FINDING THE DETERMINANTS OF PLAYER MARKET VALUE IN ASSOCIATION FOOTBALL USING FIFA VIDEO GAME DATA

A Thesis

by

Ozan Can Eren

Submitted to the
Graduate School of Sciences and Engineering
In Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in the
Department of Data Science

Özyeğin University
August 2021

FINDING THE DETERMINANTS OF PLAYER MARKET VALUE IN ASSOCIATION FOOTBALL USING FIFA VIDEO GAME DATA

Approved by:

Assoc. Professor Okan Örsan Özener,
Advisor
Department of Industrial Engineering
Özyeğin University

Assist. Professor Başak Altan Özener,
Department of Economics
Özyeğin University

Assist. Professor Erinç Albey,
Co-Advisor
Department of Industrial Engineering
Özyeğin University

Assoc. Professor Mehmet Güray Güler
Department of Industrial Engineering
Yıldız Technical University

Assist. Professor İhsan Yanıkoğlu,
Department of Industrial Engineering
Özyeğin University

Date Approved: August 2021

To all humanity

ABSTRACT

Association football ("football" hereafter) has been one of the most popular sports in the world in the last decades. In line with its rising popularity, football has become a growing multi-billion-dollar market. Large amounts of money are being paid as transfer fees for the top players in the big European leagues. This resulted in the emergence of an attention to the market values of football players. One of the open problems in this field is finding out the determinants of the players' market values. While some of these determinants have been researched before, the focus has usually been on the on-field performance statistics. In this study, we use a unique dataset and focus on the intrinsic features and attributes of football players that are provided in the FIFA 2015, 2016, 2017 and 2018 video games. Player market values are acquired from "transfermarkt.com". For each of the 2014-2015, 2015-2016, 2016-2017 and 2017-2018 seasons, we only analyzed the players who played for at least ninety minutes in any of the top 5 leagues in Europe (English Premier League, Spanish La Liga, German Bundesliga, Italian Serie A and French Ligue 1) during that season. We follow feature selection methods and build OLS regression models to find out the most important demographic, team-related, physical, technical and mental features and attributes that make a player playing in a sub-position of defense, midfield and attack valuable for each of the analyzed playing season. We show that players from different sub-groups ("Center Backs", "Fullbacks", "Central Defensive Midfielders", "Central Midfielders", "Wingers" and "Strikers") have different features and attributes that make them valuable. For each of the 6 playing positions in every 4 seasons, we build OLS regression models and provide the outputs.

ÖZETÇE

Futbol, son onyıllarda dünyadaki en popüler sporlardan biri haline gelmiştir. Artan popülerliği ile birlikte futbol, büyüyen milyar dolarlık bir pazar halini almıştır. Büyük avrupa liglerindeki üst düzey oyuncuların transferleri için yüksek meblağlar ödenmektedir. Bu durum, futbolcuların piyasa değerlerine dikkat çekilmesine sebep olmuştur. Bu alandaki ucu açık konulardan birisi ise futbol oyuncularının piyasa değerlerini belirleyen faktörlerin tespitidir. Daha önce bu faktörlerin bir kısmı üzerine araştırmalar yapılmış olmasına rağmen odak noktası genellikle saha içi istatistikler olmuştur. Bu tez kapsamında benzersiz bir veri seti kullanılmakta ve futbol oyuncularının FIFA 2015, 2016, 2017 ve 2018 oyunlarındaki esas özellik ve niteliklerine odaklanılmaktadır. Oyuncu piyasa değerleri "transfermarkt.com" adresinden temin edildi. 2014-2015, 2015-2016, 2016-2017 ve 2017-2018 sezonlarının her biri için Avrupa'nın en büyük 5 liginde (İngiliz Premier Ligi, İspanyol La Liga, Alman Bundes Liga, İtalyan Serie A ve Fransız Lig 1) en az doksan dakika sahada kalan futbolcular analiz edildi. Çeşitli değişken seçme metotları kullanılıp Sıradan En Küçük Kareler modelleri oluşturularak defans, orta saha ve hücum bölgelerindeki spesifik alt pozisyonlarda oynayan oyuncuları değerli kılan en önemli demografik, takımsal, fiziksel, teknik ve mental özellik ve nitelikler her sezon için tespit edildi. Farklı alt pozisyonlarda (Stoper, Bek, Merkezi Defansif Orta Saha, Merkezi Orta Saha, Kanatçı ve Forvet) oynayan oyuncuları değerli kılan farklı özellikler ve nitelikler olduğu gösterildi. Her 4 sezon ve 6 alt pozisyon için oluşturulan Sıradan En Küçük Kareler modellerinin çıktıları paylaşıldı.

ACKNOWLEDGEMENTS

First of all, I would like to offer my gratitude to my advisors Okan Örsan Özener and Erinç Albey. I would like to thank to Başak Altan who acted as a mentor to me. They constantly guided me since the beginning of my studies and showed understanding during my hard times.

I would like to thank my parents and sister for constantly believing in and encouraging me.

I am thankful for "TÜBİTAK 2210-A Genel Yurt İçi Lisansüstü Burs Programı" for supporting me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Football is the most popular sport in the world in terms of number of players and fans. It is estimated that there are 3.3-3.5 billion football fans around the world [1]. With its ever-rising popularity, football clubs are turning into "football companies" [2] generating revenues in the amounts that were never seen before. These football companies are paying astronomical amounts of money for the transfers of top players. As the transfer market grows rapidly, the question "What makes a football player valuable?" has aroused curiosity and gained importance.

In this paper, we build value assessment models by breaking down the most important demographic, player-related, physical, technical and mental features and attributes of players. The market value data was acquired from the online platform "transfermarkt.com" which is the leading website on the football transfer market. The website includes estimations of market values for almost every football player playing in a professional league in any country. Football players' market values are estimates of their transfer fees. It has been shown that market values and transfer fees are comparable [3][4]. The player market value data of transfermarkt.com has been proved to be accurate many times [5]. The economical valuation of the website seems to match the prices paid for the transferred players [4]. The website updates the market values of all players at least 2 times a year (at the end of the half season and end of the whole season). We acquired the market value data of players that belong to the end-season updates from 2015 to 2018.

So far, the research made on the topic of finding the determinants of the players' market values has usually been by building value assessment models using the realized on-field performance statistics. However, it should be noted that on-field performance statistics are not players' actual abilities or capabilities but rather indicators and probabilistic actualizations of their true skills and attributes. In this paper, we propose a value assessment model that is not directly reliant on on-field performance statistics (other than "Number of Minutes Played"). In order to do that, we used the data from the EA Sports' video game "FIFA" which includes demographic, team-related, physical, technical and mental features and attribute ratings of players. Listed in Guinness World Records as the best-selling sports video game franchise in the world, by 2021, the FIFA series had sold over 325 million copies [6]. There are more than 30 player features and attributes to be found in the video game. Examples of the features and attributes are weight, height, age, team, ball control, dribbling, aggression, vision, sliding tackle, short passing, finishing, heading accuracy etc. The game's database is updated on a weekly basis. We acquired the data that was updated right before the end-season player market value updates of the transfermarkt.com. To our knowledge, this data set has never been used in the literature. Additionally, our results are specific to each player position which make our analysis quite valuable for not only football coaches, managers and scouts but junior players as well.

EA Sports employs a team of 25 EA Producers and hundreds of outside data contributors. The main responsibility of this team is to make sure that all player data in the video game is up to date. There is another community of over 6,000 people which consists of FIFA Data Reviewers or Talent Scouts from all over the world. This community constantly provides suggestions and alterations to the database. This detailed process ensures that the information in the video game is as accurate as possible [7]. These data reviewers use their expertise on assessing a player's actual capabilities that are not necessarily directly indicated by on-field performance statistics and assign a rating for each attribute of a player. The main difference between the realized on-field performance statistics and player attribute ratings in the video game is that on-field performance statistics are completely objective metrics. This type of data lacks the ability to indicate a player's true potential and capabilities as each player has his ups and downs throughout a playing season and doesn't always reflect his capabilities on the performance statistics fully. In contrast, the player attribute ratings in the video game are not just based on completely objective on-field performance statistics but rather include the opinions of the EA Sports' data reviewers and expert talent scouts. It can be argued that there is a latent knowledge used by these type of data reviewers and talent scouts while assigning attribute ratings for players [8]. By using this unique player attributes data, we were able to utilize this latent knowledge of the data reviewers and talent scouts of EA Sports.

Using the above-described datasets, we built multiple OLS regression models for different sub-groups of players that are playing in specific positions on the field. Our main goal was to break down the features and intrinsic skills & attributes of football players that make them valuable in each of the four playing seasons.

Our empirical analysis suggests that age, league, team rating, the international reputation of the player are significant features for all of the positions. However, other than these features, different playing positions require different set of features affecting the valuation of the player. Another interesting result we reach is that if a player is recruited by a team in the English Premier League (EPL), this player's valuation will be significantly higher even though all the features affecting the quality of his play are the same. This implies that junior players immediately started playing in the EPL likely to have a higher valuation path throughout their carrier than the rest of their peers.

CHAPTER II

PREVIOUS WORK

While the popularity of football has always been rising, the economical valuation of both teams and players have reached record numbers. As these football teams were started getting referred as "football companies" [2], the concept of "Moneyball in Football" has emerged. In various research projects, the key performance indicators and determinants for market value of football players were studied. In 2011 51 Sports Science students and 15 Performance Analysis experts came together in Hungary for an Intensive Program in Performance Analysis of Sport (IPPAS). Their main goal was to define the key Performance Indicators for different positions in association football. Every group spent hours debating and arguing over the key Performance Indicators [9]. Although the main objective was not directly to find out the determinants of the market value, the result of this study is one of the vastest in terms of the number of different key performance indicators defined for players playing in different positions in the field. However, it should be noted that the findings of this research are so generalized that almost every skill is declared to be significant for every player subgroup. In addition to that, that research was not based on any direct quantitative analysis but was based on the consensus of some experts. When it comes to the research in which the determinants of players' market values were analyzed using quantitative methods, the focus has usually been on the on-field performance statistics of players. One of these types of studies was published in 2017 in which linear regression analyses were made on a dataset of players from the top 5 European leagues [3]. The dataset included the market value data from transfermarkt.com which is the leading website on the football transfer market. In 2019, KPMG, which is one of the

Big Four accounting organizations and provides advisory services, has developed a player valuation tool using linear regression models [10]. These examples show the importance of and interest in finding the determinants of football players' market values.

CHAPTER III

DATASET

3.1 Dataset Description

The dataset that we use in our analyses is an ensemble of two datasets acquired from "transfermarkt.com" and "sofifa.com".

3.1.1 Player Market Value Data

From transfermarkt.com we collected the player market value data. This website was used as a market value data source in many previous studies [11]. We collected the market value data (in million Euros) for the football players that were in the squad of a team in the top 5 European leagues during a whole season. These 5 leagues are the English Premier League, Spanish La Liga, German Bundesliga, Italian Serie A and French Ligue 1. The data belongs to the end-season player market value update of transfermarkt.com from 2015 to 2018.

3.1.2 Player Features and Attributes Data

From "sofifa.com" we collected player-related features and attribute ratings that are to be found in the FIFA 2015, 2016, 2017 and 2018 video games. This website was used as a data source in previous research on football analytics [12]. The database is updated on a weekly basis. The player features and attributes data that we used in our analyses belongs to the weekly update that was right before the end-season player market value update of transfermarkt.com for each season. The dataset can be broken down into 5 parts: Demographic Features, Team Related Features, Physical Attributes, Technical Attributes and Mental Attributes.

Demographic Features

Age: Age of the player

EU: Defines whether the player is from a country that is in the European Union.

International Reputation: Defines how internationally well-known a player is

Team Related Features

League: Name of the country to which the player's team belongs

Team Rating: A rating determined by the average attributes of a team's players

National Team: Defines whether the player plays for his country's national team.

Mental Attributes

Aggression: Defines the aggression level of a player on pushing, pulling and tackling. Aggressive players generally win more tackles, but also risk committing fouls.

Interceptions: Defines the player's ability to read the game and intercept passes.

Positioning: Defines the player's ability to spot open spaces and move into good positions.

Vision: Defines the player's mental awareness about his teammates' positioning for passing the ball to them.

Technical Attributes

Position: The main position of the player on the field

Ball Control: Defines how well the player initially controls the ball, then how good he is at keeping it under control.

Crossing: Defines the accuracy and the quality of the player's crosses.

Curve: Defines the player's ability to curve the ball when passing and shooting.

Dribbling: Defines the player's ability to dribble/carry the ball and past an opponent while being in control.

Finishing: Defines the accuracy of the player's shots (using foot) inside the penalty area.

FK Accuracy: Determines the accuracy of the player when taking a free kick.

Heading Accuracy: Defines the player's accuracy when using the head to pass, shoot or clear the ball.

Marking: Defines the player's capability to mark players to prevent them from taking control of the ball.

Preferred Foot: The foot the player mainly uses (left or right)

Penalties: Determines the player's accuracy for the penalty kicks.

Short Passing: Defines the short passing accuracy of the player.

Long Passing: Defines the long passing accuracy of the player.

Shot Power: Defines the amount of power a player can put into a shot.

Sliding Tackle: Defines how well the player can win the ball and keep possession of the ball while making a sliding challenge without conceding a foul.

Standing Tackle: Defines how well the player can win the ball and keep possession of the ball while making a standing challenge without conceding a foul.

Volleys: Defines the player's accuracy while performing volleys (shots taken when the ball is in the air).

Weak Foot: Defines how well the player is able to use the other foot than his preferred foot.

Physical Attributes

Height: Height of the player in centimeters

Weight: Weight of the player in kilograms

Acceleration: Defines how quickly the player can reach his top speed.

Agility: Defines the player's ability to change direction at pace.

Balance: Defines the ability of the player to remain upright and steady when running, carrying and controlling the ball.

Jumping: Defines how high the player can jump.

Reactions: Defines how quickly the player reacts to certain events around him.

Sprint Speed: Defines how fast the player runs while at top speed.

Stamina: Defines the amount of time the player can sprint before slowing down. Also defines how fast he recovers his energy after a match.

Strength: Defines the player's ability of succeeding in a physical challenge/jostle.

<u>Other</u>

Minutes Played: For how many minutes the player was on the field in total in a season in the domestic league games (This is the only feature in our dataset that may be considered as an on-field statistic. It has been proven many times that playing time in domestic leagues has a very significant positive impact on both a player's salary and market value [13][14][3]. We use this feature in our analysis in order to eliminate its effect on the market value and find the true effect of other player features and attributes.)

To increase the reliability of the player attribute data and in line with previous research, we only considered the players who were on the pitch for at least ninety minutes in a given season [15] and excluded goalkeepers from our dataset [16] as their performance is assessed in a considerably different way than that of outfield players [3]. The players that were on loan were also excluded. The distribution of the market values is highly right skewed in every season. The whole descriptive statistics of the player features and attributes data and histogram of the market values are provided in the following table and figures:

Table 1: Descriptive statistics of all features and attributes (the whole data)

Numeric Variables

Variable	Measurement	Range	Mean	St. Dev.	Min.	Max.
Age	Years (float)	≥ 0	27.13	4.09	17.42	41.75
Minutes Played	Minutes	0 to 3420	1665	876.83	90	3420
Team Rating	Rating (float)	0.5 to 5	4.06	0.51	3	5
International Rep.	Rating (integer)	1 to 5	1.6	0.78	1	5
Height	Centimeters	≥ 0	181.51	6.16	160.02	203.2
Weight	Kilograms	≥ 0	75.94	6.6	55.79	101.15
Acceleration	Rating (integer)	1 to 99	70.58	12.12	21	97
Aggression	Rating (integer)	1 to 99	68.63	13.09	13	94
Agility	Rating (integer)	1 to 99	69.39	12.18	21	96
Balance	Rating (integer)	1 to 99	67.58	12.65	25	96
Ball Control	Rating (integer)	1 to 99	72.72	8.41	29	96
Crossing	Rating (integer)	1 to 99	63.56	14.22	11	93
Curve	Rating (integer)	1 to 99	60.77	15.49	12	94
Dribbling	Rating (integer)	1 to 99	68.59	12.39	14	97
Finishing	Rating (integer)	1 to 99	55.56	18	10	95
FK Accuracy	Rating (integer)	1 to 99	54.25	16.31	10	94
Heading Acc.	Rating (integer)	1 to 99	64.58	12.09	22	95
Interceptions	Rating (integer)	1 to 99	60.66	19.78	10	93
Jumping	Rating (integer)	1 to 99	69.4	10.99	26	96
Long Passing	Rating (integer)	1 to 99	65.63	10.76	22	95
Long Shots	Rating (integer)	1 to 99	60.12	16.09	11	94
Marking	Rating (integer)	1 to 99	57.32	20.98	11	94
Penalties	Rating (integer)	1 to 99	57.89	13.52	12	96
Positioning	Rating (integer)	1 to 99	62.07	16.36	12	95
Reactions	Rating (integer)	1 to 99	72.47	7.17	38	96
Short Passing	Rating (integer)	1 to 99	72.09	7.96	27	93
Shot Power	Rating (integer)	1 to 99	68.53	12.36	13	96
Sliding Tackle	Rating (integer)	1 to 99	58.39	21.45	10	95
Standing Tackle	Rating (integer)	1 to 99	61.26	20.6	11	93
Sprint Speed	Rating (integer)	1 to 99	71.18	11.46	21	96
Stamina	Rating (integer)	1 to 99	72.97	9.9	31	97
Strength	Rating (integer)	1 to 99	70.63	11	24	95
Vision	Rating (integer)	1 to 99	63.47	13.96	13	95
Volleys	Rating (integer)	1 to 99	54.53	16.99	11	93
Weak Foot	Rating (integer)	1 to 5	3.14	0.67	1	5
Market Value	Million Euros	≥ 0	9.31	14.86	0.05	200

Binary & Categorical Variables

Variable	Level	Count
EU	1	5675
	0	2635
League	England	1743
	Germany	1626
	Spain	1609
	France	1720
	Italy	1612
National Team	1	1591
	0	6719
Position	CB	1692
	LB	765
	RB	780
	LWB	29
	RWB	26
	CDM	813
	CM	1196
	CAM	411
	LM	500
	RM	513
	LW	178
	RW	181
	CF	104
	ST	1122
Preferred Foot	Left	2156
	Right	6154

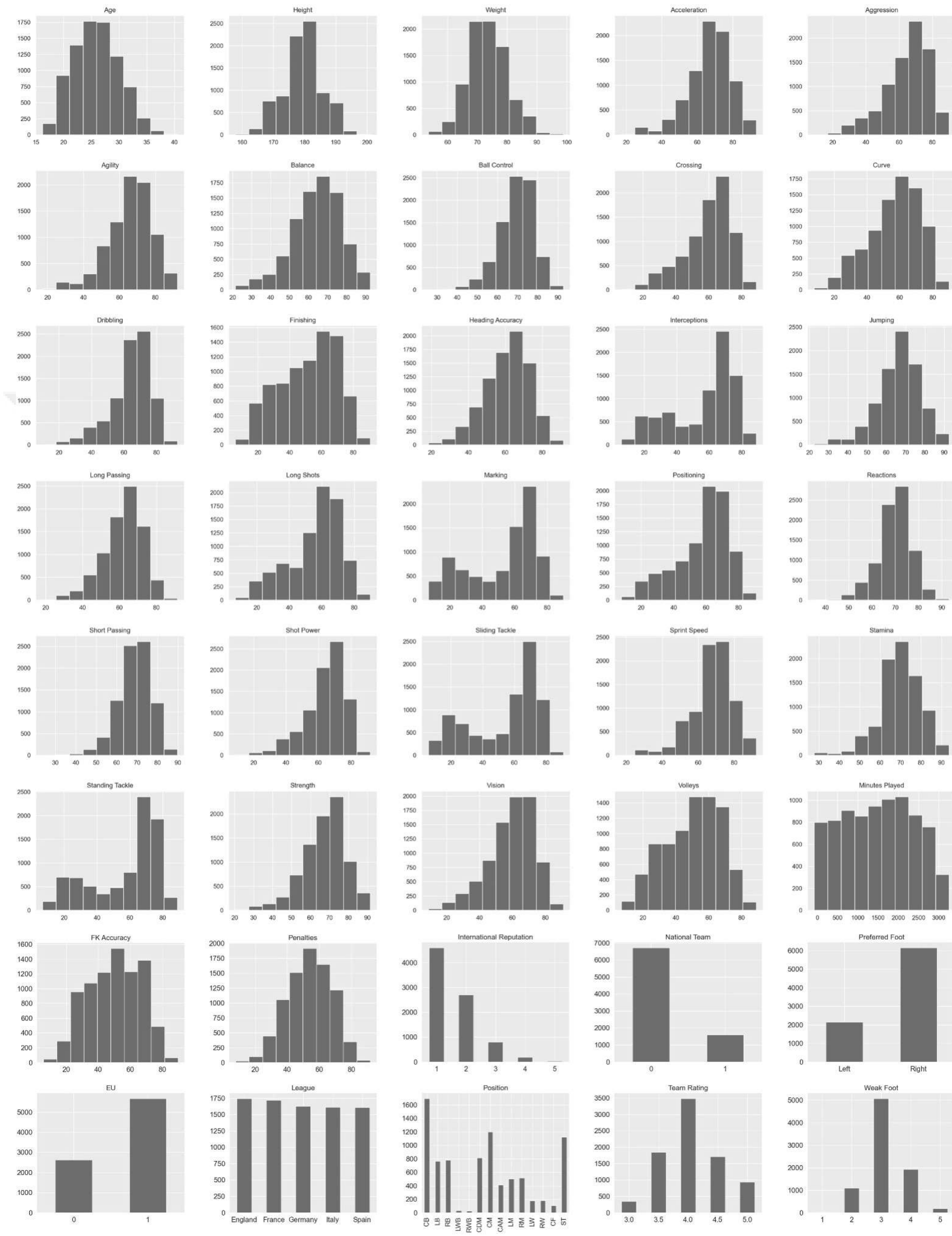

Figure 1: Histograms of all features and attributes

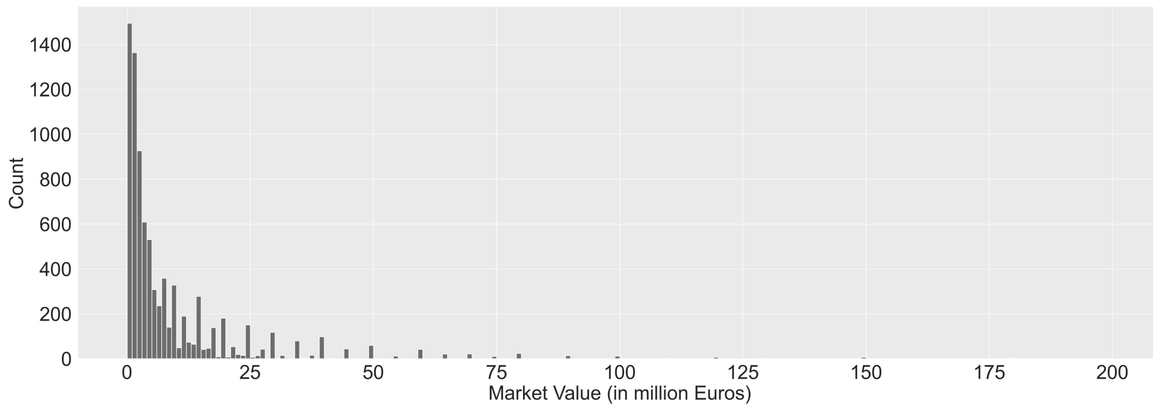

Figure 2: Histogram of market values

3.2 Dataset Collection

In order to acquire the player market value data from "transfermarkt.com", we developed our own web scraper using Python programming language. The player features and attributes data from "sofifa.com" was also acquired by using our own web scraper. In order to combine the two datasets, we needed a unique feature that is the same in both datasets. The first idea that comes to mind is using the feature "Player Name". However, this feature was not the same in the two datasets as some player names were reported as full names in one dataset and abbreviated in the other. Some players' names were abbreviated in both datasets but in different ways. So, we created a new set of features that is unique and the same in both datasets which consists of the "Team Name" and "Jersey Number". This set of features is unique for each player and therefore can be used as a unique key to combine the two datasets. This dataset combining method itself is a value-add as this sort of linkage problem has previously hindered some researchers in this field to create and work on large datasets [4]. Using this unique set of features, we combined the two datasets created the final one which is used in our analyses.

3.3 *Preparing the Data*

3.3.1 Dividing Data into Sub-Groups

For each of the four seasons before building OLS models, we divided our data into subgroups based on the main playing positions of the players. It is already proven in previous research that football players playing in different main positions have different attributes that are desired [9]. Other than goalkeepers, there are three different main positions: Defense, Midfield and Attack. However, it should be noted that dividing the players into four positions is a simplification [8]. As it's also suggested in previous research, it is possible and more accurate to divide the playing positions even further [9]. So, we divide our data into six sub-groups that are:

- **Center Backs**
- **Fullbacks** (consist of Left/Right Backs & Left/Right Wing Backs)
- **Central Defensive Midfielders**
- **Central Midfielders** (consist of Central Midfielders & Central Attacking Midfielders)
- **Wingers** (consist of Left/Right Midfielders & Left/Right Wingers)
- **Strikers** (consist of Strikers & Center Forwards)

The descriptive statistics of every sub-group are as follows:

Table 2: The mean values of numeric variables for every sub-group of players (the whole data)

Variable	Center Backs	Fullbacks	Central Defensive Midfielders	Central Midfielders	Wingers	Strikers
Age	28.06	27.05	27.60	26.76	26.13	27.22
Minutes Played	1844.04	1637.45	1694.06	1685.44	1541.72	1545.73
Team Rating	4.07	4.07	4.04	4.10	4.06	4.02
International Rep.	1.60	1.44	1.58	1.69	1.61	1.69
Height	186.82	179.45	182.23	179.62	177.73	183.10
Weight	81.15	74.00	76.31	73.75	72.31	77.99
Acceleration	59.72	75.22	63.07	69.85	81.35	73.41
Aggression	76.42	72.40	77.63	67.42	56.36	62.30
Agility	55.90	71.11	65.31	73.41	79.71	71.67
Balance	55.18	70.71	65.62	72.20	76.33	66.10
Ball Control	62.96	71.43	73.49	77.87	77.48	75.30
Crossing	47.16	73.13	59.80	68.82	72.87	58.86
Curve	42.85	62.50	57.70	69.40	70.36	63.22
Dribbling	51.58	68.71	66.25	75.22	78.71	73.46
Finishing	34.63	45.08	50.28	63.64	68.41	76.62
FK Accuracy	39.52	50.45	55.17	65.35	62.20	55.44
Heading Acc.	75.11	63.39	64.77	58.34	53.96	71.53
Interceptions	75.80	71.83	76.14	63.72	41.78	32.02
Jumping	74.42	71.14	69.82	64.91	63.79	72.06
Long Passing	62.27	64.50	72.92	74.88	66.02	54.34
Long Shots	41.54	54.25	61.77	70.66	68.90	68.68
Marking	76.10	71.28	70.82	56.49	37.25	27.81
Penalties	45.76	51.40	56.52	63.25	63.74	70.45
Positioning	38.63	60.13	58.33	70.32	72.20	77.28
Reactions	71.10	71.41	73.26	73.91	71.97	73.92
Short Passing	66.09	70.70	76.43	78.65	73.60	68.99
Shot Power	58.45	64.35	69.54	73.14	72.75	76.46
Sliding Tackle	75.75	74.27	71.96	58.85	38.34	26.55
Standing Tackle	78.06	74.24	76.03	63.40	41.19	30.99
Sprint Speed	63.42	76.24	62.70	68.23	80.36	74.47
Stamina	67.25	77.35	78.15	75.05	71.95	70.15
Strength	79.56	69.05	74.33	66.64	61.39	73.50
Vision	47.53	57.96	69.03	75.56	70.77	64.93
Volleys	35.92	46.59	51.19	62.78	63.97	71.39
Weak Foot	2.89	2.93	3.10	3.36	3.33	3.30
Value	7.95	6.09	7.49	11.34	11.49	11.49

Table 3: Count values of binary and categorical variables for every sub-group of players (the whole data)

Variable	Level	Center Backs	Fullbacks	Central Defensive Midfielders	Central Midfielders	Wingers	Strikers
EU	1	1136	1194	509	1150	909	777
	0	556	406	304	457	463	449
League	England	353	339	162	334	294	261
	Germany	318	306	186	278	279	259
	Spain	330	335	166	282	285	211
	France	330	346	195	308	293	248
	Italy	361	274	104	405	221	247
National Team	1	354	230	137	360	256	254
	0	1338	1370	676	1247	1116	972
Position	CB	1692	-	-	-	-	-
	LB	-	765	-	-	-	-
	RB	-	780	-	-	-	-
	LWB	-	29	-	-	-	-
	RWB	-	26	-	-	-	-
	CDM	-	-	813	-	-	-
	CM	-	-	-	1196	-	-
	CAM	-	-	-	411	-	-
	LM	-	-	-	-	500	-
	RM	-	-	-	-	513	-
	LW	-	-	-	-	178	-
	RW	-	-	-	-	181	-
	CF	-	-	-	-	-	104
	ST	-	-	-	-	-	1122
Preferred Foot	Left	376	714	108	325	439	194
	Right	1316	886	705	1282	933	1032
# Players		***1692***	***1600***	***813***	***1607***	***1372***	***1226***

3.3.2 Transformations on the Independent Variables

Before building OLS models for each season and playing position, we made a transformation on the data. Instead of using the variable "Age" directly, we used "Squared Age" in our models as its relationship with the player market value was found to be non-linear in previous research [16]. All of the independent variables were then also normalized (scaled between 0 and 1).

3.3.3 Transformation on the Response Variable

We build OLS regression models for each of the 6 playing position in every 4 seasons, which makes 24 models in total. Before building these models, we considered making transformations on the response variable (player market value) as it is highly right skewed.

The first type of transformation considered was using the natural logarithm of the response variable which is a common practice while dealing with a highly skewed response variable and in line with previous research in this field [3][17][18]. So, the resulting linear regression equation would be:

$$ln(MarketValue) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_n x_n + \epsilon \tag{1}$$

We also considered making a Box-Cox Transformation on the market value data for every model. The Box-Cox method suggests a value for the lambda just by taking the response variable as an input [19]. So, the resulting linear regression equation would be:

$$(MarketValue^{\lambda} - 1)/\lambda = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_n x_n + \epsilon, \quad if\ \lambda \neq 0 \tag{2}$$

It should be noted that natural logarithmic transformation is actually a special case of Box-Cox transformation when $\lambda = 0$. The λ values that the Box-Cox Transformation suggested were very close to 0 (between 0.0001 and 0.07) for every model. As a result, we have decided to make natural logarithmic transformation on the response variable for every model built for each playing position in each season. Finally, we have normalized the transformed response variable values.

CHAPTER IV

MODEL BUILDING

4.1 Feature Selection

For each playing position, we had to decide for the set of features to be included in the regression models. This feature selection procedure consists of three parts:

4.1.1 Feature Elimination Based on Correlation

In this part, a stepwise feature elimination procedure is conducted. At each step, a Pearson correlation matrix is calculated and the highest correlation value is detected. From the feature pair that has the highest correlation, the one that has the highest total correlation in the matrix (in absolute value) is deleted. Then, a new correlation matrix is calculated and the whole procedure is conducted again. The feature elimination stops until none of the correlation values is larger than the critical threshold of 0.90 [20]. The pseudocode of the procedure is as follows:

Algorithm 1: Correlation based feature elimination

```
corr = np.abs(X.corr());
np.fill_diagonal(corr.values, 0);
while corr.max().max() ≥ 0.90 do
    index_1 = corr.stack().index[np.argmax(corr.values)][0];
    index_2 = corr.stack().index[np.argmax(corr.values)][1];
    sum_1 = corr[index_1].sum();
    sum_2 = corr[index_2].sum();
    if sum_1 > sum_2 then
        X.drop(index_1, axis=1, inplace=True);
    else
        X.drop(index_2, axis=1, inplace=True);
    end
    corr = np.abs(X.corr());
    np.fill_diagonal(corr.values, 0);
end
```

4.1.2 Feature Selection by Backward Elimination

We conduct logarithmic transformation on the response variable and fit an OLS Regression model on the data using the features selected in the previous part (Feature Elimination Based on Correlation). Then, we conduct a "Backward Elimination" procedure by deleting the variable that has the largest p-value at each step. The procedure is continued until all of the variables left in the model are significant (has a p-value that is less than or equal to 0.10).

4.1.3 Selecting the Common Features

After conducting the previously described two feature elimination procedures, we end up with different sets of features for each of the four seasons. We detect the features that ended up in at least two of the four seasons. These features are used to build models for all of the four playing seasons.

4.2 Model Specification Tests

After choosing the set of features for each of the 6 playing position and building models for 4 seasons, all of the 24 models were subjected to multiple specification tests. We conducted tests for heteroskedasticity, endogeneity, residual normality and multicollinearity. All of the models passed the White Test against heteroskedasticity and Anderson-Darling Test for residual normality at a significance level of 0.05. VIF values, which were checked against any multicollinearity, were all below the critical threshold of 10 [21]. However, 2 of the 24 models have failed to pass the Ramsey RESET Test for endogeneity. One of these models was for Central Defensive Midfielders and the other one was for Strikers. For Central Defensive Midfielders, in addition to the variable “Team Rating”, we also considered including the “Squared Team Rating” as an additional variable in the models. After adding this variable, the model which have previously failed the Ramsey RESET Test was able to pass it. For Strikers, in addition to the variable “Squared Age” we also considered including the “Cubed Age” as an additional variable in the models. After adding this variable, the model which have previously failed the Ramsey RESET Test was able to pass it.

CHAPTER V

RESULTS

For all of the playing positions we conducted feature selection and built models for four playing seasons following the procedure explained in Chapter 4. In line with previous research, we found that "Age" (squared) and "Minutes Played" are significant determinants of the market value ($p \leq 0.01$) for every player sub-group in all four seasons [3]. "Team Rating", which is determined by the average attribute ratings of the players in a team, was also found to be significant ($p \leq 0.10$) in all seasons for every player subgroup. Almost all of the "League" variables were found to be significant ($p \leq 0.10$) determinants of the market value for every player subgroup in four seasons. Additionally, all else being equal, if a player is in English Premier League, then his valuation is higher than the case where the player is in any other league. This is significant for all playing positions. While early studies left the popularity factor of players to the error term (e.g., [13]), there is an "International Reputation" variable in our data, which is a rating that defines how internationally well-known a player is. This variable was found to be a significant ($p \leq 0.10$) determinant of the market value for almost every player sub-group in four seasons. The detailed OLS regression outputs and findings can be found in the following sections.

5.1 Valuation Analysis of Center Backs

Table 4: OLS Regression outputs of Center Backs

Season	2014-2015	2015-2016	2016-2017	2017-2018
Age^2	-0.526***	-0.606***	-0.579***	-0.720***
	(0.025)	(0.028)	(0.023)	(0.029)
Aggression	0.041	0.017	0.061***	0.073**
	(0.034)	(0.032)	(0.022)	(0.033)
Agility	0.068**	0.068**	0.080***	0.019
	(0.029)	(0.030)	(0.022)	(0.025)
Heading Accuracy	0.149***	0.181***	0.104***	0.099***
	(0.040)	(0.040)	(0.031)	(0.036)
Interceptions	0.057	0.039	0.124***	0.225***
	(0.055)	(0.045)	(0.040)	(0.047)
International Rep.	0.038	0.098***	0.063***	0.121***
	(0.024)	(0.025)	(0.021)	(0.029)
League_France	-0.042***	-0.038***	-0.040***	-0.063***
	(0.014)	(0.015)	(0.012)	(0.015)
League_Germany	-0.046***	-0.053***	-0.041***	-0.067***
	(0.014)	(0.014)	(0.011)	(0.014)
League_Italy	-0.074***	-0.070***	-0.066***	-0.084***
	(0.014)	(0.015)	(0.012)	(0.014)
League_Spain	-0.028**	-0.040***	-0.060***	-0.084***
	(0.014)	(0.014)	(0.011)	(0.013)
Marking	0.070	0.141**	0.098*	0.002
	(0.054)	(0.062)	(0.053)	(0.063)
Minutes Played	0.178***	0.149***	0.173***	0.152***
	(0.019)	(0.019)	(0.016)	(0.020)
National Team	0.023**	0.022*	0.017*	0.006
	(0.011)	(0.011)	(0.010)	(0.012)
Sliding Tackle	0.160***	0.101**	0.085**	0.083*
	(0.043)	(0.042)	(0.037)	(0.050)
Stamina	0.087***	0.162***	0.095***	0.098***
	(0.029)	(0.035)	(0.026)	(0.035)
Standing Tackle	0.173***	0.232***	0.153***	0.190***
	(0.054)	(0.077)	(0.047)	(0.070)
Strength	0.116***	0.019	0.066***	0.013
	(0.036)	(0.028)	(0.022)	(0.035)
Team Rating	0.147***	0.139***	0.126***	0.123***
	(0.023)	(0.027)	(0.023)	(0.031)
Intercept	0.070**	0.102**	0.216***	0.235***
	(0.036)	(0.041)	(0.028)	(0.035)
Adjusted R^2	0.888	0.873	0.904	0.864
N	331	328	328	340

Notes: * $p \leq .10$ ** $p \leq .05$ *** $p \leq .01$; standard errors in parantheses

Table 5: Significant Features of Center Backs

Season	2014-2015	2015-2016	2016-2017	2017-2018
Age	***	***	***	***
Aggression			***	**
Agility	**	**	***	
Heading Accuracy	***	***	***	***
Interceptions			***	***
International Rep.		***	***	***
League	***	***	***	***
Marking		**	*	
Minutes Played	***	***	***	***
National Team	**	*	*	
Sliding Tackle	***	**	**	*
Stamina	***	***	***	***
Standing Tackle	***	***	***	***
Strength	***		***	
Team Rating	***	***	***	***
Intercept	**	**	***	***

Both sliding and standing tackling, which are essential defensive abilities, seem to be significant determinants of the market value for center backs in all of the four seasons. Heading accuracy, that is important to clear the ball in defense, was also found to be significant in all seasons. Interceptions, which is another crucial defensive ability, has become significant in the last two seasons. The aggression level has become significant in the last two seasons. Stamina level was found to be significant in all four seasons. The fact that a center back plays for his country's national team and his international reputation rating were found to be important determinants of the market value in most of the analyzed seasons. When the sums of the coefficients of these two features are calculated, it was realized that they are the largest at the end of the 2015-2016 and 2017-2018 seasons that are right before there are World and European Cup tournaments.

5.2 Valuation Analysis of Fullbacks

Table 6: OLS Regression outputs of Fullbacks

Season	2014-2015	2015-2016	2016-2017	2017-2018
Age^2	-0.463***	-0.422***	-0.424***	-0.495***
	(0.034)	(0.025)	(0.026)	(0.028)
Ball Control	0.084	0.133***	0.218***	0.257***
	(0.052)	(0.049)	(0.049)	(0.072)
Crossing	0.052	0.085*	0.072	0.135**
	(0.046)	(0.044)	(0.045)	(0.054)
Interceptions	0.091**	0.093**	0.072	0.154***
	(0.038)	(0.046)	(0.044)	(0.053)
International Rep.	0.091***	0.074**	0.090***	0.092***
	(0.029)	(0.029)	(0.030)	(0.034)
League_France	-0.056***	-0.085***	-0.102***	-0.042***
	(0.014)	(0.014)	(0.014)	(0.016)
League_Germany	-0.061***	-0.090***	-0.089***	-0.063***
	(0.014)	(0.013)	(0.014)	(0.016)
League_Italy	-0.065***	-0.087***	-0.082***	-0.077***
	(0.015)	(0.014)	(0.014)	(0.016)
League_Spain	-0.064***	-0.102***	-0.088***	-0.140***
	(0.014)	(0.014)	(0.014)	(0.015)
Marking	0.109**	0.091**	0.080**	0.070
	(0.043)	(0.041)	(0.040)	(0.062)
Minutes Played	0.197***	0.170***	0.168***	0.197***
	(0.020)	(0.019)	(0.019)	(0.022)
Reactions	0.115***	0.203***	0.151***	0.027
	(0.036)	(0.048)	(0.037)	(0.050)
Sliding Tackle	0.077*	0.081	0.066	0.119**
	(0.041)	(0.050)	(0.043)	(0.047)
Sprint Speed	0.117**	0.071**	0.115***	0.112***
	(0.054)	(0.033)	(0.032)	(0.040)
Stamina	0.095***	0.177***	0.091***	0.150***
	(0.035)	(0.033)	(0.032)	(0.037)
Team Rating	0.171***	0.156***	0.151***	0.236***
	(0.025)	(0.026)	(0.028)	(0.032)
Intercept	0.090**	0.039	0.115***	-0.146***
	(0.037)	(0.026)	(0.029)	(0.046)
Adjusted R^2	0.842	0.857	0.839	0.843
N	302	329	335	323

Notes: * $p \leq .10$ ** $p \leq .05$ *** $p \leq .01$; standard errors in parantheses

Table 7: Significant Features of Fullbacks

Season	2014-2015	2015-2016	2016-2017	2017-2018
Age	***	***	***	***
Ball Control		***	***	***
Crossing		*		**
Interceptions	**	**		***
International Rep.	***	**	***	***
League	***	***	***	***
Marking	**	**	**	
Minutes Played	***	***	***	***
Reactions	***	***	***	
Sliding Tackle	*			**
Sprint Speed	**	**	***	***
Stamina	***	***	***	***
Team Rating	***	***	***	***
Intercept	**		***	***

Both marking and intercepting, which are defensive abilities, seem to be significant determinants of the market value for fullbacks in most of the four seasons. Reactions, which is an ability that defines how quickly the player reacts to certain events around him, was also found to be a significant determinant of the player market value in most seasons. The ball control ability seems to become significant in and gain importance over the last three seasons which means that this skill has become more desired in the player market. Stamina was found to be significant in all of the seasons which makes perfect sense as fullbacks make long sprints and dribbles forward and have to maintain high stamina throughout an entire game. Crossing skill seems to have gained importance over time which might be an indication of the change in the dynamics of football as fullbacks have started to contribute more to attack. It can also be seen that the international reputation of the player was a significant determinant of the market value in all of the four seasons.

5.3 Valuation Analysis of Central Defensive Midfielders

Table 8: OLS Regression outputs of Central Defensive Midfielders

Season	2014-2015	2015-2016	2016-2017	2017-2018
Age^2	-0.465***	-0.501***	-0.460***	-0.464***
	(0.033)	(0.034)	(0.035)	(0.041)
Height	0.072*	0.005	0.015	0.139***
	(0.041)	(0.039)	(0.039)	(0.052)
Interceptions	0.341***	0.190***	0.093*	0.274***
	(0.060)	(0.057)	(0.051)	(0.081)
International Rep.	0.048	0.087**	0.089**	0.090*
	(0.038)	(0.034)	(0.040)	(0.047)
League_France	-0.072***	-0.103***	-0.072***	-0.035
	(0.019)	(0.020)	(0.019)	(0.024)
League_Germany	-0.080***	-0.081***	-0.055***	-0.066***
	(0.018)	(0.020)	(0.018)	(0.022)
League_Italy	-0.056***	-0.016	-0.022	-0.052*
	(0.021)	(0.022)	(0.022)	(0.026)
League_Spain	-0.043**	-0.040**	-0.060***	-0.091***
	(0.020)	(0.019)	(0.019)	(0.023)
Marking	0.102**	0.177***	0.186***	0.054
	(0.047)	(0.044)	(0.048)	(0.072)
Minutes Played	0.227***	0.166***	0.122***	0.175***
	(0.027)	(0.025)	(0.025)	(0.031)
National Team	0.044***	0.035**	0.016	0.010
	(0.016)	(0.017)	(0.015)	(0.020)
Reactions	0.099	0.034	0.159***	0.121*
	(0.061)	(0.063)	(0.045)	(0.065)
Stamina	0.155***	0.142***	0.130***	0.120***
	(0.035)	(0.038)	(0.040)	(0.044)
Strength	0.059	0.063	0.169***	0.115*
	(0.058)	(0.045)	(0.050)	(0.063)
Team Rating	-0.785**	-0.915**	-0.351	-0.498
	(0.327)	(0.385)	(0.395)	(0.451)
Team Rating^2	0.931***	1.021***	0.695*	0.764*
	(0.323)	(0.366)	(0.379)	(0.429)
Vision	0.276***	0.175***	0.201***	0.082
	(0.054)	(0.047)	(0.046)	(0.056)
Weight	-0.068	-0.045	-0.113***	-0.094*
	(0.042)	(0.041)	(0.041)	(0.051)
Intercept	-0.059	0.215***	0.082	0.022
	(0.049)	(0.050)	(0.054)	(0.067)
Adjusted R^2	0.877	0.855	0.884	0.875
N	183	167	160	149

Notes: * $p \leq .10$ ** $p \leq .05$ *** $p \leq .01$; standard errors in parantheses

Table 9: Significant Features of Central Defensive Midfielders

Season	2014-2015	2015-2016	2016-2017	2017-2018
Age	***	***	***	***
Height	*			***
Interceptions	***	***	*	***
International Rep.		**	**	*
League	***	***	***	***
Marking	**	***	***	
Minutes Played	***	***	***	***
National Team	***	**		
Reactions			***	*
Stamina	***	***	***	***
Strength			***	*
Team Rating	**	**		
Team Rating	***	***	*	*
Vision	***	***	***	
Weight			***	*
Intercept		***		

The number of Central Defensive Midfielders is already small and becoming less each season so it might be argued whether it's meaningful to draw conclusions from the table above. Nevertheless, the OLS regression outputs suggest that reacting ability has become significant and more important in the last two seasons. Intercepting and marking, which are key defensive abilities, were found to be significant determinants of the player market value in most of the analyzed seasons. Stamina level of the player was found to be significant in all of the seasons. Vision was found to be significant in the first three seasons which means that central defensive midfielders who can read the game better and have good playmaking skills were more valuable. International reputation of the player has become significant in the last three seasons and has gained importance.

5.4 Valuation Analysis of Central Midfielders

Table 10: OLS Regression outputs of Central Midfielders

Season	**2014-2015**	**2015-2016**	**2016-2017**	**2017-2018**
Acceleration	0.007	0.066**	0.069**	0.064***
	(0.038)	(0.032)	(0.029)	(0.022)
Age^2	-0.428***	-0.424***	-0.433***	-0.415***
	(0.025)	(0.026)	(0.023)	(0.022)
Ball Control	0.155**	0.230***	0.193***	0.160***
	(0.060)	(0.064)	(0.050)	(0.055)
Finishing	0.138***	0.048	0.062**	0.097***
	(0.038)	(0.037)	(0.028)	(0.032)
International Rep.	0.137***	0.061*	0.084***	0.054**
	(0.034)	(0.032)	(0.021)	(0.021)
League_France	-0.043***	-0.039***	-0.059***	-0.050***
	(0.014)	(0.015)	(0.012)	(0.011)
League_Germany	-0.065***	-0.062***	-0.074***	-0.070***
	(0.015)	(0.014)	(0.012)	(0.011)
League_Italy	-0.013	-0.029**	-0.034***	-0.042***
	(0.014)	(0.014)	(0.011)	(0.011)
League_Spain	-0.029**	-0.073***	-0.061***	-0.083***
	(0.015)	(0.015)	(0.012)	(0.011)
Minutes Played	0.163***	0.219***	0.123***	0.156***
	(0.020)	(0.020)	(0.017)	(0.016)
National Team	0.008	0.028**	0.011	0.025***
	(0.012)	(0.012)	(0.009)	(0.009)
Positioning	0.124***	0.088**	0.083**	0.012
	(0.046)	(0.039)	(0.041)	(0.037)
Preferred Foot_Right	-0.017	-0.021*	-0.005	-0.014
	(0.011)	(0.011)	(0.009)	(0.008)
Reactions	0.073	0.107***	0.218***	0.119***
	(0.045)	(0.040)	(0.038)	(0.035)
Short Passing	0.155***	0.111*	0.158***	0.105*
	(0.052)	(0.058)	(0.047)	(0.048)
Stamina	0.192***	0.163***	0.136***	0.094***
	(0.033)	(0.033)	(0.030)	(0.024)
Team Rating	0.156***	0.276***	0.118***	0.184***
	(0.026)	(0.026)	(0.024)	(0.023)
Vision	0.078	0.132**	0.144***	0.176***
	(0.065)	(0.064)	(0.052)	(0.052)
Intercept	-0.015	-0.114***	-0.026	0.087***
	(0.036)	(0.034)	(0.030)	(0.027)
Adjusted R^2	0.877	0.900	0.902	0.899
N	305	312	331	328

Notes: * p ≤ .10 ** p ≤ .05 *** p ≤ .01; standard errors in parantheses

Table 11: Significant Features of Central Midfielders

Season	2014-2015	2015-2016	2016-2017	2017-2018
Acceleration		**	**	***
Age	***	***	***	***
Ball Control	**	***	***	***
Finishing	***		**	***
International Rep.	***	*	***	**
League_France	***	***	***	***
Minutes Played	***	***	***	***
National Team		**		***
Positioning	***	**	**	
Preferred Foot		*		
Reactions		***	***	***
Short Passing	***	*	***	*
Stamina	***	***	***	***
Team Rating	***	***	***	***
Vision		**	***	***
Intercept		***		***

Stamina level, ball control and short passing abilities are found to be significant in every season which is meaningful as central midfielders are playmakers and their main goals are to keep control of the ball and pass it accurately. Vision level was found to be significant in most of the seasons and gained importance each year which means that it has become more valuable that central midfielders can read the game and are aware about their teammates' positioning for passing the ball to them. Acceleration, finishing, positioning and reacting attributes seem to be significant determinants of central midfielders' market value in most of the seasons. International reputation was found to be significant in all seasons which means that popular players are more valuable in the market. The fact whether a central midfielder plays for his country's national team was found to be a significant determinant of the market value in the end of the 2015-2016 and 2017-2018 seasons that are right before the World and European Cup tournaments.

5.5 *Valuation Analysis of Wingers*

Table 12: OLS Regression outputs of Wingers

Season	**2014-2015**	**2015-2016**	**2016-2017**	**2017-2018**
Age^2	-0.435***	-0.326***	-0.337***	-0.396***
	(0.031)	(0.027)	(0.026)	(0.027)
Ball Control	0.178***	0.045	0.151**	0.090
	(0.055)	(0.064)	(0.071)	(0.065)
Crossing	0.076**	0.085**	0.124***	0.058*
	(0.036)	(0.035)	(0.046)	(0.035)
Curve	0.021	0.026	0.000	0.057*
	(0.029)	(0.030)	(0.031)	(0.030)
Dribbling	0.085	0.164***	0.015	0.107*
	(0.054)	(0.061)	(0.064)	(0.056)
Finishing	0.082**	0.260***	0.167***	0.090**
	(0.032)	(0.045)	(0.047)	(0.044)
International Rep.	0.107***	0.140***	0.111***	0.088***
	(0.028)	(0.029)	(0.031)	(0.033)
League_France	-0.055***	-0.048***	-0.061***	-0.031**
	(0.013)	(0.014)	(0.014)	(0.013)
League_Germany	-0.059***	-0.081***	-0.084***	-0.081***
	(0.012)	(0.013)	(0.012)	(0.013)
League_Italy	-0.002	-0.028*	-0.082***	-0.050***
	(0.014)	(0.015)	(0.014)	(0.014)
League_Spain	-0.042***	-0.063***	-0.103***	-0.096***
	(0.012)	(0.012)	(0.013)	(0.013)
Minutes Played	0.184***	0.150***	0.152***	0.168***
	(0.018)	(0.019)	(0.018)	(0.019)
Reactions	0.151***	0.144***	0.222***	0.173***
	(0.038)	(0.043)	(0.054)	(0.045)
Short Passing	0.112***	0.208***	0.128**	0.031
	(0.041)	(0.051)	(0.052)	(0.040)
Sprint Speed	0.081***	0.087***	0.135***	0.098***
	(0.030)	(0.025)	(0.029)	(0.033)
Stamina	0.093***	0.162***	0.135***	0.134***
	(0.030)	(0.033)	(0.033)	(0.031)
Team Rating	0.110***	0.140***	0.133***	0.199***
	(0.022)	(0.026)	(0.026)	(0.028)
Intercept	0.028	-0.163***	-0.110***	0.052
	(0.030)	(0.032)	(0.037)	(0.035)
Adjusted R^2	0.890	0.899	0.893	0.891
N	273	279	271	280

Notes: * $p \leq .10$ ** $p \leq .05$ *** $p \leq .01$; standard errors in parantheses

Table 13: Significant Features of Wingers

Season	2014-2015	2015-2016	2016-2017	2017-2018
Age	***	***	***	***
Ball Control	***		**	
Crossing	**	**	***	*
Curve				*
Dribbling		***		*
Finishing	**	***	***	**
International Rep.	***	***	***	***
League	***	***	***	***
Minutes Played	***	***	***	***
Reactions	***	***	***	***
Short Passing	***	***	**	
Sprint Speed	***	***	***	***
Stamina	***	***	***	***
Team Rating	***	***	***	***
Intercept		***	***	

In each season, either ball control or dribbling skill is found to be significant which is in line with the main goal of the winger players that is to keep control of the ball and carry it forward. Crossing, which is mainly the responsibility of wingers and is crucial as they cross the ball inside the penalty area, was found to be a significant determinant of the market value in most of the seasons. Short passing ability was found to be significant in most of the seasons. Sprint speed, that is important for the wingers as they carry the ball forward and run past the opponent player, was found to be a significant trait in all seasons. Stamina level, that is crucial for the wingers as they make many long runs during a game, seems to be significant in every season. Finishing ability and international reputation were also found to be significant determinants of the market value in every season.

5.6 *Valuation Analysis of Strikers*

Table 14: OLS Regression outputs of Strikers

Season	2014-2015	2015-2016	2016-2017	2017-2018
Acceleration	0.104***	0.134***	0.111***	0.059*
	(0.029)	(0.029)	(0.038)	(0.034)
Age^2	0.689***	0.788**	0.616*	0.594**
	(0.263)	(0.310)	(0.331)	(0.261)
Age^3	-1.224***	-1.306***	-1.291***	-1.174***
	(0.269)	(0.326)	(0.347)	(0.268)
Ball Control	0.033	0.090*	0.189***	0.110**
	(0.046)	(0.054)	(0.056)	(0.049)
Finishing	0.134***	0.158***	0.261***	0.177***
	(0.043)	(0.045)	(0.061)	(0.050)
Heading Accuracy	0.061*	0.074*	0.099**	0.005
	(0.032)	(0.040)	(0.048)	(0.040)
Interceptions	0.036	0.015	0.063**	0.002
	(0.024)	(0.028)	(0.031)	(0.026)
International Rep.	0.090***	0.103***	0.147***	0.125***
	(0.027)	(0.029)	(0.037)	(0.034)
League_France	-0.068***	-0.084***	-0.084***	-0.068***
	(0.012)	(0.015)	(0.017)	(0.015)
League_Germany	-0.067***	-0.088***	-0.082***	-0.076***
	(0.013)	(0.014)	(0.016)	(0.013)
League_Italy	-0.028*	-0.070***	-0.072***	-0.081***
	(0.014)	(0.016)	(0.017)	(0.015)
League_Spain	-0.039***	-0.069***	-0.117***	-0.123***
	(0.013)	(0.016)	(0.017)	(0.014)
Long Passing	0.070**	0.065*	0.046	0.005
	(0.032)	(0.035)	(0.034)	(0.033)
Long Shots	0.016	0.097**	0.067	0.025
	(0.036)	(0.049)	(0.049)	(0.035)
Minutes Played	0.208***	0.161***	0.130***	0.188***
	(0.020)	(0.024)	(0.023)	(0.022)
Positioning	0.235***	0.236***	0.064	0.161**
	(0.056)	(0.062)	(0.069)	(0.065)
Reactions	0.097**	0.041	0.100	0.110**
	(0.047)	(0.058)	(0.061)	(0.055)
Shot Power	0.111***	0.059	0.059	0.086**
	(0.033)	(0.037)	(0.056)	(0.033)
Team Rating	0.150***	0.174***	0.130***	0.179***
	(0.024)	(0.032)	(0.034)	(0.031)
Intercept	-0.070**	-0.022	-0.045	0.136***
	(0.035)	(0.038)	(0.051)	(0.046)
Adjusted R^2	0.906	0.902	0.870	0.889
N	248	226	253	240

Notes: * $p \leq .10$ ** $p \leq .05$ *** $p \leq .01$; standard errors in parantheses

Table 15: Significant Features of Strikers

Season	2014-2015	2015-2016	2016-2017	2017-2018
Acceleration	***	***	***	*
Age	***	***	***	***
Ball Control		*	***	**
Finishing	***	***	***	***
Heading Accuracy	*	*	**	
Interceptions			**	
International Rep.	***	***	***	***
League	***	***	***	***
Long Passing	**	*		
Long Shots		**		
Minutes Played	***	***	***	***
Positioning	***	***		**
Reactions	**			**
Shot Power	***			**
Team Rating	***	***	***	***
Intercept	**			***

In every season, the finishing ability of strikers was found to be a significant determinant of the market value with an increasing importance. Heading accuracy, which is crucial to score goals with the head, was a significant determinant in the first three seasons. These are in line with the main objective of the strikers as they are the main scorers of their teams. Acceleration ability is important as a striker wants to run past the opponent defenders and seems to be a significant determinant of the market value in all seasons. Positioning ability, that is necessary as a striker wants to freely receive passes from his teammates, were found to be a significant trait in most of the seasons. Ball control was found to be significant in the last three seasons. International reputation of a striker seems to be a significant factor of the market value in every season with an increasing effect.

CHAPTER VI

CONCLUSION

Conducting OLS regression analyses on unique datasets of football players played in the top five European leagues, we identified the most important determinants of players' market values throughout four playing seasons. It can be said that this whole research validated the accuracy of the FIFA video game while assigning ratings to different player skills as the determinants of the market value found in our analyses are in line with the dynamics of football. Players playing in different positions can choose to follow training programs in which they focus on improving their skills that were found to be significant and most important determinants of market value in this paper. This can enable players increase their market value quickly. Football teams can also benefit from these analyses by using them as decision support tools while making transfers. With ever-increasing amounts of data, we think that football clubs will gradually adopt data analytics.

Bibliography

[1] Sourav, "Top 10 most popular sports in the world [updated 2020]." Available at https://sportsshow.net/top-10-most-popular-sports-in-the-world/.

[2] E. Amir and G. Livne, "Accounting, valuation and duration of football player contracts," *Journal of Business Finance & Accounting*, vol. 32, no. 3-4, pp. 549–586, 2005.

[3] O. Müller, A. Simons, and M. Weinmann, "Beyond crowd judgments: Data-driven estimation of market value in association football," *European Journal of Operational Research*, vol. 263, no. 2, pp. 611–624, 2017.

[4] M. He, R. Cachucho, and A. J. Knobbe, "Football player's performance and market value.," in *Mlsa@ pkdd/ecml*, pp. 87–95, 2015.

[5] S. Herm, H.-M. Callsen-Bracker, and H. Kreis, "When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community," *Sport Management Review*, vol. 17, no. 4, pp. 484–492, 2014.

[6] EA Sports, "Electronic arts announces multiplatform EA SPORTS FIFA global expansion." Available at https://www.ea.com/news/ea-sports-fifa-global-expansion.

[7] R. Murphy, "FIFA player ratings explained: How are the card number stats decided?." Available at https://www.goal.com/en-ae/news/fifa-player-ratings-explained-how-are-the-card-number-stats/1hszd2fgr7wgf1n2b2yjdpgynu.

[8] G. Kumar, "Machine learning for soccer analytics," *University of Leuven*, 2013.

[9] M. D. Hughes, T. Caudrelier, N. James, A. Redwood-Brown, I. Donnelly, A. Kirkbride, and C. Duschesne, "Moneyball and soccer-an analysis of the key performance indicators of elite male soccer players by position," 2012.

[10] KPMG, "KPMG player valuation tool." Available at https://www.footballbenchmark.com/methodology/player_valuation.

[11] E. Franck and S. Nüesch, "Talent and/or popularity: what does it take to be a superstar?," *Economic Inquiry*, vol. 50, no. 1, pp. 202–216, 2012.

[12] L. Cotta, P. de Melo, F. Benevenuto, and A. Loureiro, "Using fifa soccer video game data for soccer analytics," in *Workshop on large scale sports analytics*, 2016.

[13] F. Carmichael and D. Thomas, "Bargaining in the transfer market: theory and evidence," *Applied Economics*, vol. 25, no. 12, pp. 1467–1476, 1993.

[14] P. Garcia-del Barrio and F. Pujol, "Hidden monopsony rents in winner-take-all markets—sport and economic contribution of spanish soccer players," *Managerial and Decision Economics*, vol. 28, no. 1, pp. 57–70, 2007.

[15] L. Brandes and E. Franck, "Social preferences or personal career concerns? field evidence on positive and negative reciprocity in the workplace," *Journal of Economic Psychology*, vol. 33, no. 5, pp. 925–939, 2012.

[16] A. Bryson, B. Frick, and R. Simmons, "The returns to scarce talent: Footedness and player remuneration in european soccer," *Journal of Sports Economics*, vol. 14, no. 6, pp. 606–628, 2013.

[17] A. T. Yiğit, B. Samak, and T. Kaya, "Football player value assessment using machine learning techniques," in *International Conference on Intelligent and Fuzzy Systems*, pp. 289–297, Springer, 2019.

[18] S. Majewski, "Identification of factors determining market value of the most valuable football players," *Journal of Management and Business Administration. Central Europe*, vol. 24, no. 3, pp. 91–104, 2016.

[19] G. E. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 26, no. 2, pp. 211–243, 1964.

[20] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, *Using multivariate statistics*, vol. 5. Pearson Boston, MA, 2007.

[21] J. F. Hair, "Multivariate data analysis," 2009.