidente del Consiglio del 4 marzo) that every social event will take

place without public attendance, this is to say football matches would be played behind closed doors. This was the first decision to actually constrain the schedule of football events.

Events studied	Dates	Event Description
Evt0, Evt1, Evt2, Evt3	10th, 12th, 13th, 19th of March	League temporal suspension
Evt4	4th of March	Italian PM announcement - Playing Behind Closed Doors
Evt10, Evt11	21st of April, 28th of April	Erdevisie (Netherlands), Ligue 1 (France) Cancellations
Evt20, Evt21	6th of May, 13th of May	Bundesliga (Germany), Turkish League Restarting Decisions

Table 6.11: Events covered by this study/analysis

The criteria for testing whether an event is significant depends if the event:

1. Has an immediate significant impact. This could be realised from the AAR (Average Abnormal Return) t-stat if AAR t-stat > Threshold.
2. Has a lasting effect recognised by the significant figures of the CAAR (Cumulative Average Abnormal Return) t-stat > Threshold at least for a certain number of days following the event.

### 6.3.5 Market Model

Several models can be used to determine abnormal returns, but the most regular is the market model, also called a single-factor model. To start with, an estimation period ranging between two hundred fifty days before the event date (-250) up till twenty (20) trading days preceding event dates is used to estimate the model of normal returns. Then, the returns of stock for a club *i* at date *t* is given by the following relation:

$$R_{it} = \text{Log}\left(\frac{P_{it}}{P_{it-1}}\right) \dots\dots\dots (1)$$

Depending on the market, we consider the predicted stock returns follow the following function:

$$R_{it}' = \alpha_i + \beta_i * R_{mt} + \varepsilon_{it} \dots\dots\dots (2)$$

Where  $\alpha_i$  is a constant,  $R_{it}$  being the  $i^{\text{th}}$  observed stock return at time t, while  $R_{mt}$  is the market return at time t. Concerning the random error term  $\varepsilon_{it}$  at time t, we assume that it is uncorrelated to the market return  $R_{mt}$  and stock returns  $R_{it}$ . The risk measure of the stock for a market is characterised by  $\beta_i$ , the coefficient generated by the regression analysis. After calculating the observable stock return, we move forward towards calculating the abnormal return as follows:

$$AR_{it} = R_{it} - R_{it}' \dots\dots\dots (3)$$

$$AR_{it} = R_{it} - (\alpha_i + \beta_i * R_{mt}) \dots\dots\dots (4)$$

After calculating the Abnormal Returns (AR) for all events i the Average Abnormal Return (AAR) is estimated in the following way:

$$AAR_{it} = \frac{1}{N} \sum_{i=1}^N AR_{it} \dots\dots\dots (5)$$

Where N is the total number of events. Finally, the Cumulative Average Abnormal Returns (CAAR) is defined as the sum over the T days in the event window [-20,20]:

$$CAAR_{it} = \sum_{i=1}^T AAR_{it} \dots\dots\dots (6)$$

The events are not happening at the same time; for this reason, we assume independence of abnormal returns. To measure the statistical significance of the AAR and the CAAR, we apply the following parametric t-test:

$$\text{T-test (AAR)} = \text{AAR} / [\sigma_{\text{AAR}} / \sqrt{N}] \dots\dots\dots (7)$$

$$\text{T-test (CAAR)} = \text{CAAR} / [\sigma_{\text{CAAR}} / \sqrt{N}] \dots\dots\dots (8)$$

- $\sigma_{\text{AAR}}$  = is the cross-sectional standard deviation of average abnormal return calculated over estimation period (40 days)
- $\sigma_{\text{CAAR}}$  = is the cross-sectional Standard deviation of cumulative average abnormal return calculated over estimation period (40 days)
- N = Number of observations.

### 6.3.6 Robustness Check

Finally, to test for the robustness of the results, we try to test using different event windows keeping the estimation period the same. In general, event studies may use specific windows that comply with the nature of the study, i.e. Ardekani et al. (2019) used (-1 day, 3 days) in their testing which needs a monetary cycle of one year. Others may use any customized window like (-1,+1), (-3,+3), (-5,+5), (-10,+10), (-15,15), (-20,20) ... Testing using different windows may assure whether the results generated are robust enough or not.

## 6.4 Results

Two different analyses have been performed: one to study the unexpected match effect and another to examine the effect of incoming transfers of superstars. According to the efficient market hypothesis, the share prices are just a reflection of all available information; thus, stocks movement is reflected on the first day following the flow of information. An investor can just take a higher risk to earn higher returns. The following analysis demonstrates this hypothesis.

### 6.4.1 Expected Match Result

In this section, results of matches which have met the predictions without surprises have been tested.

### 6.4.2 Unexpected Match Events

In this category, we test positive and negative match surprises. A win after a loss or draw expectation is considered a positive surprise, while a loss or a draw after a win expectation is regarded as a negative surprise. A win surprise produces a surge in abnormal returns by 1% in one day, and it approaches 2% within five days from gameday (Figures 6.4 and 6.5) while a loss surprise plunges abnormal stock returns by around 1.5% within two days (Figure 6.4). It is necessary to mention that the abnormal returns are observed on the first trading day after the game. Typically, since the study considers national league matches that are played continuously every weekend, with little exceptions when interrupted with Champions League matches that are played in midweek, it was good to restrict the estimation window to five days before and after the match.

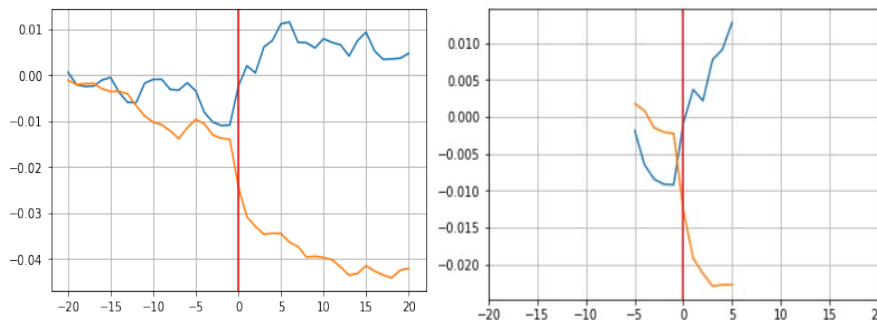


Figure 6.4: Positive (blue) & Negative (orange) Surprises

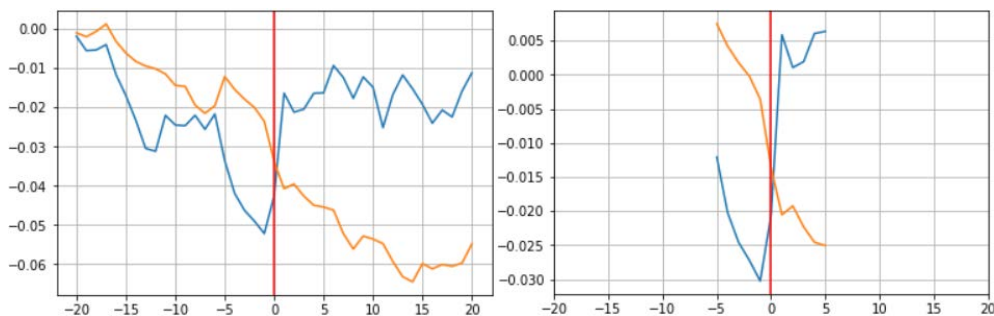




Figure 6.5: Positive (blue) & Negative (orange) Surprises with prediction ratio > 2

Both win and loss surprises affected the stock price significantly on the first two trading days (see Table 6.12) after the weekend that the game was played. This can be justified because football matches are continuous ongoing events and therefore the effect of a match does not last so long after the day of the match unless maybe it is a final game, but that also needs to be investigated in future.

	AAR (%)	t-test
(+2) Positive surprise: Win unexpected	0.008418	4.013801***
(-2) Negative: Loss: win expected	-0.00724	-4.23062***
(-1) Negative: Draw: win expected	-0.00453	-1.93505*

Table 6.12: Abnormal Returns-National League Matches on the first trading day

\*\*\* Indicates statistical significance at the 1% level, \* at 10%

Both win and loss surprises are statistically significant at the 1% confidence level for the two trading days following the game (see Tables 6.13, 6.14, 6.15 for details). Moreover, we could not ignore that significant negative effect of a surprise caused by a match draw when a club was expected to win. The decline in the average abnormal return is around 1% during the first two trading days after the match. Overall, we may conclude that results were significant on the first two trading days after the matches in cases of a win, draw, and loss due to the fast market adjustment triggered by the continuity of events during a season. When we restrict the study to matches with winning or losing odds ratio greater than two (Figure 6.14), the stock market responds faster and more aggressively to a surprise (both positive and negative), with slightly higher response to the unexpected loss. That matches the conclusion of Scholtens and Peenstra (2009). This can be justified that the public responds emotionally in a more sensitive way to the losses as suggested by Reilly and Gilbourne (2003), then Pain and Harwood (2004).

Negative Surprise (Loss when Win was expected)						
Days	AAR	CAAR	S(AAR)	S(CAAR)	t-stat(AAR)	t-stat(CAAR)
-20	-0.00083	-0.00083	0.028478	0.028478	-0.324413	-0.324413
-19	-0.0014	-0.00223	0.034447	0.04178	-0.455306	-0.596517
-18	0.000646	-0.00158	0.018318	0.045872	0.394543	-0.385755
-17	-0.00066	-0.00225	0.024973	0.052707	-0.296963	-0.476432
-16	-0.00243	-0.00467	0.030392	0.05326	-0.893348	-0.981259
-15	-0.00073	-0.0054	0.022718	0.057138	-0.357561	-1.056828
-14	0.003454	-0.00195	0.025489	0.060342	1.515233	-0.360659
-13	0.001071	-0.00088	0.022007	0.065704	0.544264	-0.148934
-12	-0.00336	-0.00423	0.026391	0.072596	-1.42143	-0.651531
-11	-0.00236	-0.00659	0.025299	0.072529	-1.043317	-1.016065
-10	-0.00056	-0.00715	0.019044	0.077959	-0.326263	-1.024981
-9	0.000895	-0.00625	0.0278	0.085358	0.359843	-0.818942
-8	0.001401	-0.00485	0.022577	0.091199	0.693539	-0.594797
-7	-0.00207	-0.00692	0.024889	0.098214	-0.928377	-0.787583
-6	0.002987	-0.00393	0.021581	0.09509	1.547477	-0.462256
-5	0.00121	-0.00272	0.027862	0.09742	0.485709	-0.312287
-4	0.001452	-0.00127	0.034045	0.109882	0.476875	-0.129118
-3	-0.00514	-0.00641	0.035266	0.112734	-1.629039*	-0.635455
-2	-0.00173	-0.00814	0.021813	0.113733	-0.885483	-0.799705
-1	0.000893	-0.00724	0.019427	0.122197	0.514153	-0.662571
0	-0.00724	-0.01448	0.019137	0.120451	-4.230618***	-1.344328
1	-0.00743	-0.02192	0.03343	0.117941	-2.485956***	-2.077562
2	-0.00106	-0.02297	0.025896	0.120019	-0.455418	-2.139853
3	-0.00128	-0.02425	0.022546	0.117802	-0.635113	-2.301683
4	0.001582	-0.02267	0.019915	0.12471	0.888082	-2.032376
5	0.00016	-0.02251	0.026776	0.127335	0.066934	-1.976404
6	-0.00307	-0.02558	0.024359	0.130173	-1.410298	-2.197217
7	0.001033	-0.02455	0.026088	0.135653	0.442687	-2.023314
8	-0.00165	-0.0262	0.021232	0.13533	-0.867862	-2.164301
9	0.000795	-0.0254	0.022124	0.133498	0.401698	-2.127428
10	0.002375	-0.02303	0.026697	0.13638	0.994503	-1.8878
11	-0.00179	-0.02482	0.021414	0.137774	-0.93576	-2.014146
12	-0.00068	-0.0255	0.016652	0.138963	-0.454035	-2.051316
13	0.000874	-0.02462	0.026149	0.14278	0.373543	-1.928066
14	0.002312	-0.02231	0.03037	0.141944	0.851302	-1.757285

15	-0.00156	-0.02387	0.026261	0.137822	-0.662994	-1.936163
16	0.000334	-0.02353	0.022587	0.139895	0.165171	-1.880804
17	-0.00271	-0.02624	0.020414	0.139638	-1.482857	-2.101055
18	0.000909	-0.02533	0.025578	0.139352	0.397179	-2.032472
19	-3.90E-05	-0.02537	0.023801	0.139607	-0.018518	-2.031905
20	0.001583	-0.02379	0.028341	0.148094	0.62463	-1.795932

Table 6.13: Negative Surprise (Loss instead of Win)

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 5% level, \* at 10%, S:Standard Deviation

Negative Surprise (Draw when Win was expected)						
Days	AAR	CAAR	S(AAR)	S(CAAR)	t-stat(AAR)	t-stat(CAAR)
-20	-0.00077	-0.00077	0.0249	0.0249	-0.370293	-0.370293
-19	-0.0046	-0.00536	0.021956	0.031304	-2.511274	-2.055928
-18	0.001933	-0.00343	0.017219	0.034188	1.347318	-1.203921
-17	0.000796	-0.00263	0.023046	0.037005	0.414365	-0.854193
-16	0.001468	-0.00117	0.022234	0.040467	0.792335	-0.345787
-15	-0.00352	-0.00468	0.038535	0.053649	-1.094694	-1.047107
-14	-0.0004	-0.00508	0.030321	0.06008	-0.157902	-1.01472
-13	0.000625	-0.00446	0.0304	0.077393	0.246738	-0.690809
-12	0.000138	-0.00432	0.023778	0.081416	0.06946	-0.636384
-11	0.002442	-0.00188	0.026859	0.085748	1.090844	-0.262553
-10	0.003504	0.001628	0.025569	0.090558	1.644723*	0.215776
-9	0.002114	0.003743	0.019711	0.090779	1.28718	0.494733
-8	0.002591	0.006334	0.031171	0.097968	0.997485	0.775801
-7	0.00203	0.008364	0.024368	0.107359	0.999923	0.934896
-6	-0.00014	0.008225	0.024994	0.099268	-0.066804	0.994279
-5	0.000817	0.009042	0.029252	0.101448	0.335332	1.069596
-4	-0.00025	0.008792	0.025915	0.108985	-0.116116	0.968017
-3	-0.00089	0.007898	0.027743	0.108146	-0.386417	0.8764
-2	-0.00278	0.005115	0.028422	0.110909	-1.17512	0.553425
-1	0.004471	0.009586	0.024282	0.121563	2.209545**	0.946272
0	-0.00453	0.005055	0.028096	0.126669	-1.93505**	0.478932
1	-0.00564	-0.00058	0.027255	0.122863	-2.480908**	-0.05657
2	0.002154	0.001575	0.023102	0.127646	1.118817	0.148036
3	-2.70E-05	0.001548	0.019271	0.129889	-0.016835	0.142981
4	-0.0013	0.000249	0.020149	0.129802	-0.77325	0.023049

5	-0.00235	-0.0021	0.031416	0.128657	-0.898949	-0.196253
6	-0.00201	-0.00411	0.027278	0.136058	-0.882442	-0.362499
7	-0.00245	-0.00656	0.02677	0.142176	-1.095979	-0.55326
8	-0.00448	-0.01104	0.022599	0.141013	-2.380291**	-0.939285
9	0.000457	-0.01058	0.0219	0.137499	0.250244	-0.923431
10	-0.0026	-0.01318	0.026774	0.141309	-1.164001	-1.11908
11	0.000015	-0.01316	0.023469	0.141787	0.007872	-1.114009
12	-0.00382	-0.01698	0.020976	0.144202	-2.182482	-1.412813
13	-0.00084	-0.01781	0.02264	0.144444	-0.443096	-1.479903
14	0.000731	-0.01708	0.026689	0.144991	0.32878	-1.413798
15	-0.00167	-0.01875	0.028775	0.149804	-0.694656	-1.50181
16	-0.001	-0.01974	0.023814	0.156352	-0.501212	-1.515247
17	-0.00044	-0.02018	0.02711	0.158548	-0.195076	-1.527623
18	-0.00017	-0.02036	0.017153	0.158462	-0.122006	-1.541659
19	-0.00025	-0.02061	0.026966	0.159971	-0.110324	-1.54571
20	0.002609	-0.018	0.032578	0.160572	0.961084	-1.344932

Table 6.14: Negative Surprise (Draw instead of Win)

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 5% level, \* at 10%, S:Standard Deviation

<b>Positive Surprise (Win when Loss was expected)</b>						
<b>Days</b>	<b>AAR</b>	<b>CAAR</b>	<b>S(AAR)</b>	<b>S(CAAR)</b>	<b>t-stat(AAR)</b>	<b>t-stat(CAAR)</b>
-20	0.000553	0.000553	0.027287	0.027287	0.290135	0.290135
-19	-0.00269	-0.00214	0.019201	0.030569	-2.004815	-1.000278
-18	-0.00039	-0.00253	0.019122	0.034204	-0.292634	-1.057553
-17	0.000082	-0.00244	0.021426	0.042505	0.054962	-0.823329
-16	0.001317	-0.00113	0.02363	0.049258	0.79782	-0.327714
-15	0.000553	-0.00058	0.032628	0.061088	0.242631	-0.134659
-14	-0.00326	-0.00383	0.021569	0.062376	-2.161841**	-0.879435
-13	-0.00215	-0.00598	0.019175	0.065725	-1.601672*	-1.301892
-12	-8.70E-05	-0.00606	0.021398	0.069714	-0.057905	-1.245174
-11	0.004253	-0.00181	0.020741	0.072797	2.936203***	-0.355882
-10	0.000818	-0.00099	0.024845	0.077612	0.471279	-0.182941
-9	0.000016	-0.00098	0.023697	0.083	0.009547	-0.168338
-8	-0.0022	-0.00317	0.024788	0.090028	-1.268014	-0.504321
-7	-0.00015	-0.00332	0.023732	0.093312	-0.091781	-0.509918
-6	0.001591	-0.00173	0.019669	0.096351	1.158076	-0.257421
-5	-0.00178	-0.00351	0.026903	0.097578	-0.946863	-0.51524

-4	-0.00465	-0.00816	0.019545	0.097455	-3.406269***	-1.199027
-3	-0.00212	-0.01028	0.023256	0.099755	-1.302354	-1.474998
-2	-0.00077	-0.01105	0.023644	0.103039	-0.46685	-1.535112
-1	0.000133	-0.01091	0.020984	0.105933	0.091013	-1.475144
0	0.008418	-0.0025	0.030028	0.113696	4.013801***	-0.314365
1	0.004453	0.001957	0.030663	0.111474	2.079336**	0.251336
2	-0.00152	0.000437	0.023653	0.115524	-0.919853	0.054189
3	0.005607	0.006045	0.031095	0.125964	2.582001**	0.687076
4	0.001433	0.007477	0.023999	0.130933	0.854739	0.817668
5	0.003613	0.01109	0.036696	0.143664	1.40953	1.105243
6	0.000398	0.011488	0.033291	0.156422	0.171289	1.051554
7	-0.0044	0.007085	0.040053	0.147378	-1.573904	0.688342
8	-0.00011	0.006975	0.020368	0.14998	-0.077574	0.665862
9	-0.00112	0.005853	0.025321	0.157196	-0.63466	0.533067
10	0.001994	0.007847	0.03298	0.166194	0.865769	0.676015
11	-0.0008	0.007052	0.028782	0.170408	-0.395632	0.592473
12	-0.00053	0.006519	0.021328	0.170984	-0.357322	0.545905
13	-0.00242	0.004098	0.02205	0.17118	-1.572451	0.34273
14	0.003354	0.007451	0.024304	0.173521	1.975812	0.614846
15	0.001822	0.009274	0.030643	0.178295	0.851355	0.744703
16	-0.00409	0.00518	0.026017	0.177395	-2.252732	0.418097
17	-0.00181	0.003369	0.026367	0.180082	-0.983332	0.267883
18	0.000085	0.003454	0.02406	0.183504	0.050517	0.269511
19	0.000148	0.003602	0.020405	0.182898	0.103964	0.282002
20	0.00107	0.004673	0.035498	0.184018	0.431698	0.363563

Table 6.15: Positive Surprise (Win instead of Loss)

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 5% level, \* at 10%, S: Standard Deviation

### 6.4.3 Transfers Events

Despite the shortage in the literature concerning the effect of players' transfers, few researchers like Stadtmann (2006) concluded that transfers do not affect the abnormal returns. Our study on players transfers shows a clear positive effect on the abnormal returns of the listed football club. The results (Figure 6.6) shows a clear 5% positive abnormal returns on the publicly traded football clubs, but the t-test shows that this is insignificant (Tables 6.16 and 6.17). Even if we

shorten the event window (Table 6.18), the effects remain insignificant. The study shows that the share price starts soaring a few days before the signing of the superstar players. This could be justified by the rumours and news preceding the transfer announcement. For this reason, it is necessary to have larger estimation windows in case of transfers events, unlike unexpected match results events. Although the effect differs from one club to another, as shown in Figure 6.7, it shows that Italian clubs display convergence in their patterns while the different patterns of Manchester United curve may be due to the overlapping events of the games and the transfers in previous years before the recent decision by the Premier League to close the summer transfer window just before the start of the season, i.e. Paul Pogba's transfer occurred amidst two games before and after the signing. If we segment the transfers into over-priced and under-priced transfers, we notice that over-priced transfers have no significant effect (Table 6.19), whereas the under-priced transfers reveal a significant positive surge in the abnormal returns by 5% during three days period (Table 6.20). Overall, there is a remarkable significant surge in the share price a few days before the transfer signing day while results (see Tables 6.16, 6.17, 6.18) show no significance on the day of signing, like expected (unexpected) match results, or the first day after signing except for a vector of under-priced superstars. Our findings may refer to the excessive rumours and media coverage following over-priced superstars, which may leak a lot of information and details about the contract details and approximate value in advance causing an earlier surge in share price, more aggressive than the day of signing.

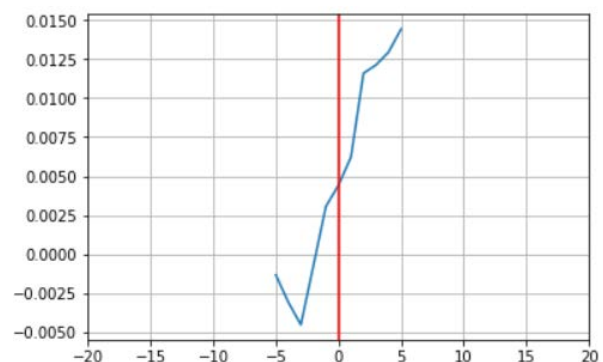
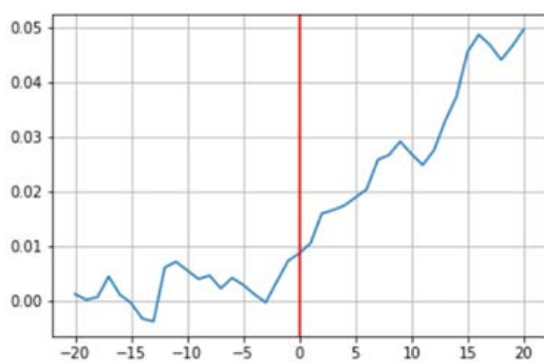


Figure 6.6: All Incoming Transfers with different event windows

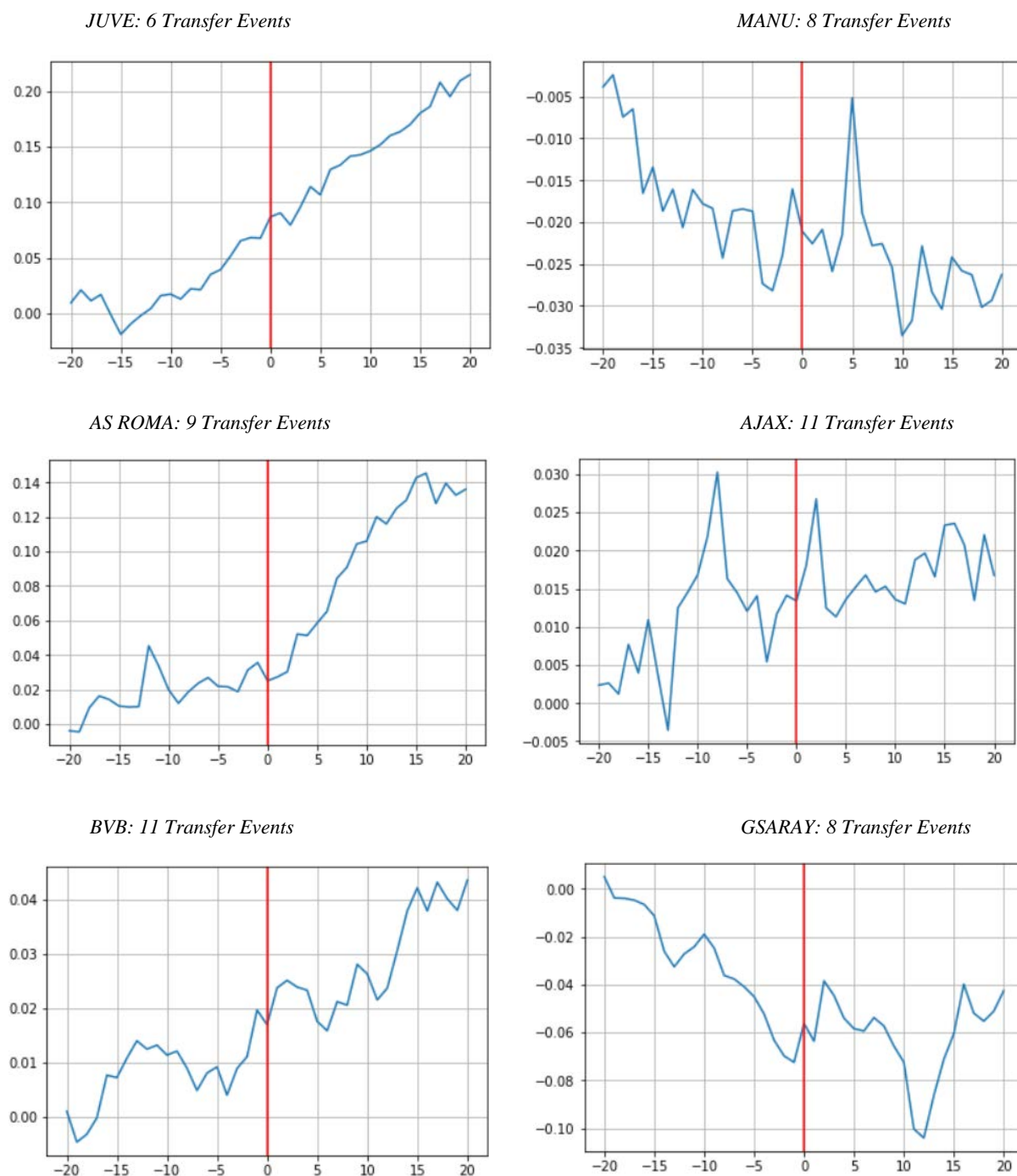


Figure 6.7: Incoming Transfers effect on different clubs' share prices

Transfer Category	AR (%)	t-test
Overall Transfers	0.001354	0.274068
Over-priced Transfers	-0.00732	-0.852
Under-priced Transfers	0.013018	2.6147***

Table 6.16: Transfers Abnormal Returns at t=0  
 \*\*\* Indicate statistical significance at the 1% level

Transfers Events - All Clubs						
Days	AAR	CAAR	S(AAR)	S(CAAR)	t-stat (AAR)	t-stat (CAAR)
-20	0.001267	0.001267	0.020622	0.020622	0.447398	0.447398
-19	-0.00104	0.000229	0.017451	0.029719	-0.433302	0.056011
-18	0.000498	0.000727	0.016675	0.033092	0.217385	0.15984
-17	0.003777	0.004504	0.020019	0.036967	1.373574	0.886941
-16	-0.00335	0.001154	0.019884	0.039983	-1.226345	0.210172
-15	-0.00153	-0.00037	0.028646	0.039973	-0.387578	-0.06753
-14	-0.00283	-0.0032	0.020281	0.041127	-1.015087	-0.566208
-13	-0.00051	-0.00371	0.015773	0.048063	-0.236206	-0.562016
-12	0.009826	0.006115	0.040083	0.056427	1.784656*	0.789003
-11	0.001061	0.007176	0.023368	0.057356	0.330477	0.910869
-10	-0.00156	0.005614	0.01428	0.057008	-0.796402	0.71694
-9	-0.00159	0.00402	0.014406	0.059879	-0.805354	0.488806
-8	0.000636	0.004656	0.018823	0.065252	0.24599	0.519512
-7	-0.00236	0.002296	0.018091	0.06368	-0.949911	0.262473
-6	0.001946	0.004242	0.020462	0.061143	0.692505	0.50512
-5	-0.0013	0.002939	0.015316	0.063382	-0.619613	0.337548
-4	-0.00173	0.001214	0.01958	0.07383	-0.641409	0.119678
-3	-0.00148	-0.00026	0.021357	0.086058	-0.502617	-0.022064
-2	0.003845	0.003584	0.024959	0.09898	1.12148	0.263606
-1	0.003769	0.007353	0.020703	0.100678	1.325306	0.531694
0	0.001354	0.008707	0.035975	0.101758	0.274068	0.622946
1	0.001837	0.010544	0.019238	0.096384	0.695232	0.796443
2	0.005411	0.015956	0.029835	0.092483	1.320439	1.256012
3	0.000636	0.016592	0.027796	0.107115	0.166672	1.127697
4	0.000816	0.017408	0.021623	0.116691	0.274715	1.086057
5	0.001471	0.018879	0.01813	0.117645	0.590606	1.168263
6	0.001504	0.020383	0.017992	0.124173	0.608636	1.195036
7	0.005393	0.025776	0.027888	0.128223	1.40779	1.463477
8	0.000884	0.02666	0.029616	0.121236	0.217358	1.600921*
9	0.002466	0.029126	0.018891	0.123293	0.950268	1.719814*



10	-0.0023	0.026829	0.022328	0.126429	-0.749045	1.54487
11	-0.00201	0.024816	0.024475	0.131363	-0.598606	1.375306
12	0.002661	0.027477	0.022639	0.134215	0.855615	1.490407
13	0.005368	0.032845	0.01851	0.135233	2.11147**	1.768194*
14	0.004391	0.037237	0.023124	0.135934	1.382562	1.994255*
15	0.00821	0.045447	0.022655	0.136507	2.638185***	2.423736**
16	0.003157	0.048603	0.027236	0.131875	0.843767	2.683117***
17	-0.00189	0.046713	0.039656	0.135911	-0.347074	2.502183**
18	-0.00269	0.044021	0.024139	0.135359	-0.811826	2.367609**
19	0.002531	0.046552	0.025157	0.139631	0.732413	2.427134**
20	0.003001	0.049553	0.016121	0.140106	1.355413	2.574853**

Table 6.17: Transfers events

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 1% level, \* at 10%, S: Standard Deviation

Transfers – Over-priced						
Days	AAR	CAAR	S(AAR)	S(CAAR)	t-stat(AAR)	t-stat(CAAR)
-5	-0.00134	-0.00134	0.015175	0.015175	-0.6444	-0.6444
-4	-0.00175	-0.0031	0.019398	0.026511	-0.65799	-0.85029
-3	-0.00145	-0.00454	0.021168	0.039988	-0.49796	-0.82732
-2	0.003786	-0.00076	0.024867	0.052737	1.108458	-0.10466
-1	0.003815	0.003056	0.020729	0.052761	1.339714	0.421732
0	0.001349	0.004405	0.035864	0.064384	0.273788	0.498103
1	0.001816	0.006221	0.019226	0.056605	0.687568	0.8001
2	0.005382	0.011603	0.029811	0.053636	1.314443	1.574949
3	0.000531	0.012134	0.027736	0.066301	0.139394	1.332411
4	0.000823	0.012958	0.02153	0.077607	0.278433	1.21554
5	0.00148	0.014438	0.018009	0.076488	0.598328	1.374191

Table 6.18: Transfers [-5,+5]

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 1% level, \* at 10%

Transfers – Over-priced						
Days	AAR	CAAR	S(AAR)	S(CAAR)	t-stat(AAR)	t-stat(CAAR)
-20	-0.00226	-0.00226	0.015509	0.015509	-0.757152	-0.757152
-19	-0.000528	-0.002788	0.013523	0.017008	-0.203022	-0.851843
-18	-0.003489	-0.006278	0.009315	0.019179	-1.946487	-1.70079
-17	0.001356	-0.004922	0.011629	0.022265	0.605711	-1.148696
-16	-0.002045	-0.006967	0.02233	0.03709	-0.47592	-0.976068
-15	0.002275	-0.004692	0.034443	0.043768	0.343159	-0.557099
-14	-0.002065	-0.006757	0.02309	0.040342	-0.464614	-0.870329
-13	-0.000724	-0.007481	0.01727	0.048175	-0.217884	-0.806933

-12	0.012051	0.00457	0.053476	0.061516	1.171019	0.386031
-11	-0.001095	0.003475	0.026013	0.059255	-0.218659	0.304769
-10	-0.006452	-0.002977	0.012917	0.053699	-2.595645	-0.288062
-9	-0.001746	-0.004723	0.014798	0.053931	-0.613142	-0.455058
-8	0.001335	-0.003388	0.016475	0.057648	0.420955	-0.305412
-7	-0.001663	-0.005051	0.019933	0.054651	-0.433393	-0.480234
-6	0.003965	-0.001086	0.022376	0.049787	0.920669	-0.113371
-5	-0.002144	-0.00323	0.017198	0.050803	-0.647718	-0.330367
-4	0.000194	-0.003036	0.021953	0.065553	0.045986	-0.240634
-3	0.001956	-0.001079	0.022509	0.078847	0.451641	-0.071127
-2	0.006008	0.004929	0.025921	0.092108	1.204454	0.278066
-1	0.00134	0.00627	0.020735	0.09129	0.335925	0.356856
<b>0</b>	<b>-0.007322</b>	<b>-0.001053</b>	<b>0.044659</b>	<b>0.101082</b>	<b>-0.851966</b>	<b>-0.054118</b>
1	0.004025	0.002973	0.018532	0.092708	1.128683	0.166614
2	-0.002511	0.000461	0.023735	0.088121	-0.549814	0.027198
3	-0.003964	-0.003503	0.014602	0.094157	-1.410687	-0.193319
4	0.000143	-0.00336	0.017876	0.103612	0.04161	-0.168499
5	0.00453	0.001171	0.017672	0.105171	1.332079	0.057836
6	-0.000471	0.0007	0.01899	0.109743	-0.128806	0.033137
7	0.00793	0.00863	0.023083	0.104529	1.785071	0.428994
8	0.005976	0.014606	0.028909	0.104109	1.07409	0.728974
9	0.004435	0.019041	0.014406	0.103425	1.599642	0.956608
10	-0.002306	0.016734	0.013966	0.104897	-0.857985	0.828954
11	0.001194	0.017929	0.021889	0.09348	0.283494	0.996582
12	0.003765	0.021693	0.013116	0.093211	1.491461	1.209317
13	0.002475	0.024169	0.015406	0.092375	0.834869	1.359495
14	0.001013	0.025182	0.024101	0.098353	0.21851	1.330405
15	0.005554	0.030736	0.020927	0.101941	1.37913	1.566701
16	0.001636	0.032373	0.017978	0.113309	0.472991	1.484566
17	-0.005928	0.026445	0.041932	0.098502	-0.734603	1.395003
18	0.002298	0.028743	0.024517	0.107013	0.487113	1.395653
19	-0.000035	0.028708	0.024507	0.099534	-0.007335	1.498716
20	0.001437	0.030146	0.017804	0.101589	0.419438	1.541914

Table 6.19: Over-priced Transfers

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 5% level, \* at 10%, S: Standard Deviation

Transfers – Under-priced						
Days	AAR	CAAR	S(AAR)	S(CAAR)	t-stat(AAR)	t-stat(CAAR)
-20	0.002951	0.002951	0.017686	0.017686	0.727313	0.727313
-19	-0.007191	-0.00424	0.015771	0.023885	-1.987551	-0.773807
-18	0.008342	0.004102	0.021462	0.040217	1.694319	0.444597
-17	0.001657	0.005759	0.027139	0.034519	0.266061	0.727168
-16	-0.000349	0.00541	0.016066	0.035847	-0.094697	0.657789
-15	-0.00366	0.00175	0.011437	0.037171	-1.39468	0.205218

-14	-0.005175	-0.003425	0.018795	0.04388	-1.200143	-0.3402
-13	-0.001264	-0.004689	0.015553	0.049436	-0.354227	-0.413413
-12	0.007066	0.002378	0.018503	0.051368	1.664671	0.201758
-11	0.003031	0.005409	0.016949	0.050912	0.779549	0.463085
-10	0.001821	0.00723	0.011566	0.051938	0.686177	0.606743
-9	-0.003021	0.004209	0.015354	0.05515	-0.857501	0.332668
-8	-0.000995	0.003214	0.018587	0.064128	-0.233234	0.218491
-7	-0.000381	0.002834	0.010124	0.068859	-0.163925	0.17938
-6	-0.001572	0.001262	0.017612	0.070188	-0.389041	0.078365
-5	0.00133	0.002591	0.010475	0.073264	0.553271	0.154182
-4	-0.003177	-0.000585	0.019164	0.082581	-0.722619	-0.030902
-3	-0.009774	-0.010359	0.020322	0.094683	-2.096405**	-0.476902
-2	0.003498	-0.006862	0.026259	0.113872	0.580599	-0.262654
-1	0.005851	-0.001011	0.02259	0.119641	1.129017	-0.036816
<b>0</b>	<b>0.013018</b>	<b>0.012007</b>	<b>0.021701</b>	<b>0.111548</b>	<b>2.614733**</b>	<b>0.469193</b>
1	-0.000875	0.011132	0.020401	0.11033	-0.187029	0.43979
2	0.016125	0.027257	0.035979	0.104643	1.953576*	1.135383
3	0.010377	0.037634	0.039588	0.130237	1.142612	1.259581
4	-0.002627	0.035008	0.026073	0.142226	-0.439135	1.072901
5	-0.001106	0.033902	0.020735	0.143931	-0.232415	1.026707
6	0.004357	0.038259	0.018728	0.152022	1.01397	1.096982
7	-0.004534	0.033725	0.019888	0.157855	-0.993735	0.931247
8	-0.008723	0.025002	0.03011	0.142782	-1.262804	0.763257
9	0.002259	0.02726	0.024497	0.152122	0.401931	0.781116
10	-0.003046	0.024215	0.025463	0.152427	-0.521407	0.692455
11	-0.006125	0.01809	0.027453	0.17173	-0.97247	0.459159
12	0.001479	0.019569	0.033727	0.180847	0.191112	0.471655
13	0.007965	0.027533	0.020315	0.18145	1.70903*	0.661425
14	0.003898	0.031431	0.021383	0.175085	0.794528	0.782504
15	0.002054	0.033485	0.016577	0.168474	0.540066	0.866354
16	0.004149	0.037634	0.037237	0.142134	0.485652	1.154136
17	0.001442	0.039076	0.022246	0.147999	0.282492	1.150863
18	-0.004079	0.034997	0.015526	0.150501	-1.145136	1.013594
19	0.004263	0.039259	0.022472	0.155424	0.826814	1.101031
20	0.005789	0.045049	0.01579	0.153448	1.598163	1.279666

Table 6.20: Under-priced Transfers

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 1% level, \* at 10%, S: Standard Deviation

#### 6.4.4 COVID 19 Events

The accelerated spread of the pandemic has created many remarkable events that could be studied. We tried to cover as many as possible in our analysis to determine which events had a larger significant effect on the individual clubs, top five football leagues, Italian clubs, Turkish

clubs, and finally on all the clubs, in our sample, aggregated. In finding which event has a significant effect, we looked at the events that:

- 1- Happen before the stock market crashed because then stocks are driven by the crash.
- 2- Have an actual effect as per the figures and statistics.

According to this criterion, the Italian prime ministers' (PM) decree *Decreto ministeriale del Presidente del Consiglio del 4 marzo*<sup>8</sup> to impose playing behind closed doors seems to have a significant effect on the listed football clubs. The stocks of Italian clubs fell significantly by 35% after the announcement (Table 6.21), the French club Olympique Lyon was affected negatively as well (Figure 6.8) by losing at least 15% within three days. All other football clubs' stocks fell except Manchester United's one which is listed on the New York Stock Exchange in the United States where the pandemic's peak effect arrived at a later stage than Europe. In general, a significant loss of 30%-40% was recorded when aggregating all clubs in the sample or when testing the effect of the pandemic on the top five prestigious leagues (English Premier League, Spanish La Liga, German Bundesliga, Italian Serie A, and French Ligue 1) (Table 6.22). Testing for all the clubs in the sample generates similar results in terms of significant losses (Table 6.23). We cannot ignore the Serie A suspension decision on March 10<sup>th</sup>, one day after a big crash in international financial markets (Table 6.9), but such decision has left Italian clubs to suffer by a significant decline (Figure 6.9) of about 10% within 5 days (Table 6.24).

*All Clubs Aggregated*

*Top 5 European Leagues*

---

<sup>8</sup> <http://www.governo.it/sites/new.governo.it/files/DPCM4MARZO2020.pdf>

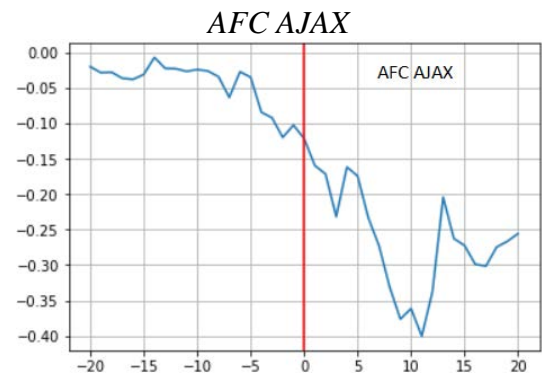
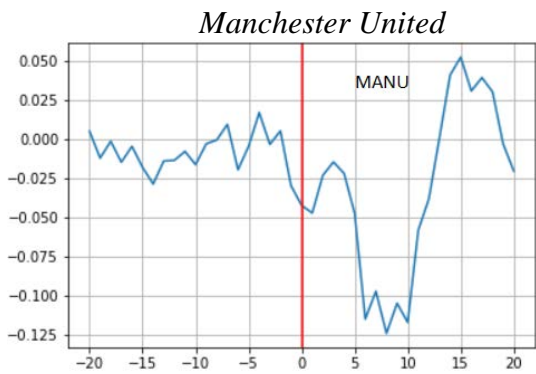
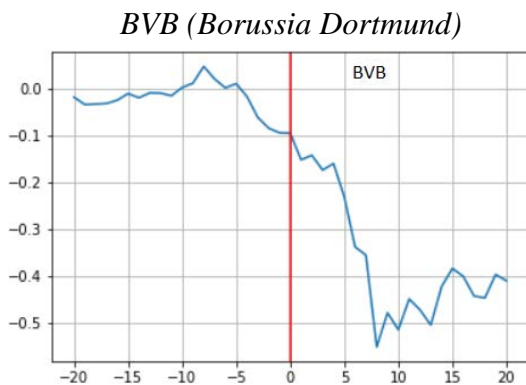
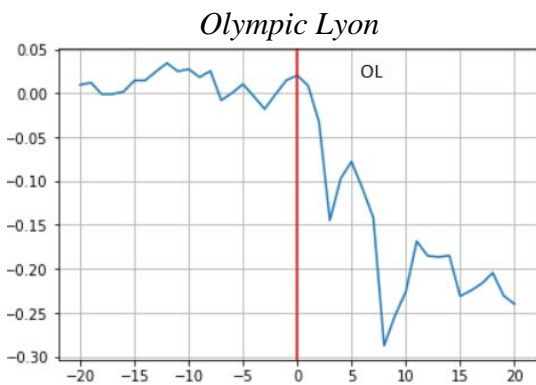
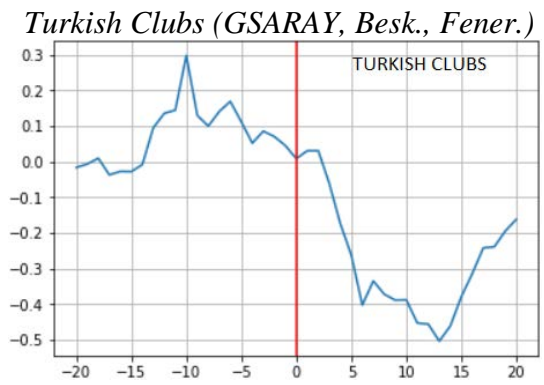
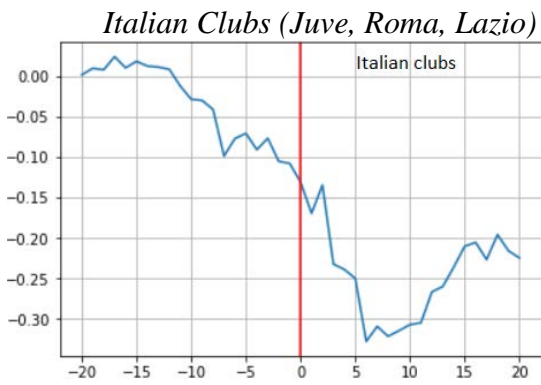
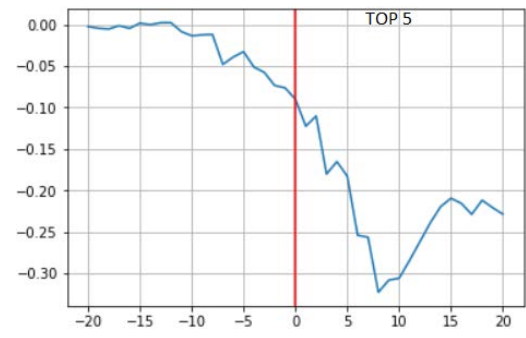
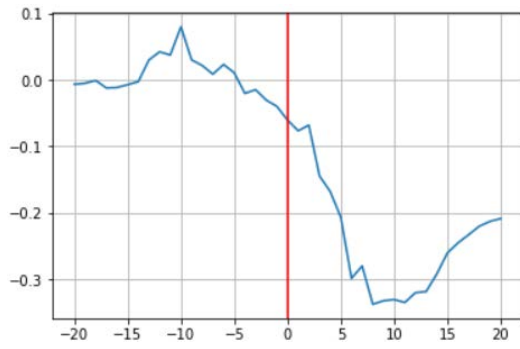


Figure 6.8: Clubs Reaction to the Italian PM's announcement (playing behind closed doors)

Effect of the Italian PM Decree on the Italian Clubs						
	AAR	CAAR	S(AAR)	S(CAAR)	t-stat (AAR)	t-stat(CAAR)
-20	-0.00523	-0.00523	0.016918	0.016918	-0.53531	-0.53531
-19	-0.01358	-0.01881	0.027161	0.010952	-0.86588	-2.97436
-18	-0.01069	-0.0295	0.009665	0.018969	-1.91643	-2.69371
-17	0.030672	0.001171	0.059539	0.040823	0.892283	0.049698
-16	0.009221	0.010392	0.009367	0.037782	1.705055	0.476423
-15	0.019017	0.029409	0.018587	0.04885	1.772057	1.042739
-14	-0.0018	0.027607	0.015899	0.059905	-0.19627	0.798218
-13	0.001108	0.028716	0.008327	0.068173	0.230486	0.729563
-12	0.031417	0.060133	0.061603	0.112976	0.883336	0.921899
-11	-0.00722	0.052916	0.012343	0.103288	-1.01271	0.887351
-10	-0.01609	0.036823	0.014797	0.092691	-1.88379	0.688081
-9	-0.00048	0.036348	0.010909	0.103571	-0.07537	0.60786
-8	-0.00593	0.030422	0.012616	0.109348	-0.81358	0.48188
-7	-0.06474	-0.03432	0.013562	0.112792	-8.26799***	-0.52697
-6	0.02811	-0.00621	0.013275	0.106509	3.667652***	-0.10093
-5	0.019675	0.013469	0.011403	0.116533	2.988639**	0.200188
-4	-0.02094	-0.00748	0.016229	0.126594	-2.23528**	-0.10227
-3	0.015763	0.008288	0.017079	0.110921	1.598598*	0.129419
-2	-0.04486	-0.03657	0.02033	0.092025	-3.82166***	-0.68826
-1	0.005621	-0.03095	0.016406	0.107433	0.59342	-0.49894
0	-0.05752	-0.08847	0.017039	0.105042	-5.84718***	-1.45879
1	-0.04803	-0.1365	0.00771	0.11017	-10.7909***	-2.14604**
2	0.018719	-0.11778	0.093481	0.067658	0.34683	-3.01525***
3	-0.13105	-0.24884	0.07933	0.052216	-2.86136**	-8.2541***
4	-0.01517	-0.26401	0.01783	0.046072	-1.47378	-9.92525***
5	-0.01283	-0.27684	0.020052	0.028549	-1.10797	-16.7954***
6	-0.12538	-0.40221	0.067165	0.068358	-3.23317***	-10.1912***
7	-0.0018	-0.40401	0.060004	0.12152	-0.05188	-5.75843***
8	-0.03344	-0.43745	0.092121	0.20976	-0.62869	-3.61212***
9	-0.00401	-0.44146	0.033064	0.242592	-0.21002	-3.15189***
10	0.005261	-0.43619	0.013366	0.255882	0.681788	-2.95256**
11	-0.00226	-0.43845	0.0118	0.261319	-0.3316	-2.90611***
12	0.045645	-0.39281	0.012531	0.268412	6.309154***	-2.53477**
13	0.014621	-0.37819	0.005984	0.265075	4.231857***	-2.47114**
14	0.036189	-0.342	0.042652	0.223115	1.469596	-2.65494**
15	0.041898	-0.3001	0.046004	0.186209	1.577441	-2.79142**
16	0.004352	-0.29575	0.016524	0.202617	0.456202	-2.52817**
17	-0.0336	-0.32935	0.008611	0.210965	-6.75872***	-2.704**
18	0.039258	-0.29009	0.008599	0.219419	7.907631***	-2.28992**
19	-0.01325	-0.30334	0.028393	0.227451	-0.80818	-2.30994***
20	-0.02132	-0.32466	0.022111	0.213892	-1.6701	-2.62902**

Table 6.21: Effect of Italian PM Decree on the Italian Clubs

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 1% level, \* at 10%, S: Standard Deviation

Effect of Italian PM Decree on Top 5 Leagues						
	AAR	CAAR	S(AAR)	S(CAAR)	t-stat (AAR)	t-stat(CAAR)
-20	-0.00197	-0.00197	0.012332	0.012332	-0.39074	-0.39074
-19	-0.00472	-0.00668	0.020068	0.015413	-0.57551	-1.06197
-18	-0.01106	-0.01774	0.008886	0.019275	-3.04789	-2.25424
-17	0.019048	0.001309	0.039543	0.026619	1.179921	0.120444
-16	0.003714	0.005023	0.011459	0.027953	0.793857	0.44013
-15	0.011434	0.016457	0.015634	0.034789	1.791466	1.158724
-14	-0.00206	0.014395	0.01291	0.043899	-0.39124	0.803215
-13	0.000376	0.014771	0.00937	0.051264	0.098309	0.70579
-12	0.017789	0.03256	0.042138	0.078098	1.034113	1.021236
-11	-0.00089	0.031674	0.014653	0.070694	-0.14812	1.097494
-10	-0.0082	0.023472	0.014149	0.06493	-1.42003	0.88549
-9	0.008118	0.03159	0.023004	0.076452	0.864401	1.012129
-8	-0.00867	0.022921	0.017625	0.072976	-1.20479	0.769368
-7	-0.03796	-0.01504	0.036516	0.081508	-2.54657	-0.45205
-6	0.012691	-0.00235	0.024776	0.07863	1.254689	-0.07325
-5	0.004253	0.001901	0.030711	0.073567	0.339187	0.063305
-4	-0.02592	-0.02402	0.038882	0.080481	-1.63315	-0.73114
-3	0.00249	-0.02153	0.01848	0.070001	0.330043	-0.75347
-2	-0.01864	-0.04017	0.033743	0.063462	-1.35294	-1.55047
-1	-0.00244	-0.04261	0.021379	0.075348	-0.27965	-1.38524
0	-0.03394	-0.07655	0.029594	0.086664	-2.80911**	-2.16362**
1	-0.02927	-0.10582	0.023104	0.10106	-3.1035**	-2.56494**
2	-0.00171	-0.10753	0.066963	0.074324	-0.06248	-3.54388***
3	-0.08695	-0.19448	0.080741	0.105881	-2.63788**	-4.49924***
4	-0.00056	-0.19504	0.029513	0.125742	-0.04628	-3.79943***
5	-0.03313	-0.22817	0.059889	0.132991	-1.35511	-4.20255***
6	-0.08822	-0.31639	0.0719	0.185965	-3.00546**	-4.16743***
7	-0.02915	-0.34555	0.078554	0.204003	-0.9091	-4.14901***
8	-0.04713	-0.39268	0.072988	0.203533	-1.58181	-4.72584***
9	0.006023	-0.38666	0.026156	0.228484	0.564063	-4.14519***
10	0.010636	-0.37602	0.023545	0.235882	1.106502	-3.90473***
11	0.024196	-0.35182	0.029974	0.25734	1.977342*	-3.34883***
12	0.011941	-0.33988	0.043913	0.249152	0.666042	-3.34149***
13	0.027057	-0.31283	0.029147	0.243637	2.273841	-3.1451***
14	0.024492	-0.28833	0.036763	0.221471	1.631891	-3.18899***
15	0.024723	-0.26361	0.051184	0.183988	1.183141	-3.50952***
16	-0.00947	-0.27308	0.024762	0.191164	-0.93659	-3.49911***
17	-0.01898	-0.29205	0.024243	0.210137	-1.91729	-3.40437***
18	0.031355	-0.2607	0.02782	0.196771	2.760728	-3.2453***
19	-0.01938	-0.28007	0.0207	0.193531	-2.29264	-3.54485***
20	-0.00674	-0.28682	0.033993	0.186149	-0.48593	-3.77416***

Table 6.22: Effect of Italian PM Decree on the Top Five Leagues

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 5% level, \* at 10%, S: Standard Deviation

Effect of Italian PM Decree on All clubs						
	AAR	CAAR	S(AAR)	S(CAAR)	t-stat(AAR)	t-stat(CAAR)
-20	-0.00735	-0.00735	0.013992	0.013992	-1.6618	-1.6618
-19	-0.00037	-0.00773	0.017565	0.018175	-0.06726	-1.34437
-18	-0.00216	-0.00988	0.013963	0.024834	-0.48851	-1.2585
-17	-0.00223	-0.01212	0.044788	0.034139	-0.15757	-1.12221
-16	0.005313	-0.0068	0.009678	0.032632	1.735867	-0.6592
-15	0.007946	0.001143	0.015374	0.039647	1.634358	0.091193
-14	0.006176	0.007319	0.014878	0.042161	1.312625	0.54896
-13	0.030356	0.037676	0.062137	0.069815	1.544914	1.706522
-12	0.023224	0.060899	0.068373	0.1298	1.074104	1.483666
-11	0.001915	0.062814	0.016155	0.137515	0.374929	1.444476
-10	0.042483	0.105297	0.081187	0.17159	1.654726	1.940553
-9	-0.04668	0.058613	0.089297	0.111292	-1.65324	1.665431
-8	-0.0161	0.042511	0.020928	0.102443	-2.43298	1.312263
-7	-0.01378	0.028733	0.086903	0.178273	-0.50138	0.509671
-6	0.019507	0.04824	0.045875	0.208847	1.344699	0.730431
-5	-0.01512	0.03312	0.037482	0.205007	-1.27562	0.510889
-4	-0.0387	-0.00558	0.034103	0.209836	-3.58817***	-0.08403
-3	0.00884	0.003264	0.028257	0.228158	0.989252	0.045234
-2	-0.01862	-0.01536	0.02727	0.225366	-2.15922	-0.21548
-1	-0.00613	-0.02149	0.027273	0.209889	-0.71127	-0.32379
0	-0.03478	-0.05627	0.02441	0.207488	-4.50582***	-0.85763
1	-0.0125	-0.06877	0.035107	0.2261	-1.12544	-0.96178
2	-0.0041	-0.07286	0.051532	0.221677	-0.2513	-1.03939
3	-0.08581	-0.15867	0.063114	0.226993	-4.29922***	-2.21043
4	-0.02749	-0.18616	0.067241	0.206436	-1.29281	-2.85164
5	-0.02175	-0.20791	0.048477	0.211765	-1.41869	-3.10465***
6	-0.13074	-0.33865	0.083583	0.219695	-4.94655***	-4.87451***
7	-0.00159	-0.34024	0.076406	0.223185	-0.06568	-4.82075***
8	-0.04583	-0.38606	0.059112	0.230103	-2.45159*	-5.30563***
9	-0.00852	-0.39459	0.039337	0.226728	-0.68519	-5.50348***
10	0.011467	-0.38312	0.023871	0.218883	1.519141	-5.53506***
11	-0.00866	-0.39178	0.053859	0.227981	-0.50857	-5.43432***
12	0.012612	-0.37917	0.039228	0.227053	1.016647	-5.28087***
13	0.016958	-0.36221	0.055586	0.232004	0.964746	-4.93704***
14	0.024159	-0.33805	0.033447	0.217841	2.284167*	-4.90731***
15	0.03984	-0.29821	0.046998	0.196382	2.680685*	-4.80202***
16	0.011756	-0.28646	0.041483	0.197036	0.896213	-4.5974***
17	0.008119	-0.27834	0.049491	0.217119	0.518772	-4.0539***
18	0.025204	-0.25313	0.032071	0.223183	2.485199*	-3.58663***
19	0.002438	-0.2507	0.035138	0.232479	0.219365	-3.41005***
20	0.006131	-0.24456	0.040994	0.247922	0.472957	-3.11945**

Table 6.23: Effect of Italian PM Decree on the All Clubs

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 5% level, \* at 10%, S: Standard Deviation



Effect of the League (Serie A) Suspension Decision on the Italian stocks						
	AAR	CAAR	S(AAR)	S(CAAR)	t-stat(AAR)	t-stat(CAAR)
-20	0.002624	0.002624	0.014015	0.014015	0.45861	0.45861
-19	0.009149	0.011773	0.015909	0.027633	1.408685	1.043583
-18	0.013497	0.02527	0.042463	0.047453	0.778597	1.304421
-17	-0.00383	0.021437	0.010905	0.044581	-0.86109	1.17781
-16	0.005834	0.027271	0.046116	0.054705	0.309889	1.221077
-15	-0.00362	0.023649	0.011053	0.048048	-0.80253	1.205646
-14	-0.01136	0.01229	0.014637	0.04257	-1.90098	0.707179
-13	-0.02857	-0.01628	0.035158	0.067569	-1.99047	-0.59016
-12	0.012203	-0.00408	0.023163	0.056416	1.29045	-0.17701
-11	-0.02074	-0.02482	0.044002	0.048031	-1.15477	-1.26582
-10	0.005644	-0.01918	0.030587	0.058374	0.451974	-0.80471
-9	0.018705	-0.00047	0.014413	0.051053	3.17878	-0.02267
-8	-0.03089	-0.03136	0.024065	0.061979	-3.14364	-1.23928
-7	0.011989	-0.01937	0.01746	0.062203	1.681899	-0.7627
-6	-0.05147	-0.07083	0.021496	0.062807	-5.86454***	-2.76249*
-5	-0.02084	-0.09168	0.029228	0.083557	-1.74687	-2.68752*
-4	-0.01816	-0.10984	0.078933	0.089671	-0.56349	-3.00028*
-3	-0.08125	-0.19108	0.055653	0.088009	-3.57604***	-5.3183***
-2	0.005504	-0.18558	0.067809	0.130709	0.198811	-3.47777***
-1	-0.06573	-0.25131	0.070159	0.098435	-2.29491*	-6.25374***
0	<b>-0.05667</b>	<b>-0.30798</b>	<b>0.066171</b>	<b>0.147852</b>	<b>-2.09766*</b>	<b>-5.10235***</b>
1	-0.0126	-0.32058	0.042553	0.171773	-0.72532	-4.57147***
2	-0.06489	-0.38547	0.078962	0.196319	-2.01282	-4.80947***
3	-0.00898	-0.39445	0.048202	0.227596	-0.45651	-4.24523***
4	-0.0101	-0.40455	0.057337	0.252584	-0.43144	-3.92318***
5	-0.0065	-0.41105	0.024987	0.270676	-0.63706	-3.71977***
6	0.024519	-0.38653	0.02386	0.283215	2.517169*	-3.34302***
7	0.004999	-0.38153	0.015021	0.289002	0.815146	-3.23371***
8	0.033715	-0.34781	0.026175	0.284553	3.155185**	-2.99404*
9	0.027203	-0.32061	0.030911	0.280573	2.155646	-2.79902*
10	0.014177	-0.30643	0.026129	0.275498	1.329035	-2.72454*
11	0.004667	-0.30177	0.047541	0.27677	0.240442	-2.67071*
12	0.020612	-0.28115	0.023844	0.287317	2.117509*	-2.39694*
13	-0.0228	-0.30395	0.020122	0.287685	-2.77543	-2.58801*
14	0.010208	-0.29375	0.033832	0.287038	0.739094	-2.50672*
15	-0.00645	-0.30019	0.020454	0.284825	-0.77233	-2.58166*
16	-0.00248	-0.30267	0.020993	0.276669	-0.28924	-2.67972*
17	-0.00202	-0.30469	0.028603	0.262332	-0.17262	-2.84499*
18	0.010169	-0.29452	0.014515	0.256509	1.715951	-2.81248*
19	0.013548	-0.28097	0.037222	0.231442	0.891568	-2.97369*
20	0.012152	-0.26882	0.022942	0.21162	1.297421	-3.11158**

Table 6.24: Effect of Italian League suspension on Italian Clubs

\*\*\* Reveals statistical significance at the 1% level, \*\* at the 1% level, \* at 10%, S: Standard Deviation

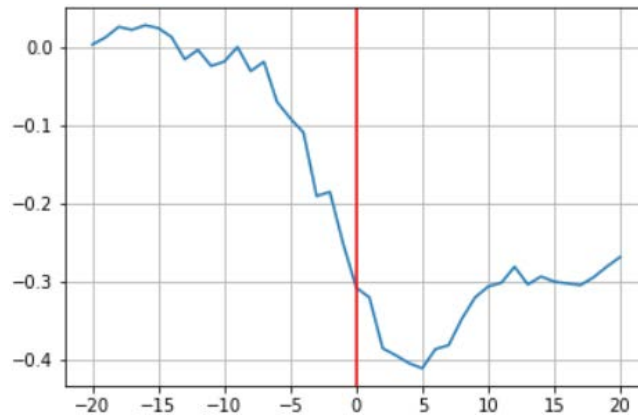


Figure 6.9: Effect of Italian League suspension on Italian Clubs

## 6.5 Conclusion

Listed football clubs, like other publicly traded companies, are affected by news and events surrounding the environment they perform in. In this study, we have drawn attention to few events that affect the fluctuation of the share price of Manchester United, Juventus, AS Roma, AFC Ajax, Borussia Dortmund, and Galatasaray collectively and individually. On the one hand, we concluded on a positive (negative) effect for an unexpected win (loss) on share prices. On the other hand, we noticed that under-priced incoming transfers (purchases) have a positive effect on the market value of football clubs in a contradiction to the findings of Fotaki et al. (2009), who found that the share price is affected by the sales contracts instead, but not purchasing contracts (incoming transfers). The pandemic outbreak COVID-19 slumped unexpectedly on the global economy in general and on the football industry in particular. Thus, it was necessary to look at its effects on the listed football clubs. The Italian clubs, like others in the continent, suffered an enormous decline in their market values amidst the emergence of the pandemic. Still, such declines were recovered gradually within a few weeks after that. Left to mention that some shareholders could have different motives than business people who invest for the sake of profit. This category of people is mostly loyal fans who buy shares to

keep them for their valuable meaning as a symbolic sense of belonging. Thus, they are not shaken by match results or players transfers to decide between buying or selling their shares, making a difference between an investor and a fan. Finally, our results are in line with Sholten and Peenstra (2012) in terms of unexpected match effects in the national leagues, while it contrasts Fotaki et al. (2009) and Bell et al. (2012) when it comes to players incoming and outgoing transfers. The results generated in this study may open other scopes of research in future on further possible effects of other segments of the transfer market on the stock market and the wealth of the shareholders.

## CHAPTER SEVEN

### 7 MAIN FINDINGS AND CONCLUSION

#### 7.1 Introduction

The fundamental aim of studying the dynamics of the pricing function of football players was to explore the determinants of players' transfer fees and salaries. Such a topic would not have been of much importance had we have not seen astronomical sums of money spent on superstars almost every year. As regular reports from financial experts and sports economists about potential bubble burst in the transfer market occupied sports media houses' stories, this research was deemed crucial to unfold the multifaceted factors affecting the market prices of players and their effect on football clubs. By testing the significant components that affect transfer fees or salaries of football players positively or negatively, we tried to explain the variance levels between the observed and predicted values and, therefore, answer problematic economic topics related to the uncertainty governing the football industry. We succeeded in achieving a global market analysis across three segments of the industry; attackers, midfielders, and defenders using a newly suggested pricing equation which was designed to be comprehensive and responsive to previous econometric challenges. Three-dimensional search topics were designed to approach pricing functions and factors. One last chapter was dedicated to testing the effects of transfers' events, in addition to match surprising results, on the market value of football clubs. We faced a lot of challenges, but we got promising results.

#### 7.2 Challenges

##### 7.2.1 Data Issues

Dealing with different large data sets simultaneously made it harder to design convenient samples for testing. With datasets extracted from different sources, *Transfermarkt*, *Futhead*, and *Sofifa*, the challenges encountered were many. For example, combining two data sets of

two different sources could not be completed without giving up some observations, that is to say losing a lot of time and effort. That is the difference between forming, manually, a customized data set and obtaining data sets from professional providers, i.e. Opta Sports, at a high cost.

### **7.2.2 Econometric Issues**

Such challenge encountered the passage through a minefield composed of econometric flaws, i.e. selection bias and heteroscedasticity, due to the poor historical literature and the scarcity of diversified data sets that could help to provide innovative solutions.

### **7.2.3 New Pricing Functions**

Most studies in the literature had a typical repetitive approach, that is to say computing the determinants of either transfer fees or salaries. Due to the lack of open sources providing data about salaries (except for Bundesliga), less literature about estimations of the salaries than transfer fees was found. The bold challenge of this research was to take an additional further, subjective, step ahead by forming different packages containing both salaries and transfer fees. Thus, computing three different pricing functions independently.

## **7.3 Main Findings**

### **7.3.1 Variables**

In chapters 3 and 4, two different data sets were used to estimate the newly designed pricing function (package). Evaluating players based on their performance statistics was not a new approach as most of the economists who searched for the determinants of transfer fees followed a similar path while evaluating players based on subjective scores (used in video games like PlayStation games), given by experts, to evaluate them has provided this research with a tool

to compare results in the two different ways. Remarkably, the results of the two different approaches converge in terms of findings. The two approaches differed by the nature of variables with respect to positions but converged by the significance of those variables (Table 7.1), i.e. the number of goals scored by an attacker being a real characteristic versus shooting and dribbling of an attacker being the synthetic characteristics (variables). It is noteworthy to mention that some missing variables (remaining duration of contracts) in previous studies happen to be influential in this study. Moreover, the loans and free transfers were present, through their dummy variables, for the first time in a study making the pricing function of this research unique and comprehensive. The controversy about the greed of agents by media houses and media, blaming them for the hike in prices of transfer fees and salaries, made it crucial to investigate the significance of the agent's dummy variables which proved such allegations insignificant, while buying clubs' financial strength was proved significant.

<b>Real Data (Chapter 3)</b>	<b>Synthetic Data (Chapter 4)</b>
Age, Age2	Age, Age2
Duration of Contract (at Current/Buying club)	Duration of Contract (at Current/Buying club)
Remaining Duration of Contract (at Previous/Selling Club)	Remaining Duration of Contract (at Previous/Selling Club)
Market Size of Buying club (number of Transfers they perform and volume of transfers (\$))	Market Size of Buying club (number of Transfers they perform and volume of transfers (\$))
Google scores (Fame/ Marketability indicator)	Google scores (Fame/ Marketability indicator)
Goals (Strikers/Attackers)	Shooting, Dribbling, & Physicality (Strikers/Attackers)
Assists (Midfielders)	Dribbling, Passing, & Physicality (Midfielders)
Bookings (Defenders)	Defending & Physicality (Defenders)
Height (Defenders)	
Experience (Attackers, Midfielders, Defenders)	

Table 7.1: The Significant Factors of the Two Econometric Approaches of Chapters 3 and 4

### 7.3.2 Market Segmentation

It was Dobson, Gerrard and Howe (2000) who mentioned that there was some kind of segmentation among the divisions of the local leagues. In two chapters, we provided a framework for potential geographical segmentation (continental and country level) and across players' positions (Figure 7.1). Such findings have been achieved after extensive testing using the different pricing functions (packages, transfer fees, and salaries). Many characteristics were identified about different countries during testing, i.e. England could be an attractive hub for different clubs from different countries to send their players on loan for free to the most prestigious league (English Premier League) longing to boost market value and maximize profit when a player returns back from loan. Additionally, the market segment for attackers across all continents converges when we use the newly designed pricing packages.

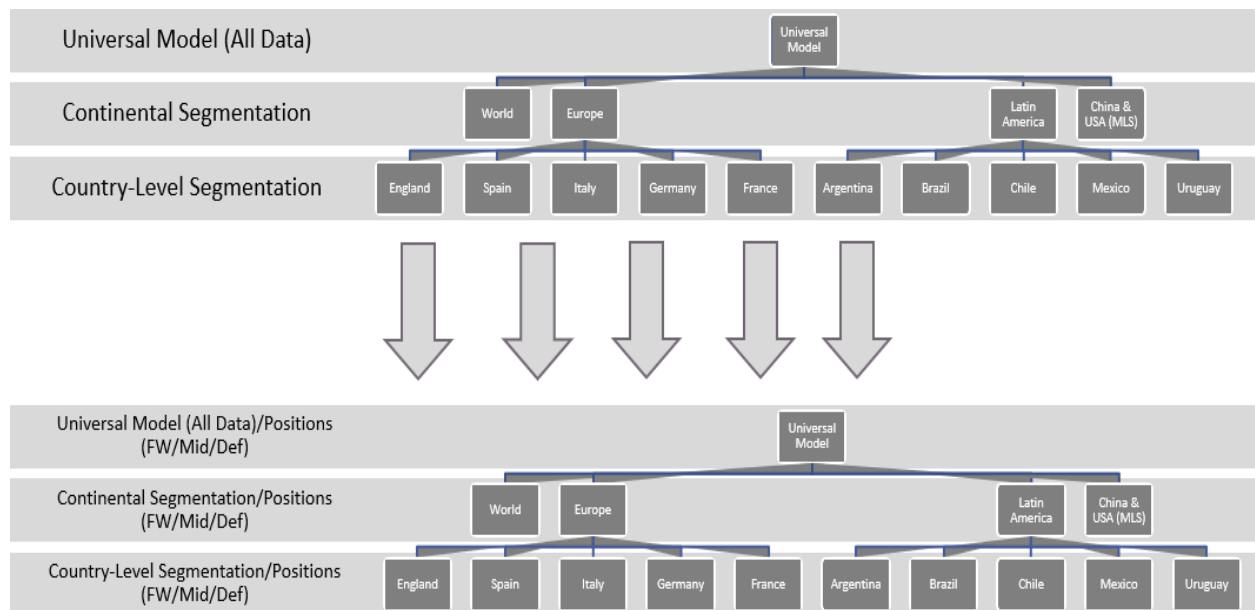


Figure 7.1: Processing Chart

At the same time, we could see sub-markets of two or three countries (e.g. Spain and France) when using the transfer fee as a pricing function, thanks to Chow test that helped to identify and confirm collective market segments. In general, the market segments of attackers and

defenders were adaptive to the model, unlike the midfielders' segment, which proved to be trickier. A larger data set could have helped to analyze further the complexities of this particular segment which is composed of offensive, defensive, and central midfielders. Achieving a consistent global market, using real and synthetic data was viable by lifting heteroskedasticity and applying Chow tests.

### **7.3.3 From Econometrics to Machine Learning**

Using three machine learning techniques was a tool to compare the findings of the econometric approach to those of the machine learning approach (Chapter 5). The flaws exposed by the econometric approaches are controlled in different ways by the machine learning techniques. Although the Generalized Linear Models (GLM), for example, helps to curb the variance and avoid overfitting, the results generated were a bit disappointing as such methodology could not investigate the importance of the factors that influence the transfer fees in the model. That led to another testing through Random Forest (RF) and Gradient Boosting Machines (GBM) that have generated good  $R^2$  of 0.82 and 0.8 respectively in the last testing phase (after validation and training). Still, the random forest algorithm could be favoured due to its slightly lower levels of the rooted-mean-square-error (RMSE) and the mean absolute error (MAE). What is interesting, as well, was the convergence of the top five important factors in both methodologies. The two methodologies can confirm what has been shown in previous chapters, using the econometric approach of the real and synthetic data, that the remaining duration of contract plays an unimportant role in determining both salaries and transfer fees. Players' position and nationality have, also, crucial special effect on transfer fees and salaries. The buying club and its purchasing activity have a highly important role to play in determining the price.



### **7.3.4 Football events' weight in Stock Market**

Furthermore, it was a significant challenge to check the effect of the transfers on the market level due to its financial importance. Clubs' budgets are between 60% - 90% occupied by transfers and salaries expenses (Andreff et al., 2008), thanks to the Financial Fair Play (FFP) regulation which is helping to regulate the market by controlling the balance between revenues and expenditures. With players' transfers that could exceed €200M (Neymar) and weekly wages that could exceed €500k (Lionel Messi), the financial risk of imbalanced budgets could increase and, consequently, affect future investment plans. Investors may consider investing in potential emerging clubs that could be profitable for them in the future. The effect of hiring the likes of Cristiano Ronaldo seems to have a positive effect on the share price of the Italian giants. Such a deal may attract investors. In general, the incoming transfers affected the market value of the clubs positively, especially undervalued transfers. Such findings fall into contradiction with Fotaki et al. (2009) who have shown that the share price is affected by the sales contracts instead. Such promising findings could be further investigated in future with a larger data set. We have also confirmed some previously proved events by Sholten and Peenstra (2012) like match surprises. A loss when a win is expected could harm the share price value with an unexpected drop in the share price of a football club, while an unexpected win can trigger a surge in the share price of the winning club.

### **7.4 Recommendations**

Many researchers believe that transfer fees and salaries have skyrocketed to unbearable levels. Top clubs that accumulate increasing profits can afford to purchase superstars at high prices, unlike smaller clubs. The winning prizes and augmented broadcasting rights favour the treasury of the winning clubs making them richer and leaving smaller clubs poorer. Consequently, the

appearance of competitive imbalance becomes a normal outcome in the long term. In order to combat the long-term consequences, we list a few recommendations that may help:

- Reviewing the progressive distribution of the prize money of the European competitions', Champions League and Europa League, winners by closing the gap between the two different competitions and within each competition.
- Ensure equal distribution of broadcasting rights within the premier divisions of the top five leagues in order to attain lower differences across the top five leagues in the world.
- increase the parachute money received by the secondary leagues to close the gap between clubs of the premier and the secondary divisions, taking into consideration the fears and concerns of Wilson, R., Ramchandani, G., & Plumley, D. (2018).
- Induce scientific market values for transfer fees and salaries based on economists and researchers annual market evaluations instead of proposing fixed caps so that a player cannot be registered through the FIFA Transfer Matching System if overvalued by illogical percentages. Clubs' over-spending could be taxed using a hierarchal scheme that may ensure a balanced and fair allocation of those taxes among most needy clubs.
- Clubs and regulators (FIFA, UEFA, continental and local Federations) should work together to support youth academies, technically and financially, to insure a long-term stability in clubs' financial statements.

## **7.5 Future Scopes**

This thesis is an elaborated framework which investigated the transfer market internal and external forces. A lot of the findings that have been achieved can integrate with other investigations by other researchers' findings to contribute to a better regularization of the industry. As the controversy goes on by FIFA, UEFA, and many other federations and regulators about reforming the transfer system, this piece of research, if supported by a more

diversified and cleaner dataset, could unlock a lot of uncertainty ruling the industry. Moreover, this research explained old and new variables influencing transfer fees and salaries. Still, the year-on-year growth of transfers' expenditure also had some other commercial reasons driven mainly by the increased broadcasting rights besides better sponsoring and global merchandising activities. This study has covered transfers lying within seasons 2007-2018 and stock market events and figures within 2000-2019. Thus, the probable consequences of the COVID-19, pandemic disease created by Coronavirus, hasn't been covered in the chapters of the thesis. Such pandemic that has triggered many fears about mysterious scenarios looming in the near future may affect the revenues of all stakeholders of the football industry and therefore the future projections of the transfers' activities of small and big clubs. Finally, one very important topic which deserves extra concentration is the transfer market of the women football market, an emerging topic of labour market inequality that needs to be addressed fairly by economists and regulators.

## 8 REFERENCES

- Addo, P.M.; Guegan, D.; Hassani, B. Credit Risk Analysis Using Machine and Deep Learning Models. *Risks* 2018, 6, 38
- Andreff, W., Aglietta, M., Drut, B., Bourse et Football. *Revue d'Economie Politique*, Dalloz, 2008, 118 (2), pp.255-96. [{halshs-00293907}](#)
- Amir, E. and Livne, G.,(2005), Accounting, Valuation and Duration of Football Player Contracts, *Journal of Business Finance & Accounting*, 32, (3-4), 549-586
- Antonioni, P., & Cubbin, J. (2000). The Bosman ruling and the emergence of a single market in soccer talent. *European Journal of Law and Economics*, 9(2), 157-173.
- Ardekani, A., Distinguin, I., Tarazi, A., (2019). Interbank network characteristics, monetary policy News” and sensitivity of bank stock returns. Working Paper. fahal-02384533f
- Ashton, J. K., Gerrard, B., Hudson, R., (2003) Economic impact of national sporting success: evidence from the London stock exchange, *Applied Economics Letters*, 10:12, 783-785, DOI: 10.1080/1350485032000126712.
- Bell, AR., Brooks, C., Matthews, D., Sutcliffe, C. (2012) Over the moon or sick as a parrot? The effects of football results on a club’s share price. *Applied Economics*.
- Bernile, G. and Lyandres, E. (2011). Understanding Investor Sentiment: The Case of Soccer. *Financial Management*, 40: 357-380.
- Big 5 – Transfer Window Analysis Winter 2019. FIFATMS.com. FIFA/TMS (2016).
- Bigsoccer, <http://www.bigsoccer.com>, accessed in February 2018
- Breiman, Leo. (2000). Some Infinity Theory for Predictors Ensembles. Technical Report. Berkeley: UC Berkeley, Vol. 577.
- Breiman, Leo. (2004). Consistency for a Sample Model of Random Forest. Technical Report 670. Berkeley: UC Berkeley, Vol. 670.

- Brown, G. W. and Hartzell, J. C. (2001) Market reaction to public information: the atypical case of the Boston Celtics, *Journal of Financial Economics*, 60, 333–70.
- Brown, S.J., Warner, J.B. (1980) Measuring security price performance. *Journal of*
- Brown, S.J., Warner, J.B. (1985) Using daily stock returns: The case of event studies. *Journal of financial economics*
- Calcio Finanza, <http://www.calciofinanza.com>, accessed in February and March 2018.
- Campbell, J.Y., Champbell, J.J., Campbell, J.W, Lo., A.W. (1997) *The econometrics of financial markets*. Princeton University Press.
- Carmichael, F. (2006). The Player Transfer System. In W. Andreff, & S. Szymanski, *Handbook of the Economics of Sports* (pp. 668 - 676). Cheltenham UK; Northampton MA, USA: Edward Elgar.
- Carmichael, F., & Thomas, D. (1993). Bargaining in the transfer market: theory and evidence. *Applied Economics*, 25(12), 1467-1476.
- Carmichael, F., Forrest, D. and Simmons, R. (1999), ‘The labour market in association football: who gets transferred and for how much?’, *Bulletin of Economic Research*, 51 (2), 125–50.
- Castellani, M., Pattitoni, P., et al. (2015) Abnormal returns of soccer teams: Reassessing the informational value of betting odds. *Journal of Sports Economics*.
- Castellano, J., Alvarez-Pastor, D., & Bradley, P. S. (2014). Evaluation of research using computerised tracking systems (amisco R and prozone R) to analyse physical performance in elite soccer: A systematic review. *Sports Medicine*, 44(5), 701–712.
- Cooper, B. and McHattie, A. (1997) *How to Invest in Sports Shares*, Batsford, London
- Das, S.R., Chen., M.Y., (2016) Soccer transfers, team efficiency and the sports cycle in the most valued European soccer leagues—have European soccer teams been efficient in trading players? *Applied Economics*.

Deloitte, <http://www.deloitte.com>, accessed in March 2018.

Demir, E., Danis, H., (2011) The effect of performance of soccer clubs on their stock prices:

Evidence from Turkey. *Emerging Markets Finance and Trade*

Dobson, S. and Gerrard, B. (1999). The determination of player transfer fees in english professional soccer. *Journal of Sport Management*, 13, pp. 259–79.

Dobson, S., Gerrard, B. and Howe, S. (2000). The determination of transfer fees in English nonleague football. *Applied Economics*, 32, pp. 1145–52.

Dobson, S. and Goddard, J. (2001), *The Economics of Football*, Cambridge: Cambridge University Press.

Dolleyi., JC (1933) Characteristics and procedure of common stock split-ups. *Harvard Business Review*

Edmans, A., Garcia, D., Norli., Ø (2014) Sports sentiment and stock returns. *The Journal of finance*

Edwards, M. (2017). “Red glory – Manchester United and me”. London: Michael O’Mara Books Limited.

Ericson, T. (2000). The Bosman case: effects of the abolition of the transfer fee. *Journal of Sports Economics*, 1(3), 203-218.

Feess, E., & Muehlheusser, G. (2003). Transfer fee regulations in European football. *European Economic Review*, 47(4), 645-668.

Feess, E., Gerfin, M., & Muehlheusser, G. (2010). The incentive effects of long-term contracts on performance-evidence from a natural experiment in European soccer. Technical Report, Mimeo: Berlin. financial economics.

Floros, C. (2014) Football and stock returns: New evidence. *Procedia Economics and Finance*.

- Fotaki, M., Markellos, R. N. and Mania, M. (2009) Human resources turnover as an asset acquisition, accumulation and divestiture process, Working Paper, Athens University of Economics and Business.
- Frick, B. (2006). Salary determination and the pay-performance relationship in professional soccer: Evidence from Germany. *Sports Economics After Fifty Years: Essays in Honour of Simon Rottenberg*. Oviedo: Ediciones de la Universidad de Oviedo, 125–146.
- Frick, B. (2007). “The football players’ labour market: Empirical evidence from the major European leagues”. *Scottish Journal of Political Economy*, 54, 422-446
- Frick, B. (2011). Performance, salaries, and contract length: empirical evidence from German soccer. *International Journal of Sport Finance*, 6(2), 87.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33:1.
- Friedman, Jerome (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*. *Journal of Statistical Software* 29: 1189-232.
- Garcia-del Barrio, P., & Pujol, F. (2007). Pay and performance in the Spanish soccer league: who gets the expected monopsony rents. Technical report, University of Navarra, Spain
- García del Barrio, P., and Szymanski, S. (2009). Goal! Profit maximization and win maximization in football leagues. *Review of Industrial Organization*, 34(1), 45-68. <https://doi.org/10.1007/s11151-009-9203-6>
- Gerrard, B. (2001), “A new approach to measuring player and team quality in professional team sports”, *European Sport Management Quarterly*, Vol. 1, No. 3, pp. 219-234.

- Gerrard, B. and Dobson, S. (2000), 'Testing for monopoly rents in the market for playing talent: evidence from English professional football', *Journal of Economic Studies*, 27 (3), 142–64.
- Google, <http://www.google.com>, accessed in June and July 2017
- Jason P. Berkowitz and Craig Depken. (2018). A rational asymmetric reaction to news: evidence from English football clubs. *Review of Quantitative Finance and Accounting*, 2018, vol. 51, issue 2, 347-374.
- Jensen, M. M., Grønbaek, K., Thomassen, N., Andersen, J., and Nielsen, J. (2014). Interactive football-training based on rebounders with hit position sensing and audio-visual feedback. *International Journal of Computer Science in Sport*, 13(1), 57–68.
- Jerome Geyer-Klingenberg, Markus Hang, Matthias Walter and Andreas Rathgeber. (2018). Do stock markets react to soccer games? A meta-regression analysis. *Applied Economics*, 2018, vol. 50, issue 19, 2171-2189.
- Ka Wai Terence Fung, Ender Demir, Chi Keung Lau and Kwok Ho Chan. (2015). Re-examining sports-sentiment hypothesis: Microeconomic evidences from Borsa Istanbul. *Journal of International Financial Markets, Institutions and Money*, 2015, vol. 34, issue C, 337-355.
- Kanyinda, A, Bouteiller, C, et Karyotis, C (2012). "Human Capital: Assessing the Financial Value of Football Players on the basis of Real Options Theory". *Journal "Investment Management and Financial Innovations"*, 9 (4).
- Kesenne S. (2002), "The monopsonistic player labour market in a win-maximising league", *European Sport Management Quarterly*, 2:3, 180-187.
- Kesenne, S. (2006). The win maximization model reconsidered : Flexible talent supply and efficiency wage. *Journal of Sports Economics*, 7, 416-427.  
doi:10.1177/1527002505279347



- Kesenne, S. (2007). The peculiar international economics of professional football in Europe. *Scottish Journal of Political Economy*, 54, 3, pp. 388–99.
- Lames, M., McGarry, T., Nebel, B., & Roemer, K. (2011). Computer science in sport special emphasis: Football (dagstuhl seminar 11271). *Dagstuhl Reports*, 1(7).
- MacKinlay, AC (1997) Event studies in economics and finance. *Journal of economic literature*.
- Matesanz, D., Holzmayer, F., Torgler, B., Schmidt, S. L., & Ortega, G. J. (2018). Transfer market activities and sportive performance in European first football leagues: A dynamic network approach. *PloS one*, 13(12), e0209362.
- Morrow, S. (1999). *The New Business of Football*, MacMillan Business, UK
- Nsolo, E., Lambrix, P., & Carlsson, N. (2018, September). Player valuation in European football. In *International Workshop on Machine Learning and Data Mining for Sports Analytics* (pp. 42-54). Springer, Cham.
- O'Donoghue, P. & Robinson, G. (2009). Validity of the prozone3 r player tracking system: A preliminary report. *International Journal of Computer Science in Sport*, 8(1), 37–53.
- Pain, M. A. and Harwood, C. G. (2004) Knowledge and perceptions of sport psychology within English soccer, *Journal of Sports Sciences*, 22, 813–26.
- Palomino, F., Renneboog, L. and Zhang, C. (2005) Stock price reactions to short-lived public information: the case of betting odds, ECGI, Working Paper No. 81
- Poli, R, Ravanel, L, Beson, R (2017). “How to Evaluate a football Player’s Transfer Value?”. CIES.
- Reilly, T. and Gilbourne, D. (2003) Science and football: a review of applied research in the football codes, *Journal of Sports Sciences*, 21, 693–705.
- Renneboog, L., Vanbrabant, P., (2000) *Share Price Reactions to Sporty Performances of Soccer Club Listed on the London Stock Exchange and the AIM*. Tilburg: Finance.

- Riley, B. and Witt, R. (1995), 'English league transfer prices: is there a racial dimension?', *Applied Economics Letters*, 2, 220–22.
- Ruijg, J. and Ophem, H. (2014). "Determinants of Football Transfers".
- Scarpello, V. & Theeke, H.A. (1989). 'Human resource accounting: a measured critique', *Journal of Accounting Literature*, 8, pp. 265--280.
- Scholtens, B, Peenstra, W. (2009) Scoring on the stock exchange? The effect of football matches on stock market returns: an event study. *Applied Economics*
- Semykina A. Woolridge J. R. 2010. Estimating Panel Data Models in the Presence of Endogeneity and Selection, *Journal of Econometrics* 157, August 2010, pp. 375-380.
- Szymanski, S. and Smith, R. (1997), 'The English football industry: profit, performance and industrial structure', *International Review of Applied Economics*, 11 (1), 135–53.
- Siegle, M., Stevens, T., & Lames, M. (2013). Design of an accuracy study for position detection in football. *Journal of Sports Sciences*, 31(2), 166–172.
- Simmons, R. (1997). Implications of the Bosman ruling for football transfer markets. *Economic Affairs*, 17, pp. 13–18
- Sportune, <http://www.sportune.fr>, accessed in February 2018
- Stadtman, Georg, (2006) Frequent News and Pure Signals: The Case of a Publicly Traded Football Club. *Scottish Journal of Political Economy*, Vol. 53, No. 4, pp. 485-504, September 2006.
- Szymanski, S. and Smith, R. (1997), 'The English football industry: profit, performance and industrial structure', *International Review of Applied Economics*, 11 (1), 135–53.
- Szymanski, S. & Kuypers, T. (1999). *Winners and Losers : The Business Strategy of Football*. Penguin, Harmondsworth.

- Tervio, M. (2004). Transfer fee regulations and player development. Mimeo, Haas School of Business, University of California, Berkeley
- Torgler, B., Schmidt, S. L., & Frey, B. S. (2006). Relative income position and performance: an empirical panel analysis.
- Torgler, B., Schmidt, S. L., & Frey, B. S. (2006). Relative income position and performance: an empirical panel analysis.
- TransferMarkt, <http://www.transfermarkt.co.uk>, accessed from January 2017 till March 2018.
- Vrooman, J. (1996), 'The baseball player's labor market reconsidered', *Southern Economic Journal*, 63, 339–60.
- Rob Wilson, Girish Ramchandani & Daniel Plumley (2018) Parachute Payments in English Football: Softening the Landing or Distorting the Balance?, *Journal of Global Sport Management*, 3:4, 351-368, DOI: [10.1080/24704067.2018.1441740](https://doi.org/10.1080/24704067.2018.1441740)
- Yaldo, L., & Shamir, L. (2017). Computational estimation of football player wages. *International Journal of Computer Science in Sport*, 16(1), 18-38.
- Zuber, RA., Yiu, P., Lamb, RP. (2005) Investor–fans? An examination of the performance of publicly traded English Premier League teams. *Applied Financial Economics*.

## 9 Appendices

### 9.1 Appendix 1: Tables and Figures for Chapter 3

Different Dependent Variables - Universal Model														
D. VARIABLES	lfee	lfee (lfee>0)	lfee (lfee>0) & (yr>13)	lfee (lfee>0) & (yr>13)	lfee (lfee>0) & (yr>13)	OP	OP (1 Var)	OP (2 Var)	SCP	SCP (1 Var)	SCP (2 Var)	SSP	SSP (1 Var)	SSP (2 Var)
<b>I. Variables</b>														
freetransfer	-4.625***					-0.560***	-0.911***	-0.819***	-0.272***	-0.529***	-0.438***	-0.0763***	-0.212***	-0.117***
tt_loan	-2.944***	-1.239***	-1.269***	-1.419***	-1.416***	-0.380***	-0.705***	-0.634***	-0.137***	-0.402***	-0.332***	0.0301	-0.106**	-0.0338
ZGOALS1	0.757***	0.121***	0.173***	0.164***	0.164***	0.288***	0.274***	0.227***	0.278***	0.271***	0.223***	0.241***	0.228***	0.179***
ZGOALS2	0.337***	0.136***	0.0783*	0.0364	0.0328	0.144***	0.134***	0.0741	0.164***	0.162***	0.102**	0.137***	0.137***	0.0751*
ZASSISTS1	0.962***	0.0389	-0.0213	-0.00133	-0.00198	0.194***	0.177***	0.103**	0.182***	0.179***	0.105**	0.188***	0.193***	0.117***
ZASSISTS2	0.398***	0.0602**	0.0722*	0.0294	0.0269	0.201***	0.106**	0.0604	0.178***	0.0784	0.033	0.159***	0.075	0.0281
ZEXP1	1.218***	0.442***	0.362***	0.358***	0.359***	0.256***	0.168***	0.160***	0.219***	0.140***	0.133***	0.160***	0.0960**	0.0880**
ZEXP2	0.598***	0.0697***	0.137***	0.126***	0.127***	0.257***	0.246***	0.157***	0.292***	0.276***	0.187***	0.281***	0.264***	0.172***
ZBAD1	0.486***	-0.0629	-0.0786	-0.0903	-0.093	0.235***	0.146*	0.0174	0.295***	0.186**	0.0577	0.259***	0.180**	0.0478
ZBAD2	0.252***	0.0490*	0.0432	0.0493	0.0494	0.148***	0.114**	0.0574	0.161***	0.129**	0.0731	0.131***	0.0888*	0.0307
<b>lRemDurDays</b>				<b>0.0994***</b>	<b>0.100***</b>		<b>0.0747***</b>	<b>0.0401***</b>		<b>0.102***</b>	<b>0.0668***</b>		<b>0.0815***</b>	<b>0.0458***</b>
<b>lsalarp</b>					<b>0.00965*</b>			<b>0.351***</b>			<b>0.352***</b>			<b>0.363***</b>
minoraiola	-0.575	-0.136	-0.192	-0.196	-0.202	-0.0698	0.0409	-0.0795	-0.116	0.0227	-0.098	-0.091	0.0734	-0.0511
mediabase	-0.137	0.754**	0.725**	0.357	0.363	0.954***	0.36	0.234	1.097***	0.451	0.325	1.065***	0.514	0.384
gestifute	0.256	0.452***	0.375***	0.16	0.172	0.305*	0.122	0.177	0.145	-0.0533	0.00131	-0.033	-0.227	-0.171
lage	38.70***	30.89***	29.77***	28.68***	27.49***	48.35***	48.28***	31.09***	-24.73***	-27.52***	-44.75***	-29.48***	-32.85***	-50.63***
lage2	-6.672***	-4.874***	-4.683***	-4.544***	-4.367***	-7.211***	-7.138***	-4.689***	2.892***	3.408***	5.863***	3.620***	4.208***	6.741***
ldur_hyp	0.732***	0.223***	0.260***	0.206***	0.206***	1.080***	1.015***	1.034***	0.170***	0.125***	0.144***	0.137***	0.0966***	0.117***
lheight	1.170***	1.032**	1.887***	2.006***	2.062***	1.710***	0.897	0.663	1.475***	0.675	0.441	1.211***	0.238	-0.00366
lgoo	-0.00764	0.0267***	0.0194***	0.0388***	0.0386***	0.0219***	0.0258***	0.0214***	0.0217***	0.0294***	0.0250***	0.0165***	0.0214***	0.0169***
lgt	0.0975***	0.0148**	0.0173*	0.0135	0.0121	0.0503***	0.00345	-0.0012	0.0488***	0.00598	0.00132	0.0471***	0.00714	0.00233
lco_cuclu	0.255***	-0.263***	-0.287***	-0.280***	-0.281***	0.0810***	0.0852***	0.0698***	0.0808***	0.0834***	0.0679***	0.0754***	0.0944***	0.0784***
lco_preclu	0.0734***	-0.0319***	-0.0472***	-0.0524**	-0.0539**	0.0559***	0.0611***	0.0205	0.0606***	0.0737***	0.0330**	0.0569***	0.0723***	0.0303**
ltot_preclu	0.0652***	0.189***	0.209***	0.199***	0.196***	0.0408***	0.0498***	0.0348***	0.0350***	0.0426***	0.0275***	0.0329***	0.0415***	0.0259***
ltot_cuclu	0.107***	0.424***	0.456***	0.483***	0.482***	0.0484***	0.0572***	0.0398***	0.0412***	0.0489***	0.0315***	0.0377***	0.0442***	0.0262***
footer_both	-0.356***	-0.389***	-0.367***	-0.214	-0.217	-0.145**	-0.0741	-0.0809	-0.0689	-0.0346	-0.0414	-0.0938	-0.067	-0.0741
footer_left	0.148**	0.332***	0.292***	0.187	0.191	0.0673	-0.0277	0.0424	0.0252	-0.0298	0.0404	0.0468	-0.00015	0.0723

Different Dependent Variables - Universal Model														
D. VARIABLES	lfee	lfee (lfee>0)	lfee (lfee>0) & (yr>13)	lfee (lfee>0) & (yr>13)	lfee (lfee>0) & (yr>13)	OP	OP (1 Var)	OP (2 Var)	SCP	SCP (1 Var)	SCP (2 Var)	SSP	SSP (1 Var)	SSP (2 Var)
footer_right	0.216***	0.333***	0.306***	0.243**	0.248**	0.0305	-0.0123	0.0555	-0.00928	-0.022	0.0459	0.00147	0.000894	0.071
nat_cont_africa	0.139***	0.104***	0.0994***	0.143***	0.134***	0.266***	0.253***	0.167***	0.235***	0.218***	0.131***	0.223***	0.205***	0.115***
nat_cont_asia	0.174**	0.0261	0.145**	0.114	0.135*	0.294***	0.208**	0.117	0.301***	0.224**	0.133	0.298***	0.220**	0.125
nat_cont_australia	-0.552***	-0.124	-0.158	-0.0748	-0.0824	-0.339***	-0.332**	-0.287**	-0.329***	-0.337**	-0.292**	-0.326***	-0.319**	-0.273**
nat_cont_europe	0.219***	-0.122***	-0.190***	-0.118***	-0.128***	0.103***	0.0506	0.0449	0.122***	0.0818**	0.0761**	0.112***	0.0733*	0.0674*
nat_cont_north_america	-1.182***	0.240***	0.325***	0.0769	0.0628	-0.207***	-0.282***	-0.220***	-0.172***	-0.237***	-0.176***	-0.134***	-0.197***	-0.133**
nat_cont_south_america	-0.473***	0.434***	0.363***	0.355***	0.354***	0.211***	0.218***	0.174***	0.237***	0.212***	0.168***	0.233***	0.181***	0.136***
pos_forward	-0.121**	0.408***	0.509***	0.478***	0.481***	0.0975**	0.145**	0.0795	0.0844**	0.133**	0.0665	0.0899**	0.131**	0.0632
pos_def	-0.202***	0.0925**	0.144**	0.0626	0.0655	-0.00784	0.0605	0.00738	0.00475	0.0746	0.0214	0.0124	0.0695	0.0146
pos_mid	-0.229***	0.268***	0.338***	0.281***	0.285***	-0.0174	0.031	-0.0183	-0.0117	0.0368	-0.0126	-0.00531	0.0228	-0.0282
y2007x	0.801***	1.254***												
y2008x	0.573***	1.236***												
y2009x	0.326**	1.194***												
y2010x	0.0467	1.063***												
y2011x	-0.0339	0.998***												
y2012x	-0.574***	1.137***												
y2013x	-0.209	1.098***	1.253***											
y2014x	-0.163	1.273***	1.465***	-0.0357	-0.036	-0.0945**	-0.128*	-0.234***	-0.138***	-0.154**	-0.259***	-0.154***	-0.162**	-0.272***
y2015x	-0.105	1.286***	1.467***	0.0678	0.0663	-0.0285	-0.0966	-0.112*	-0.0878**	-0.149**	-0.164***	-0.100**	-0.150**	-0.165***
y2016x	-0.00582	1.414***	1.597***	0.255**	0.253**	-0.538***	-0.733***	-0.728***	-0.617***	-0.795***	-0.790***	-0.684***	-0.867***	-0.862***
y2017x	-0.0865	1.527***	1.722***	0.369***	0.369***	0.0718*	0.00649	0.122**	0.0189	-0.0109	0.105*	-0.00527	-0.0247	0.0948*
y2018x	-0.338**	1.692***	1.867***	0.454***	0.451***	0.0415	-0.08	-0.138**	0.00293	-0.11	-0.168***	-0.0122	-0.107	-0.167***
Constant	-61.93***	-49.81***	-50.11***	-47.41***	-45.44***	-78.97***	-78.32***	-49.15***	59.30***	63.61***	92.85***	67.59***	73.14***	103.3***
Chi2	31519.22	203.47	117.5	109.41	110.01	84.48	120.51	245.32	88.9	51.73	193.64	41.9	16.1	137.61
Prob> Chi2	0	0	0	0	0	0	0	0	0	0	0	0	0.0001	0
Observations	84,302	11,162	4,990	2,917	2917	15,454	6,794	6794	15,454	6,794	6794	15,454	6,794	6794
R-squared	0.428	0.667	0.689	0.721	0.721	0.643	0.709	0.763	0.528	0.542	0.632	0.506	0.511	0.611
Adj. R-Squared	0.4276	0.6658	0.6869	0.7167	0.717	0.6423	0.707	0.7615	0.5269	0.539	0.6298	0.5045	0.5079	0.6083

Table AP 1.1: Basic Tests using All available Dependent Variables. Sample: All Data

## CHOW TESTS

The Chow test is  $F(k, N_1 + N_2 - 2 * k) = F(38, 1002)$  = a number below [1.45 & 1.394] for 5% critical value

$$\frac{728.515845 - (597.689279 + 93.895505)}{38} = \frac{597.689279 + 93.895505}{860 + 218 - 2 * 38}$$

Chow test is 1.400. In this case it is acceptable for the 5% critical. ONE market detected== En+It+Sp. England, Italy, and Spain accommodate one unified pricing model.

The Chow test is  $F(k, N_1 + N_2 - 2 * k) = F(38, 1260)$  = a number below [ 1.29513 & 1.34187] for 10% critical value

$$\frac{873.205988 - (728.515845 + 115.079278)}{38} = \frac{728.515845 + 115.079278}{1078 + 258 - 2 * 38}$$

Chow test is 1.16386. In this case it is acceptable for the 10% critical value. ONE market detected== En+It+Sp+Gr. England, Germany, Italy, and Spain accommodate one unified pricing model.

The Chow test is  $F(k, N_1 + N_2 - 2 * k) = F(38, 1520)$  = a number below [1.29513 & 1.34187] for 10% critical value

$$\frac{972.588594 - (873.205988 + 76.0666528)}{38} = \frac{873.205988 + 76.0666528}{1336 + 260 - 2 * 38}$$

Chow test is 0.9825. In this case it is acceptable for the 10% critical value. ONE market detected== En+It+Sp+Gr. England, France, Germany, Italy, and Spain accommodate one unified pricing model.

Latin America - Across Positions									
Variables	Forwards			Midfielders			Defenders		
	SCP	SCP	SCP	SCP	SCP	SCP	SCP	SCP	SCP
	(all)	1 add, Var	2 Add, Var	(all)	1 add, Var	2 Add, Var	(all)	1 add, Var	2 Add, Var
freetransfer	-0.204**	-0.415**	-0.180***	0.0199	-0.455**	-0.342*	-0.0368	-0.239	-0.139
tt_loan	-0.158*	-	-0.217***	0.00534	-0.387*	-0.323*	-0.115	-0.349	-0.365*
ZGOALS1	0.469***	0.887***	0.11	0.512**	-0.391	-0.152	0.491*	0.638	0.409
ZGOALS2	0.534***	0.2	-0.0447	0.430**	0.226	0.025	-0.00462	-0.289	-0.345
ZASSISTS1	0.318	0.041	0.0664	0.801***	0.621*	0.718**	0.255	0.532	0.142
ZASSISTS2	0.716***	0.352	0.0571	0.0625	0.32	0.186	-0.107	-0.477	-0.804*
ZEXP1	0.00888	-0.211	-0.0919	0.0408	0.199	0.108	0.0726	-0.413	-0.0776
ZEXP2	-0.197	0.152	0.00192	0.321**	0.546*	0.44	0.802***	0.976***	0.793***
ZBAD1	-0.631	-0.744	-0.0701	-0.531	0.183	-0.0343	-0.752	-1.305*	-0.806
ZBAD2	0.508			0.25	-0.852	-0.591	-0.0554	-0.16	-0.303
lRemDurDays		<b>0.0649</b>	<b>0.0510***</b>		<b>0.364***</b>	<b>0.283***</b>		<b>0.309***</b>	<b>0.252***</b>
lsalarp			<b>0.957***</b>			<b>0.264***</b>			<b>0.347***</b>
minoraiola				-1.075**			-	-	-
gestifute				-1.645			1.499	2.724*	1.469
lage	-14.71**	7.362	-63.31***	-57.28***	-61.06***	-62.03***	-43.83***	-	-
lage2	1.303	-1.853	8.574***	7.598***	8.322**	8.437***	5.526***	9.089***	10.92***
ldur_hyp	0.137***	0.0171	0.138***	0.210***	0.0565	0.0796	0.137**	0.00867	0.0279
lheight	-0.643	1.192	0.584	7.898***	4.22	2.909	4.806***	2.143	1.265
lgoos	0.00927	-0.0262	0.000201	0.0105	0.0498	0.04	0.0147	0.0720**	0.0537**
lgt	0.0710***	0.0892**	-0.011	0.0645***	0.00729	-0.0281	0.107***	0.0523	0.0444
lco_cuclu	-0.0197	0.138	0.0576*	0.0312	-0.0779	0.0597	0.0309	0.0175	0.0049
lco_preclu	0.0433	0.0636	-0.0191	0.0423	-0.12	-0.097	-0.011	-0.0444	-0.107
ltot_preclu	0.0124**	0.00808	0.00437	0.0200***	0.0338	0.0334*	0.0255***	0.0231	0.0267
ltot_cuclu	0.0387***	0.0198	0.00281	0.0211***	0.00532	-0.0124	0.0173***	0.0259	0.00871
footer_both	0.266	-1.331**	-0.153	0.135	0.19	0.198	0.42	0.819	0.960*
footer_left	-0.0968	0.154	0.112	0.227	-0.244	-0.415	-0.118	-0.00098	0.189
footer_right	-0.159	0.582	0.228**	0.0491	-0.56	-0.484	-0.114		
nat_cont_africa	-0.259	-0.43	-0.0805	-0.556	0.771	0.466	-0.212		
nat_cont_asia	-0.584*	-1.245**	-0.112						
nat_cont_australia	0.265***	0.236	-0.0301						
nat_cont_europe	-0.440***	0.348	0.0957	0.238***	0.321*	0.284*	0.14	0.551***	0.467***
nat_cont_north_america	-0.313*	0.43	0.0535	0.000148	-0.0218	-0.0534	0.155	0.263	0.377
nat_cont_south_america				0.0101	-0.126	-0.0972	0.252	0.213	0.273
y2013x	0.525**								
y2014x	0.116	0.265	0.164	-0.395***	-0.605	-0.273	0.0287	0.332	0.0311
y2015x	0.261	0.0527	0.038	-0.320***	-0.664	-0.38	-0.0886	0.147	-0.219
y2016x	-0.082	-0.276	0.0214	-0.539***	-1.346***	-1.078***	-0.332***	-0.208	-0.606**
y2017x	0.235	-0.235	0.082	-0.173	-1.036**	-0.706*	-0.0904	0.125	-0.295
y2018x		0.0229	0.17	0.0311	-0.506	-0.49	-0.184	0.0913	-0.332
constant	47.45***	6.693	122.1***	109.4***	119.1***	120.6***	90.97***	123.5***	145.2***
Chi2	<b>9.2</b>	<b>11.75</b>	<b>14.81</b>	<b>0.08</b>	<b>0.14</b>	<b>6.28</b>	<b>2.59</b>	<b>7.75</b>	<b>10.97</b>
Prob>Chi2	<b>0.0024</b>	<b>0.0006</b>	<b>0.0001</b>	<b>0.7717</b>	<b>0.7086</b>	<b>0.0122</b>	<b>0.1075</b>	<b>0.0054</b>	<b>0.0009</b>
Observations	<b>1,562</b>	<b>391</b>	<b>391</b>	<b>1,353</b>	<b>259</b>	<b>259</b>	<b>1,358</b>	<b>307</b>	<b>307</b>
R-squared	<b>0.442</b>	<b>0.453</b>	<b>0.958</b>	<b>0.5</b>	<b>0.586</b>	<b>0.654</b>	<b>0.456</b>	<b>0.426</b>	<b>0.553</b>
Adj R-Squared	<b>0.4302</b>	<b>0.4022</b>	<b>0.9543</b>	<b>0.4871</b>	<b>0.5248</b>	<b>0.6012</b>	<b>0.4426</b>	<b>0.3594</b>	<b>0.4986</b>

Table AP 1.2: "Latin American Model": Tests across Positions

<b>Goals (ZGOALS1)</b>	<b>Goals - Previous 2 (ZGOALS2)</b>
Goals - Previous CL Goals - Previous CL Penalty Goals - Previous CL Qualifications Goals - Previous CL Qualifications Penalty Goals - Previous UCL Goals - Previous UCL Penalty Goals - Previous UCL Qualifications Goals - Previous UCL Qualifications Penalty Goals - Previous International Goals - Previous Local Competition Goals - Previous Local Competition Penalty Goals - Previous International Goals - Previous	Goals - Previous - 2 CL Goals - Previous - 2 CL Penalty Goals - Previous - 2 CL Qualifications Goals - Previous - 2 CL Qualifications Penalty Goals - Previous - 2 UCL Goals - Previous - 2 UCL Penalty Goals - Previous - 2 UCL Qualifications Goals - Previous - 2 UCL Qualifications Penalty Goals - Previous - 2 International Goals - Previous - 2 Local Competition Goals - Previous - 2 Local Competition Penalty Goals - Previous - 2 International Goals (Previous - 2)
<b>Assists (ZASSISTS1)</b>	<b>Assists - Previous 2 (ZASSISTS2)</b>
Assists - Previous CL Assists - Previous CL Qualifications Assists - Previous UCL Assists - Previous UCL Qualifications Assists - Previous Local Competition Assists - Previous International Assists - Previous	Assists (Previous - 2) CL Assists - Previous - 2 CL Qualifications Assists - Previous - 2 UCL Assists - Previous - 2 UCL Qualifications Assists - Previous - 2 Local Competition Assists - Previous - 2 International Assists (Previous - 2)
<b>Negative Characteristics - Previous (ZBAD1)</b>	<b>Negative Characteristics -Previous 2 (ZBAD2)</b>
CL Own Goals - Previous CL Red Cards - Previous CL Yellow cards - Previous CL Yellow/Red cards - Previous UCL Own Goals - Previous UCL Red Cards - Previous UCL Yellow cards - Previous UCL Yellow/Red cards - Previous UCL Qualifications Own Goals - Previous UCL Qualifications Red Cards - Previous UCL Qualifications Yellow Cards - Previous UCL Qualifications Yellow/Red Cards - Previous Local Competition Own Goals - Previous Local Competition Red Cards - Previous Local Competition Yellow Cards - Previous Local Competition Yellow/Red Cards - Previous	CL Own Goals - Previous - 2 CL Red Cards - Previous - 2 CL Yellow cards - Previous - 2 CL Yellow/Red cards - Previous - 2 UCL Own Goals - Previous - 2 UCL Red Cards - Previous - 2 UCL Yellow cards - Previous - 2 UCL Yellow/Red cards - Previous - 2 UCL Qualifications Own Goals - Previous - 2 UCL Qualifications Red Cards - Previous - 2 UCL Qualifications Yellow Cards - Previous - 2 UCL Qualifications Yellow/Red Cards - Previous - 2 Local Competition Own Goals - Previous - 2 Local Competition Red Cards - Previous - 2 Local Competition Yellow Cards - Previous - 2 Local Competition Yellow/Red Cards - Previous - 2
<b>Experience Previous (ZEXP1)</b>	<b>Experience Previous 2 (ZEXP2)</b>
CL App (Starting 11) - Previous CL App (Substituted On) - Previous CL App (Substituted off) - Previous	CL App (Starting 11) - Previous - 2 CL App (Substituted on) - Previous - 2 CL App (Substituted off) - Previous - 2



UCL Qualifications App (Starting 11) - Previous	UCL Qualifications App (Starting 11) - Previous - 2
UCL Qualifications App (Substituted on) - Previous	UCL Qualifications App (Substituted on) - Previous - 2
UCL Qualifications App (Substituted off) - Previous	UCL Qualifications App (Substituted off) - Previous - 2
UCL App (Starting 11) - Previous	UCL (Starting 11) - Previous - 2
UCL App (Substituted on) - Previous	UCL (Substituted on) - Previous - 2
UCL App (Substituted off) - Previous	UCL (Substituted off) - Previous - 2
UCL (minutes played) - Previous	UCL (minutes played) - Previous - 2
Local Competition App. (Starting 11) - Previous	Local Competition App. (Starting 11) - Previous - 2
Local Competition App. (Substituted on) - Previous	Local Competition App. (Substituted on) - Previous - 2
Local Competition App. (Substituted off) - Previous	Local Competition App. (Substituted off) - Previous - 2
International App - Previous	International App - Previous - 2
International (minutes played) - Previous	International (minutes played) - Previous - 2

Table AP 1.3: Independent Variables Used to form the Index Variables

I. Variables	Description
freetransfer	Dummy for players moving after finishing previous contract
tt_loan	Dummy for players moving on loan
ZGOALS1	Aggregate (after regression) number of goals in previous first year
ZGOALS2	Aggregate (after regression) number of goals in previous second year
ZASSISTS1	Aggregate (after regression) number of assists in previous first year
ZASSISTS2	Aggregate (after regression) number of assists in previous second year
ZEXP1	Aggregate (after regression) experience factor in previous second year
ZEXP2	Aggregate (after regression) experience in previous second year
ZBAD1	Aggregate (after regression) number of bookings in previous second year
ZBAD2	Aggregate (after regression) number of bookings in previous second year
lRemDurDays	ln (logarithm) of the remaining duration of previous contract at the time of current contract
lsalarp	ln (logarithm) of the previous salary at previous club during previous contract
minoraiola	Dummy for players having minoraiola as their agent
mediabase	Dummy for players having mediabase as their agent
gestifute	Dummy for players having gestifute as their agent
lage	ln (logarithm) of players age
lage2	ln (logarithm) of players age squared (lage squared)
lagem	ln (logarithm) of players whose age is below 24 years old
lagem2	ln (logarithm) of players whose age is below 24 years old - squared (lagem squared)
lagep	ln (logarithm) of players whose age is above 24 years old (lagep)
lagep2	ln (logarithm) of players age is above 24 years old - squared (lagep squared)
ldur	ln (logarithm) of duration of contract
ldur_hyp	ln (logarithm) of hypothetical duration of contract
lheight	ln (logarithm) of player's height
lgoo	ln (logarithm) of player's search frequency on google
lgt	ln (logarithm) of player's search frequency as a google trend
lco_cuclu	ln (logarithm) of number of transfers done by the current/buying club 2007/2008-2018/2019
lco_preclu	ln (logarithm) of number of transfers done by the previous/selling club 2007/2008-2018/2019
ltot_preclu	ln (logarithm) of the volume of transfers (£) done by the current/buying club 07/08-18/19
ltot_cuclu	ln (logarithm) of the volume of transfers (£) done by the previous/selling club 07/08-18/19
footer_both	Dummy for players playing in both feet
footer_left	Dummy of players playing in left foot
footer_right	Dummy of the players playing in right foot
nat_cont_africa	Dummy of players of African origin
nat_cont_asia	Dummy of players of Asian origin
nat_cont_australia	Dummy of players of Australian origin
nat_cont_europe	Dummy of players of European origin

nat_cont_north_america	Dummy of players of North American origin
nat_cont_south_america	Dummy of players of South American origin
pos_forward	Dummy of forward players (Strikers)
pos_def	Dummy of midfield players (Midfielders)
pos_mid	Dummy of defensive players (Defenders)
y2007x	Yearly Dummy for transfers of the 2007/2008 Two transfers windows
y2008x	Yearly Dummy for transfers of the 2008/2009 Two transfers windows
y2009x	Yearly Dummy for transfers of the 2009/2010 Two transfers windows
y2010x	Yearly Dummy for transfers of the 2010/2011 Two transfers windows
y2011x	Yearly Dummy for transfers of the 2011/2012 Two transfers windows
y2012x	Yearly Dummy for transfers of the 2012/2013 Two transfers windows
y2013x	Yearly Dummy for transfers of the 2013/2014 Two transfers windows
y2014x	Yearly Dummy for transfers of the 2014/2015 Two transfers windows
y2015x	Yearly Dummy for transfers of the 2015/2016 Two transfers windows
y2016x	Yearly Dummy for transfers of the 2016/2017 Two transfers windows
y2017x	Yearly Dummy for transfers of the 2017/2018 Two transfers windows
y2018x	Yearly Dummy for transfers of the 2018/2019 Two transfers windows
<b>CL (Not Variable)</b>	<b>Champions League (Continental; S. American, Asian,..)</b>
<b>UCL (Not Variable)</b>	<b>UEFA Champions League</b>
<b>Previous Season (Not Variable)</b>	<b>the season preceeding the date of contract</b>
<b>Second Previous Season (Not Variable)</b>	<b>Two Seasons before the contract date</b>
lgoalsprevious	ln (logarithm) of total number of Goals - in Previous season
lclgoalsprevious	ln (logarithm) of CL Goals - in Previous season
linternationalgoalsprevious	ln (logarithm) of International Goals (with National Team) - in Previous season
lclpenaltygoalsprevious	ln (logarithm) of CL Penalty Goals - in Previous season
lclqualificationsgoalsprevious	ln (logarithm) of CL Qualifications' Goals - in Previous season
lclqualificationspenaltygoalsprevious	ln (logarithm) of CL qualifications penalty Goals - in Previous season
llocalcompetitiongoalsprevious	ln (logarithm) of local competitions Goals - in Previous season
llocalcompetitionpenaltygoalsprevious	ln (logarithm) of the local competition penalty Goals - in Previous season
lEUclgoalsprevious	ln (logarithm) of UCL Goals - in Previousseason
lEUclqualificationsgoalsprevious	ln (logarithm) of UCL Penalty Goals - in Previous season
lEUclqualificationspenaltygoalsprevious	ln (logarithm) of UCL qualifications penalty Goals - in Previous season

lgoalsprevious2	In (logarithm) of total number of Goals - in second Previous season
lclgoalsprevious2	In (logarithm) of CL Goals - in second Previous season
linternationalgoalsprevious2	In (logarithm) of International Goals (with National Team) - in second Previous season
lclpenaltygoalsprevious2	In (logarithm) of CL Penalty Goals -in second Previous season
lclqualificationsgoalsprevious2	In (logarithm) of CL Qualifications' Goals - in second Previous season
lclqualificationspenaltygoalsprevious2	In (logarithm) of CLqualifications penalty Goals - in second Previous season
llcalcompetitiongoalsprevious2	In (logarithm) of local competitions Goals - in second Previous season
llocalcompetitionpenaltygoalsprevious2	In (logarithm) of local competitions penalty Goals - in second Previous season
IEUclgoalsprevious2	In (logarithm) of UCL Goals - in Previous season
IEUclqualificationsgoalsprevious2	In (logarithm) of UCL Penalty Goals - in Previous season
IEUclqualificationspenaltygoalsprevious2	In (logarithm) of UCL qualifications penalty Goals - in Previous season
lassistsprevious	In (logarithm) of total number of Assists - in Previous season
lclassistsprevious	In (logarithm) of total number of CL Assists - in Previous season
lclqualificationsassistsprevious	In (logarithm) of total number of CL Qualifications Assists - in Previous season
llocalcompetitionassistsprevious	In (logarithm) of total number of Assists in local competitions - in Previous season
linternationalassistsprevious	In (logarithm) of total number of Assists made with national team- in Previous season
IEUclqualificationsassistsprevious	In (logarithm) of total number of Assists in UCL qualification League - in Previous season
IEUclassistsprevious	In (logarithm) of number of UCL Assists - in Previous season
lassistsprevious2	In (logarithm) of total number of Assists - in second Previous season
lclassistsprevious2	In (logarithm) of number of CL Assists - in second Previous season
lclqualificationsassistsprevious2	In (logarithm) of number of CL Qualifications Assists - in second Previous season
llocalcompetitionassistsprevious2	In (logarithm) of total number of Assists in local competitions - in second Previous season
linternationalassistsprevious2	In (logarithm) of total number of Assists made with national team- in second Previous season
IEUclqualificationsassistsprevious2	In (logarithm) of number of Assists in UCL qualification League - in second Previous season
IEUclassistsprevious2	In (logarithm) of number of UCL Assists - in second Previous season
lclowngoalsprevious	In (logarithm) of # of own goals in CL - Previous season
lclredcardsprevious	In (logarithm) of # of red cards in CL - Previous season
lclyellowcardsprevious	In (logarithm) of # of yellow cards in CL - Previous season
lclyellowredcardsprevious	In (logarithm) of # of yellow Or red cards in CL - Previous season
lclqualificationsowngoalsprevious	In (logarithm) of # of own goals in CL qualifications - Previous season
lclqualificationsredcardsprevious	In (logarithm) of # of red cards in CL qualifications - Previous season

lclqualificationsyellowcardsprev	In (logarithm) of # of yellow cards in CL qualifications - Previous season
lclqualificationsyellowredcardsp	In (logarithm) of # of yellow Or red cards in CL qualifications- Previous season
llocalcompetitionowngoalspreviou	In (logarithm) of # of own goals in local competitions - Previous season
llocalcompetitionredcardspreviou	In (logarithm) of # of red cards in local competitions - Previous season
llocalcompetitionyellowcardsprev	In (logarithm) of # of yellow cards in local competitions - Previous season
llocalcompetitionyellowredcardsp	In (logarithm) of # of yellow Or red cards in local competitions - Previous season
IEUclowngoalsprevious	In (logarithm) of # of own goals in UCL - Previous season
IEUclredcardsprevious	In (logarithm) of # of red cards in UCL - Previous season
IEUclyellowcardsprevious	In (logarithm) of # of yellow cards in UCL - Previous season
IEUclyellowredcardsprevious	In (logarithm) of # of yellow Or red cards in UCL - Previous season
IEUclqualificationsowngoalspr	In (logarithm) of # of own goals in UCL qualifications - Previous season
IEUclqualificationsredcardspre	In (logarithm) of # of red cards in UCL qualifications - Previous season
IEUclqualificationsyellowcards	In (logarithm) of # of yellow cards in UCL qualifications - Previous season
IEUclqualificationsyellowredca	In (logarithm) of # of yellow Or red cards in UCL qualifications- Previous season
lclowngoalsprevious2	In (logarithm) of # of own goals in CL - Second Previous season
lclredcardsprevious2	In (logarithm) of # of red cards in CL - Second Previous season
lclyellowcardsprevious2	In (logarithm) of # of yellow cards in CL - Second Previous season
lclyellowredcardsprevious2	In (logarithm) of # of yellow Or red cards in CL - Second Previous season
lclqualificationsowngoalspr2	In (logarithm) of # of own goals in CL qualifications - Second Previous season
lclqualificationsredcardsp2	In (logarithm) of # of red cards in CL qualifications - Second Previous season
lclqualificationsyellowcard2	In (logarithm) of # of yellow cards in CL qualifications - Second Previous season
lclqualificationsyellowredcard2	In (logarithm) of # of yellow Or red cards in CL qualifications- Second Previous season
llocalcompetitionowngoalspr2	In (logarithm) of # of own goals in local competitions - Second Previous season
llocalcompetitionredcardsp2	In (logarithm) of # of red cards in local competitions - Second Previous season
llocalcompetitionyellowcard2	In (logarithm) of # of yellow cards in local competitions - Second Previous season
llocalcompetitionyellowred2	In (logarithm) of # of yellow Or red cards in local competitions - Second Previous season
IEUclowngoalsprevious2	In (logarithm) of # of own goals in UCL - Second Previous season
IEUclredcardsprevious2	In (logarithm) of # of red cards in UCL - Second Previous season
IEUclyellowcardsprevious2	In (logarithm) of # of yellow cards in UCL - Second Previous season
IEUclyellowredcardsprevious2	In (logarithm) of # of yellow Or red cards in UCL - Second Previous season

IEUclqualificationsowngoalspr2	In (logarithm) of # of own goals in UCL qualifications - Second Previous season
IEUclqualificationsredcardspre2	In (logarithm) of # of red cards in UCL qualifications - Second Previous season
IEUclqualificationsyellowcards2	In (logarithm) of # of yellow cards in UCL qualifications - Second Previous season
IEUclqualificationsyellowredca2	In (logarithm) of # of yellow Or red cards in UCL qualifications- Second Previous season
lclappstarting11previous	In (logarithm) of # of starting lineup in CL - Previous season
lclappsubstitutedonprev	In (logarithm) of # of times substituted ON in CL - Previous season
lclappsubstitutedoffprev	In (logarithm) of # of times substituted Off in CL - Previous season
lclqualificationsappstarti	In (logarithm) of # of starting lineup in CL qualifications- Previous season
lclqualificationsappsubst	In (logarithm) of # of times substituted ON in CL qualifications - Previous season
lclqualificationsappsubsof	In (logarithm) of # of times substituted Off in CL qualifications - Previous season
lclminutesplayedpreviou	In (logarithm) of # of minutes played in CL - Previous season
llocalcompetitionappstar	In (logarithm) of # of starting lineup in local competition - Previous season
llocalcompetitionappsub	In (logarithm) of # of times substituted ON in local competition - Previous season
llocalcompetitionappsof	In (logarithm) of # of times substituted Off in local competition - Previous season
linternationalapppreviou	In (logarithm) of # of appearances with national team - Previous season
linternationalminutespla	In (logarithm) of # of minutes played with national team - Previous season
lclappstarting11previous2	In (logarithm) of # of starting lineup in CL - Second Previous season
lclappsubstitutedonprev2	In (logarithm) of # of times substituted ON in CL - Second Previous season
lclappsubstitutedoffprev2	In (logarithm) of # of times substituted Off in CL - Second Previous season
lclqualificationsappstart2	In (logarithm) of # of starting lineup in CL qualifications- Second Previous season
lclqualificationsappsubst2	In (logarithm) of # of times substituted ON in CL qualifications - Second Previous season
lclqualificationsappsubsof2	In (logarithm) of # of times substituted Off in CL qualifications - Second Previous season
lclminutesplayedpreviou2	In (logarithm) of # of minutes played in CL - Second Previous season
llocalcompetitionappstar2	In (logarithm) of # of starting lineup in local competition - Second Previous season
llocalcompetitionappsub2	In (logarithm) of # of times substituted ON in local competition - Second Previous season
llocalcompetitionappsof2	In (logarithm) of # of times substituted Off in local competition - Second Previous season
linternationalapppreviou2	In (logarithm) of # of appearances with national team - Second Previous season

linternationalminutespla2	In (logarithm) of # of minutes played with national team - Second Previous season
IEUclappstarting11previous2	In (logarithm) of # of starting lineup in UCL - Second Previous season
IEUclappsubstitutedonprev2	In (logarithm) of # of times substituted ON in UCL - Second Previous season
IEUclappsubstitutedoffprev2	In (logarithm) of # of times substituted Off in UCL - Second Previous season
IEUclqualificationsappstart2	In (logarithm) of # of starting lineup in UCL qualification - Second Previous season
IEUclqualificationsappsubst2	In (logarithm) of # of times substituted ON in UCL qualification - Second Previous season
IEUclqualificationsappsubsof2	In (logarithm) of # of times substituted Off in UCL qualification - Second Previous season

Table AP 1.4: Variables Description

## 9.2 Appendix 2: Tables and Figure for Chapter 4

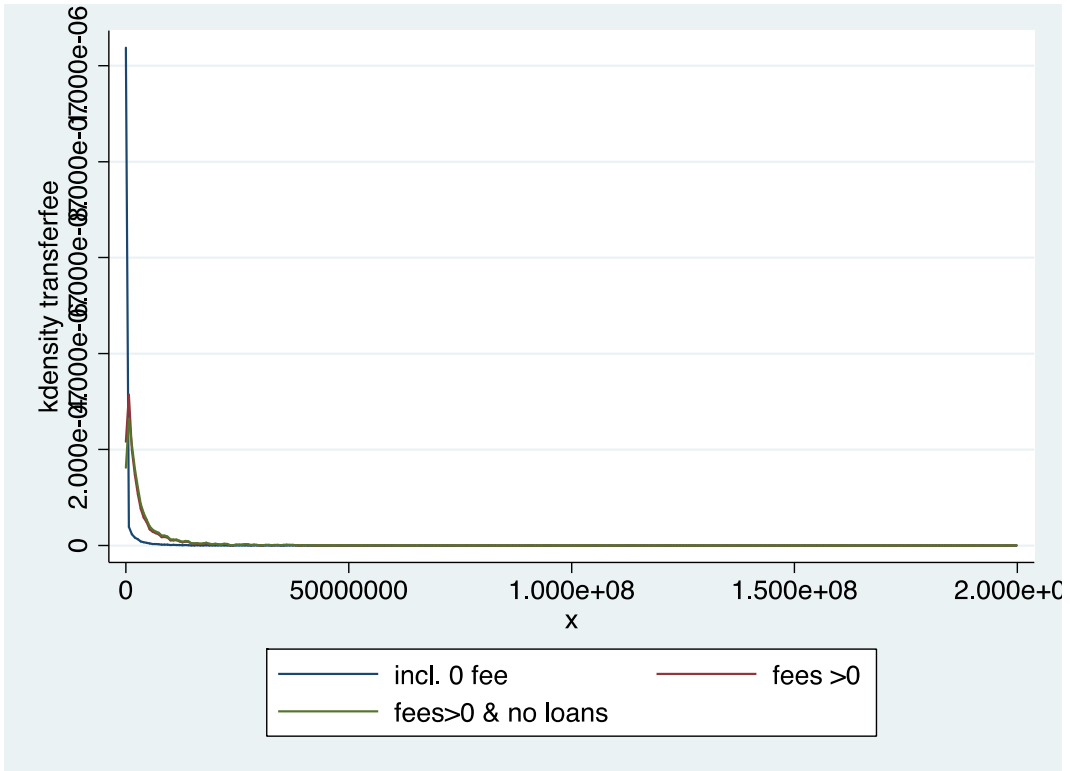
Period	Average Annual Growth Rate
1894-2017	8,1%
1967-2017	9,1%
1992-2017	11,5%
2007-2017	11,7%

**Table AP 2.1: AAGR of Max transfer fees in £2019 PPP**

Country	# obs.	Country	# obs.
Argentina	1010	Italy	2819
Brazil	155	Mexico	1344
Chile	334	Spain	948
China	124	USA	1202
England	2212	Uruguay	165
France	1355	Other	1133
Germany	1250	Total	14051

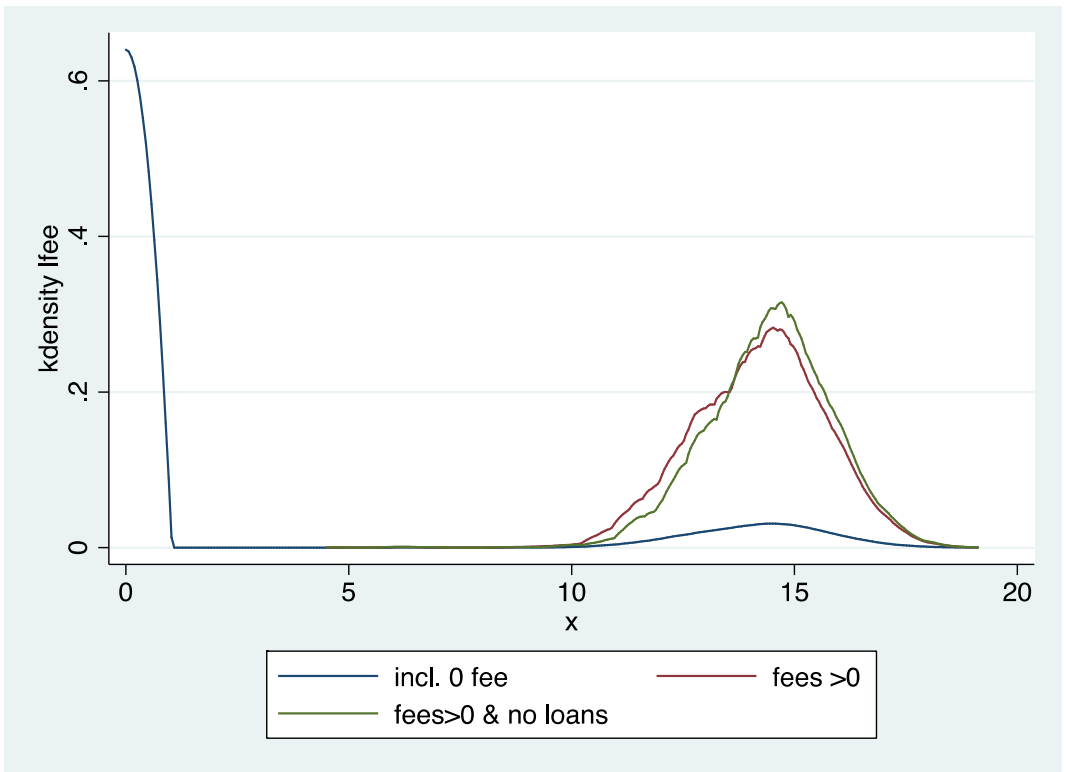
**Table AP 2.2: Unique observations in our database per country 2007-2019**





**Figure AP 2.1: Non-normality of transfer fees**

**Kernel density estimate of transfer fee**



**Kernel density estimate log of transfer fee**

Even dismissing zero fees and transfer fees linked to loans does not make the distribution normal. All tests (Jarque Beira + Shapiro-Wilk) conclude that the distribution of fees is not normal.

**Jarque Beira tests of normality for transfer fees and log transfer fees are negative:**

. sktest transferfee

```

Skewness/Kurtosis tests for Normality
----- joint -----
Variable | Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2
-----+-----
transferfee | 5.8e+04 0.0000 0.0000 . .

```

. sktest transferfee if transferfee!=0

```

Skewness/Kurtosis tests for Normality
----- joint -----
Variable | Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2
-----+-----
transferfee | 6.8e+03 0.0000 0.0000 . .

```

. sktest transferfee if transferfee!=0&tt\_loan==0

```

Skewness/Kurtosis tests for Normality
----- joint -----
Variable | Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2
-----+-----
transferfee | 5.8e+03 0.0000 0.0000 . .

```

. sktest lfee

```
Skewness/Kurtosis tests for Normality
----- joint -----
Variable | Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2
-----+-----
lfee | 5.8e+04 0.0000 0.0000 . .
```

. sktest lfee if lfee!=0

```
Skewness/Kurtosis tests for Normality
----- joint -----
Variable | Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2
-----+-----
lfee | 6.8e+03 0.0000 0.0000 . 0.0000
```

. sktest lfee if lfee!=0&tt\_loan==0

```
Skewness/Kurtosis tests for Normality
----- joint -----
Variable | Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2
-----+-----
lfee | 5.8e+03 0.0000 0.0000 . 0.0000
```

**Shapiro-Wilk tests of normality for transfer fees and log transfer fees are negative:**

. swilk transferfee

```
Shapiro-Wilk W test for normal data
Variable | Obs W V z Prob>z
-----+-----
transferfee | 57739 0.35552 1.3e+04 26.392 0.00000
```

. swilk transferfee if transferfee!=0

```
Shapiro-Wilk W test for normal data
Variable | Obs W V z Prob>z
-----+-----
transferfee | 6795 0.45279 1945.342 20.056 0.00000
```

. swilk transferfee if transferfee!=0&tt\_loan==0

```
Shapiro-Wilk W test for normal data
Variable | Obs W V z Prob>z
-----+-----
transferfee | 5801 0.47291 1629.632 19.495 0.00000
```

```
. swilk lfee
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
lfee	57739	0.97718	469.735	17.104	0.00000

```
. swilk lfee if lfee!=0
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
lfee	6795	0.99292	25.180	8.543	0.00000

```
. swilk lfee if lfee!=0&tt_loan==0
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
lfee	5801	0.98860	35.239	9.389	0.00000

### Appendix 3 – Variables Names and Description

<b>I. Variables</b>	<b>Description</b>
<b>Freetransfer</b>	Dummy for players moving after finishing previous contract
<b>tt_loan</b>	Dummy for players moving on loan
<b>Lpace</b>	ln (logarithm) of the pace
<b>Lshooting</b>	ln (logarithm) of the shooting scores
<b>Ldribbling</b>	ln (logarithm) of the dribbling
<b>Lpassing</b>	ln (logarithm) of the passing
<b>Ldefending</b>	ln (logarithm) of the defending
<b>Lphysicality</b>	ln (logarithm) of the physicality (physical strength)
<b>lsf_follow</b>	ln (logarithm) of the number of followers of players on sofifa website
<b>lRemDurDays</b>	ln (logarithm) of the remaining duration of previous contract at the time of current contract
<b>Lsalarp</b>	ln (logarithm) of the previous salary at previous club during previous contract
<b>Minoraiola</b>	Dummy for players having minoraiola as their agent
<b>Mediabase</b>	Dummy for players having mediabase as their agent
<b>Gestifute</b>	Dummy for players having gestifute as their agent
<b>Lage</b>	ln (logarithm) of players age
<b>lage2</b>	ln (logarithm) of players age squared (lage squared)
<b>Lagem</b>	ln (logarithm) of players whose age is below 24 years old
<b>lage2m2</b>	ln (logarithm) of players whose age is below 24 years old - squared (lage2 squared)
<b>Lagep</b>	ln (logarithm) of players whose age is above 24 years old (lagep)
<b>lagep2</b>	ln (logarithm) of players age is above 24 years old - squared (lagep squared)
<b>Ldur</b>	ln (logarithm) of duration of contract
<b>ldur_hyp</b>	ln (logarithm) of duration of contract
<b>Lheight</b>	ln (logarithm) of player's height
<b>Lgoo</b>	ln (logarithm) of player's search frequency on google
<b>Lgt</b>	ln (logarithm) of player's search frequency as a google trend
<b>lco_cuclu</b>	ln (logarithm) of number of transfers done by the current/buying club 2007/2008-2018/2019
<b>lco_preclu</b>	ln (logarithm) of number of transfers done by the previous/selling club 2007/2008-2018/2019
<b>ltot_preclu</b>	ln (logarithm) of the volume of transfers (£) done by the current/buying club 07/08-18/19
<b>ltot_cuclu</b>	ln (logarithm) of the volume of transfers (£) done by the previous/selling club 07/08-18/19
<b>footer_both</b>	Dummy for players playing in both feet
<b>footer_left</b>	Dummy of players playing in left foot
<b>footer_right</b>	Dummy of the players playing in right foot

<b>nat_cont_africa</b>	Dummy of players of African origin
<b>nat_cont_asia</b>	Dummy of players of Asian origin
<b>nat_cont_australia</b>	Dummy of players of Australian origin
<b>nat_cont_europe</b>	Dummy of players of European origin
<b>nat_cont_north_america</b>	Dummy of players of North American origin
<b>nat_cont_south_america</b>	Dummy of players of South American origin
<b>pos_forward</b>	Dummy of forward players (Strikers)
<b>pos_def</b>	Dummy of midfield players (Midfielders)
<b>pos_mid</b>	Dummy of defensive players (Defenders)
<b>y2007x</b>	Yearly Dummy for transfers of the 2007/2008 Two transfers windows
<b>y2008x</b>	Yearly Dummy for transfers of the 2008/2009 Two transfers windows
<b>y2009x</b>	Yearly Dummy for transfers of the 2009/2010 Two transfers windows
<b>y2010x</b>	Yearly Dummy for transfers of the 2010/2011 Two transfers windows
<b>y2011x</b>	Yearly Dummy for transfers of the 2011/2012 Two transfers windows
<b>y2012x</b>	Yearly Dummy for transfers of the 2012/2013 Two transfers windows
<b>y2013x</b>	Yearly Dummy for transfers of the 2013/2014 Two transfers windows
<b>y2014x</b>	Yearly Dummy for transfers of the 2014/2015 Two transfers windows
<b>y2015x</b>	Yearly Dummy for transfers of the 2015/2016 Two transfers windows
<b>y2016x</b>	Yearly Dummy for transfers of the 2016/2017 Two transfers windows
<b>y2017x</b>	Yearly Dummy for transfers of the 2017/2018 Two transfers windows
<b>y2018x</b>	Yearly Dummy for transfers of the 2018/2019 Two transfers windows
CL (Not Variable)	<b>Champions League (Continental; S. American, Asian...)</b>
UCL (Not Variable)	<b>UEFA Champions League</b>
Previous Season (Not Variable)	<b>the season preceeding the date of contract</b>
Second Previous Season (Not Variable)	<b>Two Seasons before the contract date</b>

Table AP 3.1: Variables' Description

### 9.3 Appendix 3: Tables and Figures for Chapter 5

	<i>Transfer Fees - Test Results - ML Models</i>								
	GLM			RF			GBM		
	Training	Validation	Test	Training	Validation	Test	Training	Validation	Test
<b>MSE</b>	3.34E+17	2.62E+17		8.91E+16	4.79E+16		2.00E+16	5.76E+16	
<b>RMSE</b>	5.78E+08	5.12E+08	6.32E+08	2.99E+08	2.19E+08	2.7E+08	1.41E+08	2.4E+08	2.77E+08
<b>MAE</b>	2.19E+08	2.1E+08	2.29E+08	68549202	61431603	67640627	33776526	70219606	71704650
<b>R^2</b>	<b>6.86E-10</b>	<b>-0.0002</b>	<b>-0.00033</b>	<b>0.733536</b>	<b>0.817407</b>	<b>0.816769</b>	<b>0.940213</b>	<b>0.780668</b>	<b>0.807629</b>

Table AP 3.2: Three Models test results – Transfer Fees.

<b>Top 30 Important Variables' per Model -Transfer Fees</b>		
<b>GLM</b>	<b>RF</b>	<b>GBM</b>
sofifa_likes	sofifa_dislikes	sofifa_dislikes
sofifa_dislikes	sofifa_currentclub	tm_nationalitycurrent
tm_previousclub.1.fc köln	tm_nationalitycurrent	sofifa_currentclub
tm_previousclub.1.fc heidenheim	potential	sofifa_likes
tm_previousclub.1.fc k'lautern	sofifa_likes	tm_durationofcontract
tm_previousclub.1.fc köln	reactions	dribbling
tm_previousclub.1.fc magdeburg	tm_durationofcontract	potential
tm_previousclub.1.fc nuremberg	tm_remaining_dur_cont_days	reactions
tm_previousclub.1.fsv mainz	tot_cuclu	tm_remaining_dur_cont_days
tm_previousclub.12 de octobre	ballcontrol	tot_cuclu
tm_previousclub.1860 munich	player_nationality	sofifa_player_position_2
tm_previousclub.NA	tot_preclu	sofifa_player_position
tm_previousclub.a.quilmes	sofifa_follows	player_nationality
tm_previousclub.a. bucaramanga	sofifa_player_position_2	ballcontrol
tm_previousclub.aalborg bk	sofifa_player_position	finishing
tm_previousclub.aalesund	dribbling	positioning
tm_previousclub.aarhus gf	co_cuclu	tm_player_position
tm_previousclub.abano	fkaccuracy	sofifa_countryofcurrentclub
tm_previousclub.abc fc	tm_player_position	tm_countryof_previousclub
tm_previousclub.aberdeen fc	sofifa_countryofcurrentclub	tot_preclu
tm_previousclub.ac ajaccio	sofifa_player_position_3	sofifa_player_position_3
tm_previousclub.ac arles	sprintspeed	sprintspeed
tm_previousclub.ac barnechea	composure	vision
tm_previousclub.ac horsens	tm_countryof_previousclub	headingaccuracy
tm_previousclub.ac le havre	virtual_rem_dur_days	composure
tm_previousclub.ac milan	positioning	instagram_follows
tm_previousclub.ac pisa	tm_endofcontract	longshots
tm_previousclub.acassuso cf	vision	volleys
tm_previousclub.accrington	finishing	mercato
tm_previousclub.acmessina	jumping	shortpassing

Table AP 3.3: Variables' Importance per three Models – Transfer Fees.



	<i>Wages - Test Results - ML Models</i>								
	<b>GLM</b>			<b>RF</b>			<b>GBM</b>		
	Training	Validation	Test	Training	Validation	Test	Training	Validation	Test
<b>MSE</b>	280.4548	275.1813		129.7613	116.2489		19.27647	121.0588	
<b>RMSE</b>	16.74678	16.58859	16.5565	11.39128	10.78188	10.27117	4.390498	11.00267	10.65392
<b>MAE</b>	10.7473	10.66691	10.65374	5.987552	5.75794	5.53866	2.420433	5.883503	5.574294
<b>R<sup>2</sup></b>	<b>0.564612</b>	<b>0.53453</b>	<b>0.555494</b>	<b>0.798554</b>	<b>0.803365</b>	<b>0.828927</b>	<b>0.970075</b>	<b>0.795229</b>	<b>0.81594</b>

Table AP 3.4: Three Models test results - Wages

<i>Top 30 Important Variables per Model - Wages</i>		
<b>GLM</b>	<b>RF</b>	<b>GBM</b>
reactions	sofifa_currentclub	reactions
sofifa_dislikes	reactions	ballcontrol
tot_cuclu	ballcontrol	sofifa_currentclub
co_cuclu	tm_nationalitycurrent	tm_nationalitycurrent
potential	sofifa_likes	tot_cuclu
sofifa_follows	tot_cuclu	mercato
sofifa_likes	tot_preclu	sofifa_player_position_2
tot_preclu	mercato	age
tm_remaining_dur_cont_days	sofifa_dislikes	player_nationality
headingaccuracy	dribbling	sofifa_likes
gkicking	positioning	sofifa_dislikes
age	age	dribbling
volleys	standingtackle	standingtackle
gkdiving	player_nationality	finishing
facebook_follows	sofifa_player_position_2	marking
facebook_likes	slidingtackle	composure
weight_kg	finishing	tm_countryof_previousclub
tm_player_position.defender.centre-back	co_cuclu	slidingtackle
twitter_follows	shotpower	co_cuclu
gkreflexes	tm_countryof_previousclub	tm_player_position
sofifa_countryofcurrentclub.England	composure	positioning
gkhandling	headingaccuracy	sofifa_player_position
tm_division2	shortpassing	vision
co_preclu	longshots	headingaccuracy
instagram_follows	sofifa_follows	tot_preclu
virtual_rem_dur_days	tm_player_position	sofifa_player_position_3
sofifa_player_position.CB	sofifa_player_position_3	gkreflexes
sofifa_endofcontract	marking	interceptions
sprintspeed	sofifa_player_position	tm_endofcontract
vision	interceptions	gkdiving

Table AP 3.5: Variables' Importance per three Models – Wages.

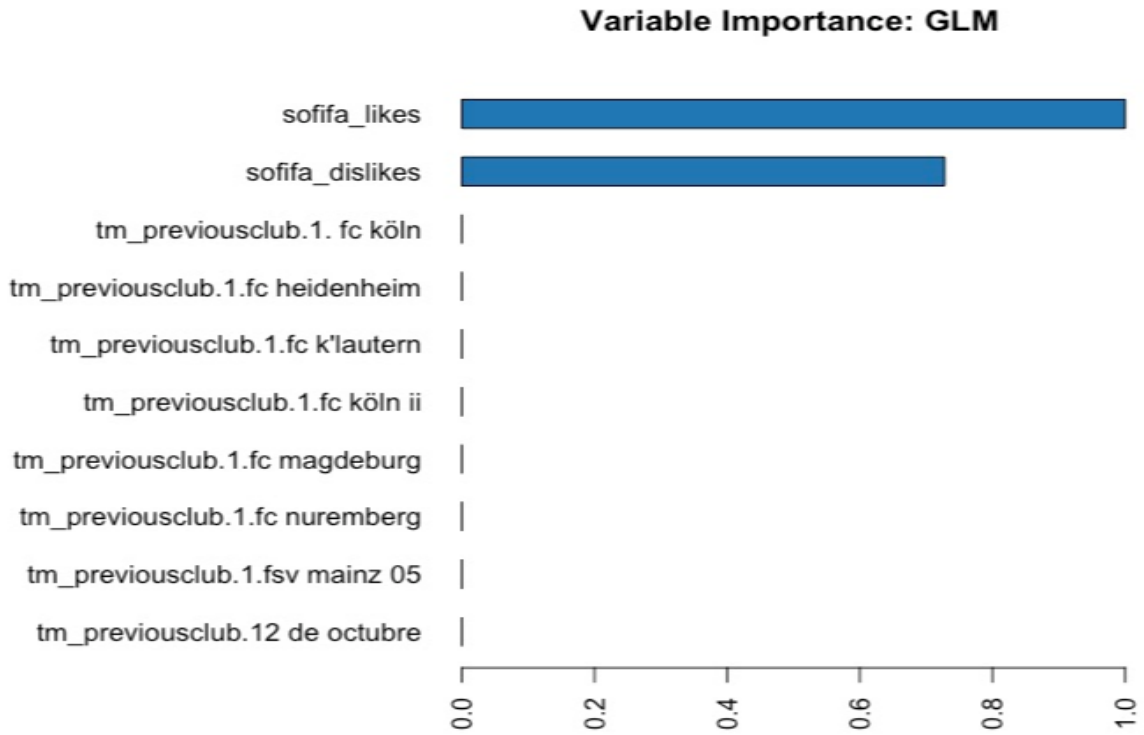


Figure AP 3.1: GLM model – Most important variables affecting Transfer fee

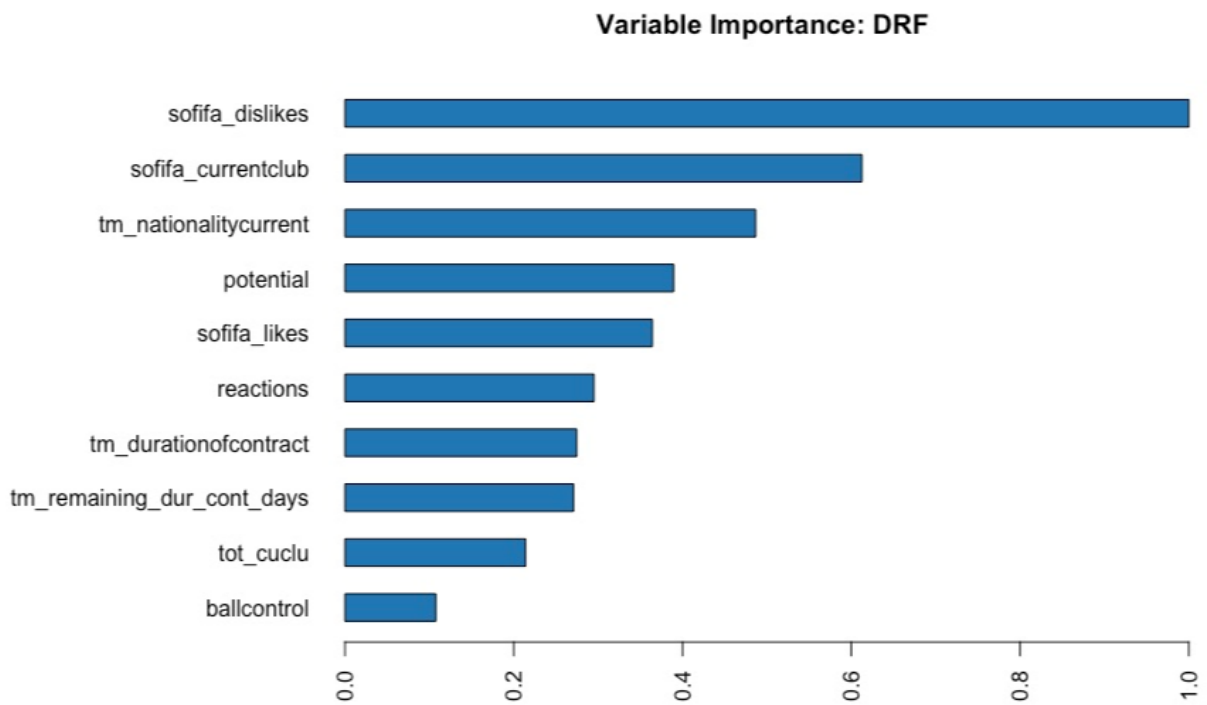


Figure AP 3.2: RF model – Most important variables affecting Transfer fee

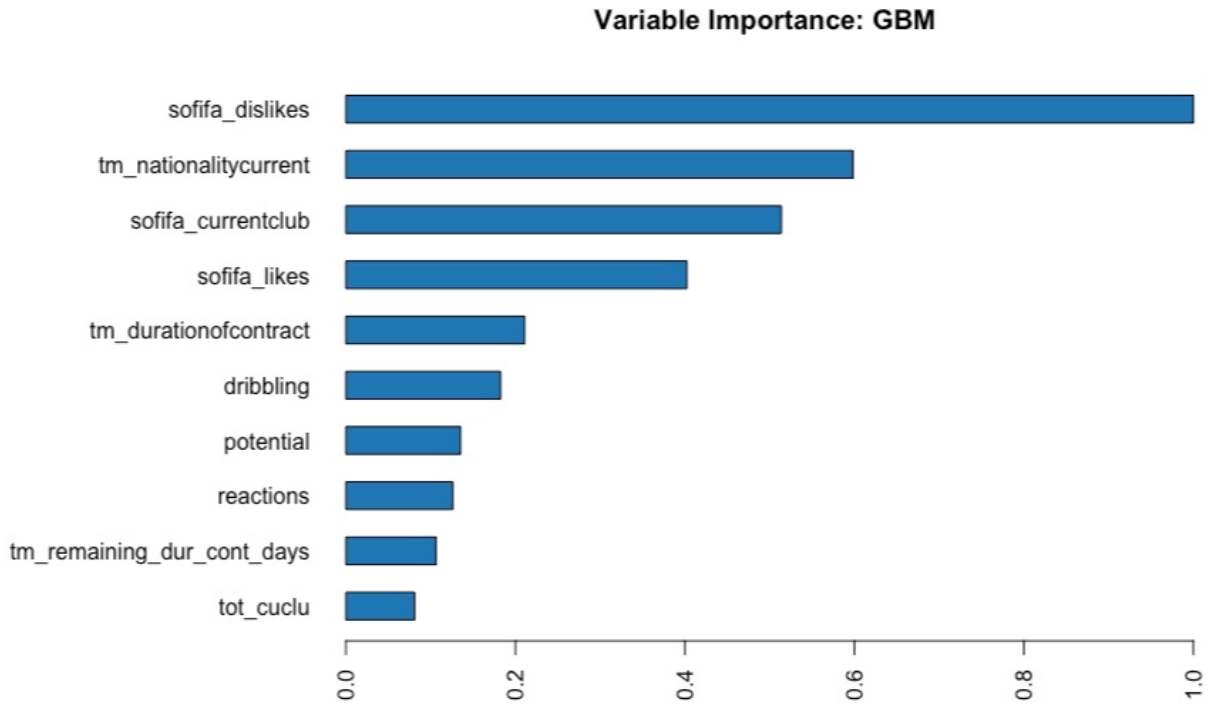


Figure AP 3.3: GBM model – Most important variables affecting Transfer fee

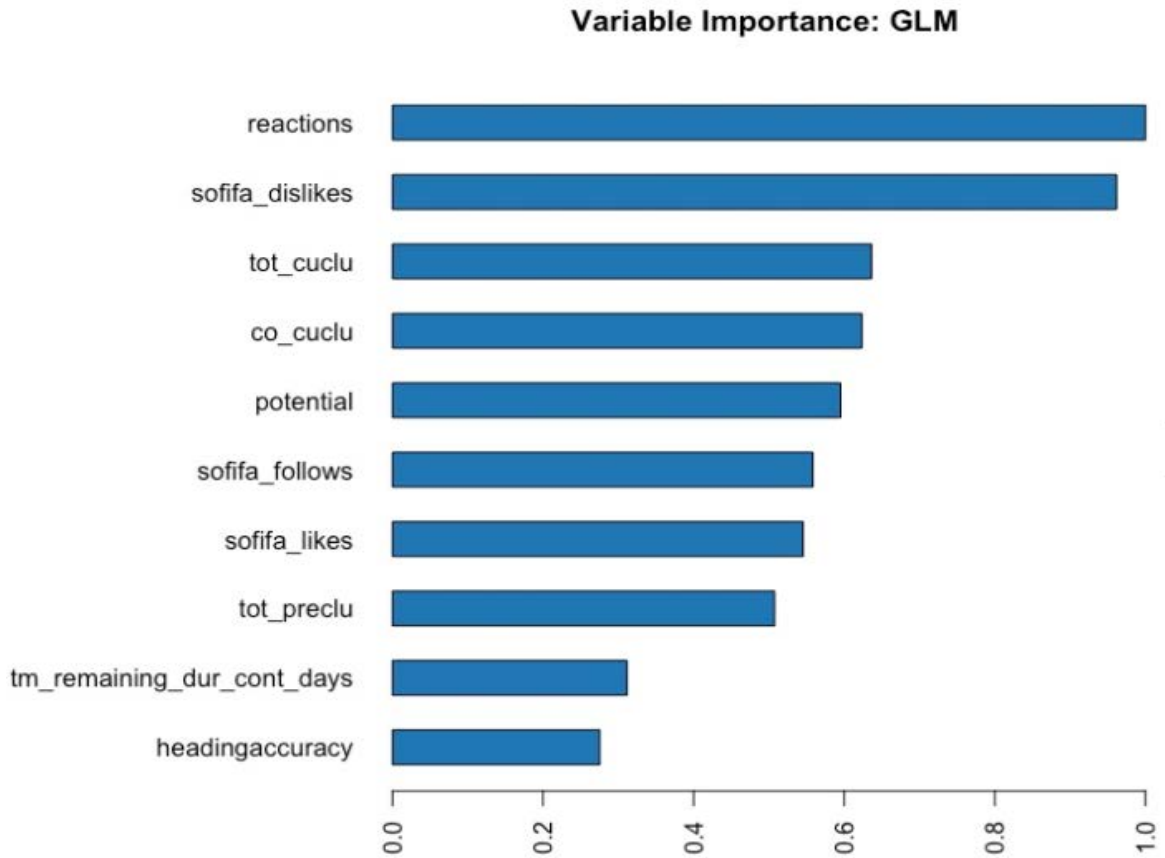


Figure AP 3.4: GLM model – Most important variables affecting Wages

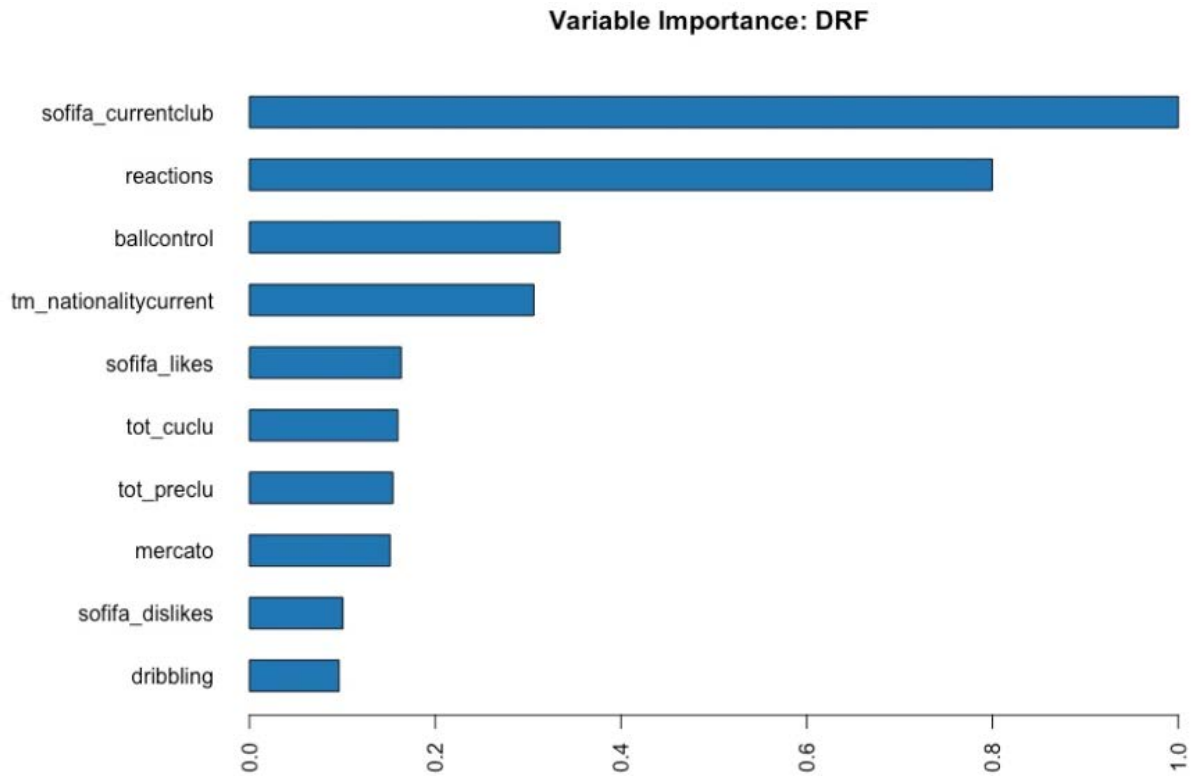


Figure AP 3.5: RF model – Most important variables affecting Wages

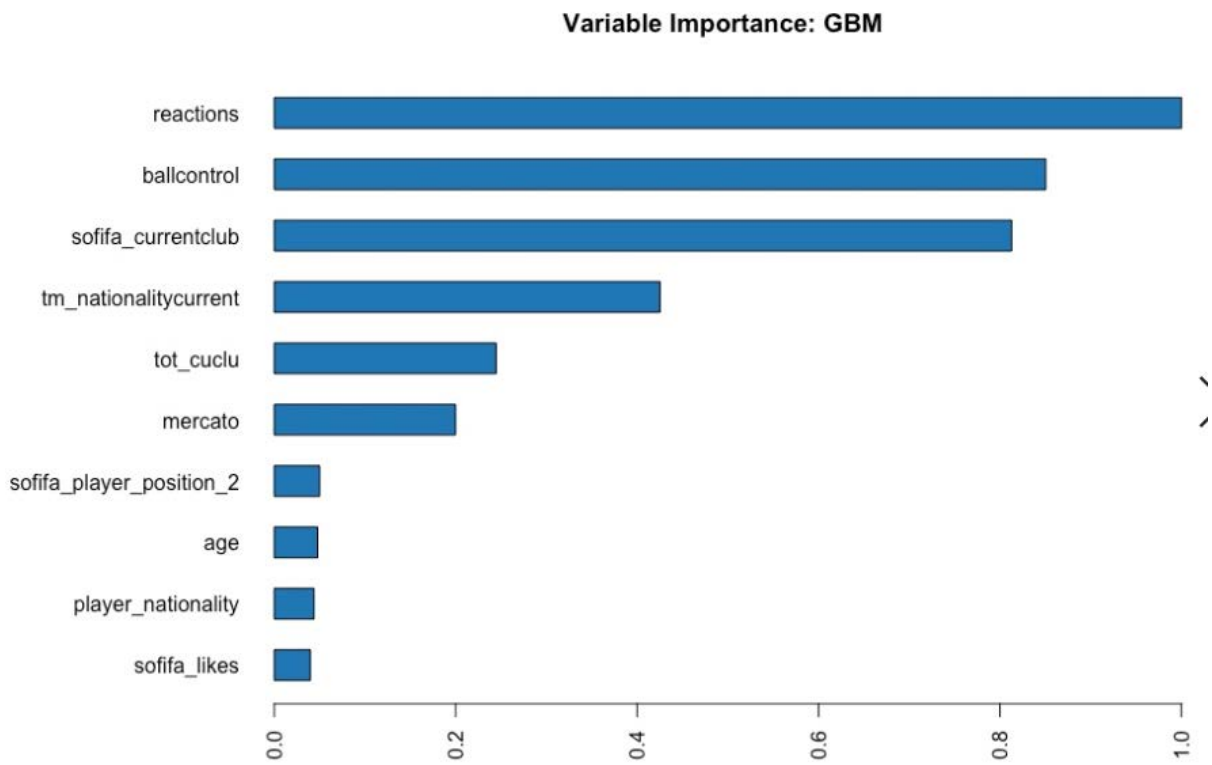


Figure AP 3.6: GBM model – Most important variables affecting Wages