CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Bachelor thesis

2021 **Jan Cvrček**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Jan Cvrček

Determinants of Football Players' Market Value

Bachelor thesis

Prague 2021

Author: Jan Cvrček

Supervisor: Mgr. Roman Kalabiška

Study programme: Economics and Finance

Academic year: 2020/2021

Abstract

This thesis investigates determinants of football players' market value in the top 5 European leagues. It focuses on the differences among defenders, midfielders and forwards. Moreover, it extends the existing knowledge by delving into the unexplored world of goalkeepers. Using the ordinary least squares method on a sample from season 2018/19, it finds several significant factors, such as goals, assists and passes accuracy. The results show that defenders seem to receive more credit for just joining the match than midfielders and forwards, indicating that the latter group is thereby expected to bring added value on the pitch. Furthermore, goalkeepers seem to reach their turning point at the age of 22.4, which is similar to the field players. Nevertheless, the peak was anticipated to be distinctly higher for goalkeepers, making this outcome surprising. Lastly, the set of the significant variables explaining the goalkeepers' market values comprises, for instance, received goals to 90 minutes ratio and team rank at the end of the season, while the proportion of successful saves turned out to be insignificant. Therefore, their market values appear to be driven to a greater extent by the overall team performance than by the statistics directly related to them. All the findings were subject to the robustness check, which suggested no significant bias by the outliers' effect.

Keywords

Football, Market Value, Top 5 European Leagues, OLS, Comparison

Abstrakt

Tato práce zkoumá faktory ovlivňující tržní cenu fotbalových hráčů v 5 nejlepších evropských soutěžích. Zaměřuje se na rozdíly mezi obránci, záložníky a útočníky. Kromě toho také rozšiřuje dosavadní výzkum o dosud neprobádaný svět brankářů. Za pomoci metody nejmenších čtverců na vzorku dat ze sezóny 2018/19 nachází několik vysvětlujících faktorů, například góly, asistence a úspěšnost přihrávek. Výsledky naznačují, že obránci jsou více odměňováni za pouhé odehrání zápasu než záložníci a útočníci, od kterých se tedy spíše očekává přidaná hodnota na hřišti. Dalším zjištěním je, že brankáři dosahují svého zlomového věku ve 22.4 letech, což je srovnatelné s hráči v poli. Nicméně dle prognóz se čekalo, že brankáři dosáhnou vrcholu mnohem později, proto je tento výsledek překvapivý. V množině relevantních proměnných k jejich tržní ceně se nachází například podíl obdržených branek na 90 minut a konečné postavení týmu v tabulce, zatímco procento úspěšných zákroků se zdá být nesouvisející. Z tohoto důvodu se jeví jejich tržní ceny více závislé na celkovém výkonu týmu než na jejich vlastních statistikách. Všechny tyto výsledky byly podrobeny testům robustnosti, které ovšem nenaznačily, že by je extrémní hodnoty výrazně zkreslily.

Klíčová slova

Fotbal, Tržní cena, Top 5 evropských lig, OLS, Porovnání

Declaration of Authorship

I hereby proclaim that I wrote my bachelor thesis on my own under the leadership of my supervisor and that the references include all resources and literature I have used.

I grant a permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, May 3, 2021

Jan Cvrček

Acknowledgments

I am immensely honored to express my profound gratitude to Mgr. Roman Kalabiška for his incessant willingness to help and for providing me with above standard supervising work. Nevertheless, I would like to thank my family, colleagues and many faculty teachers for helping me and motivating me throughout the study at the institute.

Bibliographic note

Cvrček, Jan. *Determinants of Football Players' Market Value*. Prague, 2021. (58 pp). Bachelor thesis (Bc.), Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis Supervisor Mgr. Roman Kalabiška

Bachelor's Thesis Proposal

Author	Jan Cvrček
Supervisor	Mgr. Roman Kalabiška
Proposed topic	Determinants of Football Players' Market Value

Research question and motivation:

I am going to be analyzing statistics and characteristics of soccer players in the top 5 European leagues by investigating significant drivers determining their market value, such as goals scored, assists, cards received and so on. The main research question is therefore going to be, whether there can be found these factors or not and if so, then which are the most influential ones. Furthermore, I will be testing other possible hypotheses stemming from the statistics, for instance, a correlation between their market value and their salaries. Nevertheless, I will also try to study a progression of the market values during a player's career.

This study can definitely be found interesting from an economic perspective. The soccer market is an extraordinary market with a comparison to classical market known in economy. For example, there exists an economic inefficiency regarding racial discrimination (Blaha, 2017). Therefore, this phenomenon is a subject of economic discussion quite frequently. Since the media contributed to the popularization of football, wages have grown rapidly and thus a lot of finance is centered in this sport (Dobson Stephen, Goddard John, 2001)

Contribution:

There already exist some studies regarding the market value and performance of football players but my thesis will uniquely focus on the top 5 European leagues and will further dive deep into other possible correlations. On the contrary, the existing papers study a single league (Miao He, Ricardo Cachucho and Arno Knobbe, 2015) or only the most valuable players (Sebastian Majewski, 2016) and do not provide us with other correlations. The results could be used not only by scouts and team managers but also by fans themselves, since football is popular all over the world. Nevertheless, this thesis can be also informative for economists as an example of an extraordinary market.

Methodology:

I am going to be working with the dataset obtained from sport webpages such as transfermarkt.com and whoscored.com most probably by using the web scraping technique to directly extract the data. The football leagues will be studied not only together but also separately to find out possible differences in the determinants regarding the significance in each country. I will analyze it using a proper regression that will be specified later. Ordinary Least Squares, Neural Networks or Random Forest Regression are the most probable ones to be considered. I will be using programs such as R and Python to analyze the data and to test hypotheses which are mentioned in the research questions or some others emerging during the work on this thesis.

Core bibliography:

Blaha, J. (2017). An Empirical Investigation of Wage Discrimination in Professional Football. Bachelor thesis, Charles University

Dobson, S., Goddard, J. A., Dobson, S. (2001). The economics of football (Vol. 10). Cambridge: Cambridge University Press.

He, M., Cachucho, R., Knobbe, A. J. (2015, September). Football Player's Performance and Market Value. In Mlsa@ pkdd/ecml (pp. 87-95).

He, Y. (2012). Predicting market value of soccer players using linear modeling techniques. University of Berkeley (working paper).

Majewski, S. (2016). Identification of factors determining market value of the most valuable football players. Journal of Management and Business Administration. Central Europe, 24(3), 91-104.

Poldaru, R., Roots, J., Viira, A. H. (2005). Artificial neural network as an alternative to multiple regression analysis for estimating the parameters of econometric models. Agronomy research, 3(2), 177-187.

Table of Contents

List of Figures

List of Tables

1 Introduction

As sport has been incessantly becoming a substantial part of our daily lives, it has also begun to attract a wide range of audience, among which we can find sports analysts as well. They have striven to explain the background of some athletes' success. Since football is said to be the largest sport on the planet, it could be argued that it is also the focal point of all the analyses that have been conducted. This, however, has not always been the case. For example, in the USA, where the first papers arose, baseball, basketball and American football received more credit in terms of the statisticians' interest.

Yet, towards the end of the twentieth century, the emergence of academic works about footballers took place with the scientists first investigating possible explanators of the transfer fees. Due to the lack of data availability at that time, major variables were general team-based rather than performance-oriented. Thanks to the later growth of technology and the internet, various statistics were collected and published with a higher frequency. It resulted in the origin of market values as a proxy for the transfer fees. Thenceforth, studies have been mostly done using this data as a regressand since it facilitates a greater number of observations.

Therefore, there is a plethora of articles examining footballers and their economic power in the present moment. This could potentially be very useful for a variety of reasons. It may provide us with a comparison of the general economic market and the football one, for instance, the latter's extraordinariness. Furthermore, determining the driving factors of a footballer's market value might serve for the clubs to seek underrated players on the market, as it was the case in the famous movie Moneyball. Nonetheless, the usage of such findings could also rest in reaching broad fans, which could thereby be dragged closely into the game by being aware of the concrete answers to what is behind the market values.

Nevertheless, the vast majority of those articles are based on the best football players globally and forwards for their relative ease to assess. Less famous leagues, different positions and comparative analyses have been mostly disregarded. Hence, in this thesis, we want to develop something that could widen the knowledge gained so far. Thus, on a sample from the top 5 European leagues in season 2018/19 using the ordinary least squares method (OLS), we try to delve into the comparison of the factors' significance and slope for defenders, midfielders and strikers, which has been faintly described in the literature. The research question is whether such factors exist, and if it is the case, how do they differ for each position.

Moreover, this thesis uniquely dives deep into the overlooked world of goalkeepers. To the best of our knowledge, this might be the first work to analyse their market value's determinants. Again, we hope to find some of them to explain the dependent variable.

At the beginning, we introduce the existing literature together with a brief overview of football history. Later, we describe the data-obtaining process as well as an explanation of some variables' choice. In addition, we include summary statistics for field players. The following section consists of the methodology description and our model specification. Further, we present the OLS results for field players, their robustness and a short comparison between the French and English leagues. Then, a section examining goalkeepers takes place. It encompasses their analysis, model description, the OLS results and the robustness check. At the end, we discuss the implications and conclude.

2 Current State of Knowledge

This section consists of 4 parts and its goal is to provide a detailed description of the existing literature regarding the determinants of football players' market value. The first one (2.1) focuses on a general overview of the research articles on this topic. The second one (2.2) studies papers with transfer fees as a dependent variable. In the third one (2.3), we delve into market values and the methodology of determining them by a website *www.transfermarkt.com*. Finally, the last part (2.4) encompasses pieces of research using market values as an explained variable.

2.1 General Overview

Sport has existed in the world for centuries. Some could argue that the beginning dates back to the Olympic Games in Ancient Greece. Others claim that we should consider the 19th century when the major sports disciplines emerged as the starting point. Yet, after World War II, the first academic papers primarily in the USA have started to study the economics behind collective sports (Dobson and Goddard, 2001). With growing popularity, especially thanks to the media, the sport has begun to be a centre of finance and football is no exception (Dobson and Goddard, 2001).

In football particularly, when a team desired to obtain a player from a different club, no matter whether his contract still held or not, it had to pay a transfer fee to the selling club. This was true until 1995, when a resolution called *the Bosman ruling* was decided, resulting in players leaving their club after the contract had expired without anyone paying a fee. Many economists and statisticians have therefore striven to determine the factors having an influence on the height of the transfer fee. There was, however, a significant lack of information availability in football compared to other sports (Erkmen et al., 2010). The majority of researchers in this time thus concentrated on the actual transfer fees. They mainly examined the clubs' positions, or in other words, their bargaining power since detailed players' games statistics were not known to them (Frick, 2007).

This has changed after several servers launched the market and started to provide a broad audience with statistical data of games and players. One of the currently leading ones can be considered a German website *www.transfermarkt.com* founded in 2000 that, among other services, publishes football players' market values. Since it rose in popularity, many researchers, managers and football fans have started to quote it (Herm et al., 2014). Its usefulness was proven, as the market values significantly correlate with players' salaries and experts' estimates (Torgler and Schmidt, 2007).

2.2 Transfer Fee Research

As mentioned in the previous section, examiners had to deal with little to no data availability because the statistics were not published or even measured. Carmichael and Thomas (1993) were among the first ones to regress a logarithmic transformation of a transfer fee on several explanatory variables with the help of the OLS method. Using data from the top 4 English leagues in season 1990/91, they found a few significant ones, such as the positive impact of the average attendance of the buying club in the previous season and a negative one of the division (1-4) both clubs played in. This meant that the better the league tier, the higher the fee.

On a sample of 202 observations from the best English leagues, Reilly and Witt (1995) further conducted an analysis based on the same OLS approach and contributed with a significance of more player-based variables. Those were, for instance, games played, goals scored or a dummy of being a forward. Nevertheless, these two research papers, as well as Speight and Thomas (1997), also aimed to study the influence of an arbitration procedure imposed by several countries to help resolve the disputes. Their findings, however, suggest contradictory results whether the arbitrary fees are greater or lower than the transfer fees agreed both by the selling and buying part.

Coming back to the determinants, the results of Reilly and Witt (1995) go in line with Dobson and Gerrard (1999), who above these findings add a significance of positive age and negative age^2, meaning that, *ceteris paribus*, the transfer fee peaks when a player reaches a certain age and then starts to plunge. Moreover, they also expand the existing literature by a positive influence of both international caps and under-21 international caps. The results seem to be more reliable, as the authors used more than 1000 observations.

A very similar conclusion draw Carmichael et al. (1999). They, however, argue that since the probability of being transferred is not equal for all players, the estimation conducted by the precedent authors brings biased coefficients. Therefore, they come up with a Heckman two-step procedure to control for it but bring about similar products. Furthermore, Dobson et al. (2000) take into consideration transfers in semi-professional English leagues in seasons from 1988 to 1997. Using OLS and again a log-level model, they reveal the significance of both selling and buying club attendance having a positive effect on the transfer fee. An interesting point from their study can also be found a negative impact of the stadium capacity of the buying team, in contrast to the positive influence of the latter team field's number of seats. This may be the first time that a stadium size appeared to play a role. Unfortunately, they only assume 114 observations, which raises doubts.

In their paper, Ruijg and van Ophem (2015), similarly to Carmichael et al. (1999), claim that the analysis is partially invalid because of the selectivity bias and hence in addition to OLS use a probit ordered estimates model to deal with it. They observe very similar results as the previous researchers. Moreover, they include a new variable called *Golden Substitution*, a fraction of substitutions in[1] and scoring a goal and the total number of substitutions in. This variable, however, appears to be significant only in the OLS model and not in the one uniquely used by these authors.

A further study was carried out by Ante (2019), who focused on the top 5 European leagues in the summer part of season 18/19. Among the most interesting findings, one can find a significant negative effect of weight on transfers into the German Bundesliga and the Spanish La Liga. Regarding a particular position on the field, a positive influence of weight was found for defenders and a negative one for forwards. Nevertheless, arguably their most remarkable factor is the player's popularity in terms of social media accounts like Instagram, Facebook or Twitter.

Related to this thesis, the most common method of estimation used in the relevant literature mentioned above is the log-level OLS. The only problem some of the researchers stated was that it could bring biased results due to the transfer selection. Fortunately, this will not cause us troubles since we use market values instead of the transfer fees. They also revealed some significant variables. In particular, goals, age and club ranking seemed to best predict the fees. Further, appearances in both the international and U-21 games, together with the popularity on social media, also delivered gratifying results and could serve as potential statistics for our hypotheses.

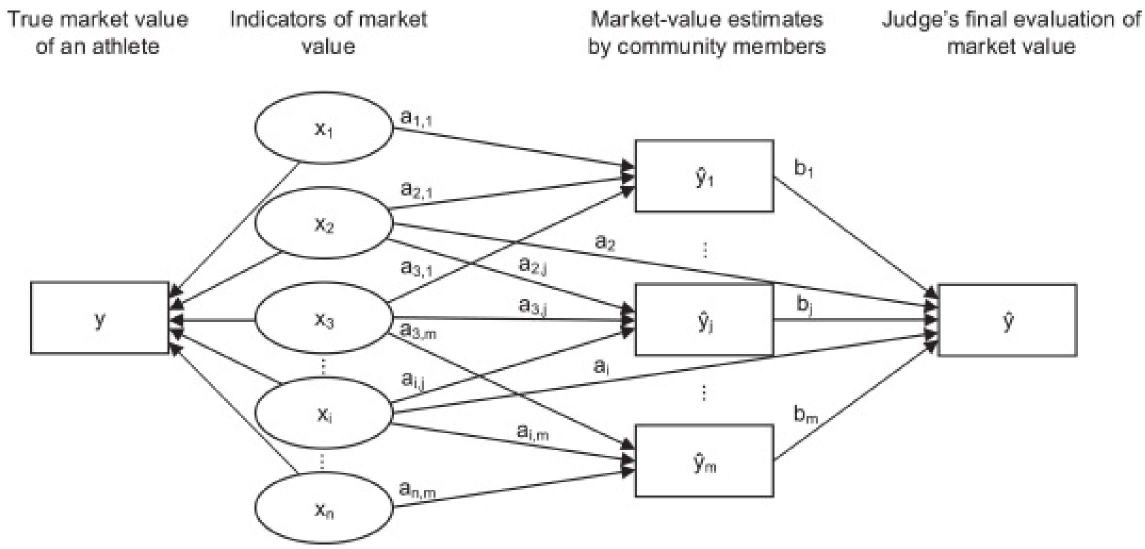
[1] A substitution in refers to a situation when a player starts on the bench and during a game goes on a pitch in exchange for a different player.

2.3 Transfermarkt Methodology

This work will be based on market values, whose height usually slightly differs from the agreed transfer fee. According to Herm et al. (2014), it is an estimate of the amount for which a team can sell the player's contract to another team. Knowing this value can therefore be informative for agents when negotiating a transfer. Secondly, it can also be useful for team managers to immediately find out which players could be potentially feasible to obtain. Last but not least, it also provides millions of football fans worldwide with these pieces of information they can speculate about and thus drags them into the game.

The most quoted website facilitating us to obtain the data is *www.transfermarkt.com* (Herm et al., 2014). It uses a so-called "wisdom of crowd", a term first mentioned by Surowiecki (2005). It consists in an idea that a fans' voice can be more efficient than the one provided by a qualified authority. Rather than relying on equality and taking a sample median or mean of all the voters' estimates, transfermarkt's methodology is based on a few experts, sometimes also referred to as judges, who give different users different weights based on their qualification in terms of experience and previous accuracy. Thus, the result is an aggregate mean of these weighted values (Herm et al., 2014). The estimation procedure is illustrated in the following diagram.

Figure 1: Transfermarkt Market Value Estimation



Source: Herm et al. (2014)

This process brings perks as well as drawbacks. Considering, for instance, purely known football players, it may happen that only a few votes are submitted. A single democratic procedure could lead to biased results since users may not be sufficiently experienced. In a worse scenario, some agents could deliberately influence the height of the value for their own purposes. Hence, weighting each input is a way to mitigate the impact of these unwanted effects and is thus more efficient (Müller et al., 2017).

On the contrary, many drawbacks emerge too. One might ask who checks the judges since they have the final word. Furthermore, as long as there are a lot of players to evaluate, it is necessary to collect a large number of data. This, however, takes some time, which results in little flexibility in updating the statistics - approximately every 9 months (Müller et al., 2017). Lastly, Herm et al. (2014) argue that well-known players tend to receive more precise values than lesser-known ones due to a comparatively higher demand for estimating it from the users' side.

There have been plenty of studies to test the validity of this website. Baan (2016), in his work, concludes that the crowd's estimation performs better than less sophisticated benchmark methods. These results align with research conducted by Peeters (2018), who aimed to test the transfermarkt's assessment on particular games. In the end, he finds out that it appears to be more precise than standard predictors like ELO rating or FIFA ranking. Moreover, it even turns out to bring financial gains when adequately applied to betting strategies.

2.4 Market Value Research

Market values allow for including more observations in the analyses, as they are estimated for the vast majority of players. Thereby, there are many articles not only discussing them but also using them as a response variable while regressing it on several predictors directly connected to the player that are published more and more often. The first factors which could come to everyone's mind could be the performance statistics on the field, such as scored goals or minutes played. Yet, it is necessary to consider the specific tasks a single player is given by his managers that vary among goalkeepers, defenders, midfielders and forwards. In order to reach the best result, each position is required to contribute to the team's achievement in a different way.

Goalkeepers should primarily protect their own goal and give other players information about the opposing team's line up, as they enjoy the best view from the back.

Nevertheless, goaltenders, as they are sometimes also referred to, play a more active role during the game nowadays in terms of having a significant amount of passes with defenders.

Moving along to them, one can further generally differentiate between left/right-backs and centre-backs. The former ought to be agile, have an excellent physical capacity and possess a remarkable ability to pass long-distance balls accurately, whereas the latter should be strong to win tackles, have a good overview of the game and be ready to accept responsibility due to being usually the last ones in front of their netkeepers.

In the middle of the pitch, we can find midfielders. Again, it can be vaguely distinguished between centre-ones and wings, playing mainly near the side lines. The first ones' frequent asset is a quick evaluation of a situation while being under pressure, distributing precise passes among team members and excellent technique. Regarding the wide midfielders, their position is very similar to the side-backs discussed earlier, i.e. to have a good stamina and high maximum speed. Moreover, they are often required to be creative when going one on one against the opponent.

Finally, at the front of the pitch, we can find forwards. Their concrete tasks may differ in particular clubs depending on the strategy, but their assessment criteria are usually scoring goals or eventually assisting to them. This facilitates academics to test their market value determinants more easily compared to other types of players, where a wider range of factors plays a significant role.

One of them is Majewski (2016), who collected data of the top 150 strikers in the world and tried to explain their market value. His findings imply that Canadian points comprising goals and assists have a strong accelerating impact. Additionally, he created new, unique variables to test their possible influence and revealed that a fraction of team market value and nationality ranking is significant in the regression and increases the player's value. Consequently, speaking in terms of a single footballer, if it is the case that he belongs to a rich club and comes from the best countries according to the FIFA ranking, it has a positive effect and thereby increases his market value. Moreover, he uses a dummy of being among the top 5 players and shows its implication on the regressor that is almost a 40 million euros growth. He claims that the reason for that could be a "goodwill" or brand these footballers bring with them.

In their article, He et al. (2015) aim to determine these factors for forwards in the Spanish La Liga. Unlike Majewski (2016), they use more methods, including a LASSO regression that seemed to fit the best. After setting the optimal threshold, they find several significant variables. Surprisingly, assists appear to contribute with the highest weight, followed by goals. This could be caused by the fact that goals are split into the ones scored from the penalty area and other ones outside it. Regarding their slope, the latter have a stronger influence than the former, which might be explained by the difficulty of scoring a goal that is generally directly proportional to the distance from the net. On the other hand, the only significant and negative variable was the number of fouls.

A further study on player positions was conducted by Richau et al. (2019), who used Boosted Regression Trees on a sample of 1897 observations from the English Premier League and excluded goalkeepers for their lack and difficulty to assess. The results vary by each position, but the most significant factor for all three of them is the average final rank of the team in the last three seasons a footballer plays or played in. Speaking of forwards, the second-highest factor is, for some people expectedly, goals. It is followed by passes and shots. the number of assists takes up the sixth place and, according to the findings, contributes approximately five times less than goals. This contradicts the implications by He et al. (2015), who suggest that assists have a higher effect than goals.

Moving to midfielders, as spoken at the beginning of this section, their main requirements are habitually passing and assisting to goals. This goes in line with the results of Richau et al. (2019), as assisting to a goal is the second important performance statistic after the team ranking. Next in line is the number of duels followed by the number of passes. In contrast to forwards, goals are not that valued for midfielders since they are roughly five times less determining than assists.

Finishing the implications of Richau et al. (2019), defenders' market values are majorly affected by the team rank, as it makes up more than a third of all the variables, which is the most among these three categories. Far behind it can be found age and, surprisingly, the number of shots. Nevertheless, this conclusion suggests that for defenders their performance on the pitch is quite difficult to evaluate quantitatively.

A research paper by Felipe et al. (2020) concentrates on the top 5 European leagues comprising the English, French, German, Italian and Spanish. Using the OLS, they observe that playing in the Champions League enhances the perceived value as well as a presence in the Europa League. The latter has anticipatedly a lower effect. They also find out there indeed exist differences throughout the competitions. Compared to the Premier League, playing in any of the remaining ones diminishes the market value. Spanish La Liga is situated in second place, followed by Serie A and the Bundesliga, while the French Ligue 1 is at the very bottom. Felipe et al. (2020) argue that there might be an explanation for it in terms of the heights of the television rights, which correspond to the market values by each league. Regarding a single position, stunningly, we cannot find forwards at the top because, according to the coefficients, attacking midfielders seem to receive the highest credit.

Apart from the performance on the pitch, we can also study the effect of the players' characteristics. The first one to come to one's mind is age. Young footballers are expected to improve with gaining experience, while older ones might not be that perspective at all. It could be therefore anticipated that, *ceteris paribus*, there exists a negative quadratic relationship between age and market value, meaning that to some point, the value rises and after reaching the peak, it begins to decline. Bryson et al. (2013) confirm this theory by allowing age and age^2 terms in the regression and finding the maximum point in the mid-twenties. Sinčák (2020) reveals that in the Czech Fortuna League, for instance, this point is at 26.75, which might imply that this competition is slow and thus helps relatively experienced players to stay still productive.

Another determinant is height. It can be argued that it increases the probability of scoring a goal and at the same time, it prevents one from receiving primarily during corners and free-kicks (Fry et al., 2014). Bryson et al. (2013) find height to significantly determine salaries, but it is not the case for the market values. They theorise it may be caused by little variation in the variable.

Academic papers have also focused on footedness and its possible influence. Bryson et al. (2013) concluded that if a player is able to control the ball by both of his feet and is therefore to some degree indifferent which one to use, it positively raises his salaries. Herm et al. (2014) approve this theory on the height of the perceived values. They state that the reason for this may also lay in the fact that it enables these players to perform in more positions during a match, making them more flexible.

As there exists discrimination in football (Blaha, 2017), some articles aimed to study the impact of nationality. In their work, Garcia-del-Barrio and Pujol (2007) draw a conclusion that non-Spanish European footballers were overrated, in contrast to the non-European ones, which were estimated to possess a lower value than they ought to. Medcalfe (2008), on the contrary, does not confirm this theory on a sample from the English league and thus, these two findings do not go in line with each other in general.

The research question of Korzynski and Paniagua (2016) in their paper is to what extent social media play a role in the market position of the top football players. Their empirical findings suggest that media activity, followers on Twitter especially, as well as good performance on the field are both necessary, however, insufficient conditions for high market values. Frenger et al. (2019) examine a very similar task. They use data from the Bundesliga at the end of season 2017/18 and collect the number of followers on Instagram, Twitter and Facebook for each player. Their first hypothesis that the social media "portfolio size" is influenced by age is rejected. Further, they test the effect of single media channels on the market value and draw a conclusion that only Instagram popularity is relevant and positively elevates the dependent variable.

To summarise the existing articles, we can conclude that the topic of the market values' determinants in football is thoroughly examined. Most of the research is based on the OLS and some pieces contain advanced methods but arrive at similar findings. The papers mainly study defenders, midfielders and forwards, but overlook goalkeepers, which opens us the door to delve into their analysis. The frequently used variables are goals and assists that are mostly significant, but their weights usually differ. For this reason, it cannot be concluded which one is more determining. Among other useful factors can be found age, passes, shots and tackles. The results also vary for the three field positions, as every single one requires different evaluation criteria. Regarding specific competitions, the Premier League seems to employ players with the highest market values. Showing up in the Champions and Europa League, however, has a very beneficial impact for all footballers, no matter in which country they operate. Finally, social media activity appeared to be significant as well. We will thus use this acquired knowledge and strive to extend the existing literature by analysing the differences among the positions, including the heavily sidelined goalkeepers.

3 Data

3.1 Sources

The data for this thesis were obtained from 3 sources. The vast majority was downloaded from *www.fbref.com*, an American website providing its users with sports statistics. It was founded in 2000, only focusing on baseball. Nowadays, its portfolio encompasses five sports altogether, including football. The goal is, according to the website, the following: *"We strive to work with respect, reliability with oomph, and craftsmanship, and also to promote the democratisation of sports data."* We gathered a number of information primarily related to the games' statistics.

The second source of our data was obtained from *www.transfermarkt.com*. We concentrated on the market values this website offers to view and received them through a web scraping method. Further, we used the fact that for each player, there is also his nationality and preferred foot and downloaded both of them to enrich our dataset.

The third part is a website *www.fifa.com*, where the men's nationality football ranking was scraped. These three parts were processed and merged in the programming language Python and subsequently analysed in R.

3.2 Data Choice

The following subsection describes the data obtaining process. Firstly, due to the Covid pandemic hitting the life on Earth at the beginning of 2020, the majority of sports events were postponed or even cancelled. Unfortunately, football was no exception. The aim of this thesis was to take a deep look at the top 5 European leagues, but the French one was cancelled in this season (2019/20) and the others suffered a long delay. Thus, it would be very difficult to conduct a study on such a sample. Hence, we decided to perform our analysis on a season when the matches really took place and therefore used the results from 2018/19 with all leagues finishing standardly in time. The market values were taken from the end of this season.

Secondly, we collected statistics on players who appeared at least once in the game and did not experience a transfer in order to have unbiased results.

Thirdly, since goalkeepers require special treatment and their performance criteria differ from other positions, we devoted a single section (6) for their analysis, where we present the data, methodology and results.

Fourthly, we also had to deal with the issue of choosing the appropriate number of position levels for the field ones. As described in the literature review, defenders can be divided into the centre and side ones and midfielders into defensive, offensive, centre and side ones. It would be, however, highly demanding or almost impossible to differentiate among those categories, as other factors like tactics and opponents' strength also play a role. Moreover, it happens pretty frequently that players change their positions, e.g. from side to centre ones. Therefore, it cannot be determined which is the dominant one. Thus, we will use only three levels and concentrate on defenders, midfielders, and forwards.

Furthermore, another problem was merging the 3 data frames since some missing data appeared in there. If it was the case, the observation was dropped.

Ultimately, the German Bundesliga has only 18 teams, while the other four have 20 of them. The players' absolute statistics from this league were weighted by $(x_i/34) \times 38$ in order to put the leagues on the same level, as there are 34 games in Germany and 38 in the other four leagues. As a result, it may cause some integers to be non-integers.

3.3 Data Characteristics - Field Players

After cleaning the rough dataset, we were left with 1565 observations. Initially, we will pay attention to the categorical ones and then deep dive into their numerical neighbours.

Categorical Variables

As described in the previous sections, we have three positions and five leagues. The most frequent level is defenders in the Premier League, where there are 140 of them. On the other hand, we can find Italian forwards with only 73 observations at the bottom. Detailed statistics can be seen in the following Table 1.

Table 1: Frequency by Leagues and Positions

	English	German	Spanish	Italian	French
Defender	140	114	107	117	131
Midfielder	109	105	104	101	105
Forward	103	91	91	73	74

Source: Author's computations

Regarding the preferred foot, most players have the right one as dominant. Among defenders, we can find a relatively very high number of left-footed footballers. On the contrary, the position where two-footed ones are most frequent is forwards. In the regression, we will use a dummy $both\ feet$, containing 1 if indifferent and 0 otherwise. A complete distribution is illustrated in Table 2.

Table 2: Frequency Positions and Foot in %

	Defender	Midfielder	Forward
Both	1.0	3.6	4.9
Left	33.2	20.8	22.0
Right	65.8	75.6	73.1

Source: Author's computations

Finally, we created a binary variable top 10 to capture the effect of the ten best players. Many academicians argue that these players contain hidden variables that accelerate their market values even though they do not attain diametrically better results. Among them, we might find, e.g. their brand as an attraction for fans. For this reason, it could work as a proxy for social media activity and fame in the regression.

Numerical Variables

Probably the most informative to come to one's mind is the number of goals scored in the whole season. Scoring a goal should contribute to the team's success with a significant weight since it not only brings the side closer to the victory, but it can also serve as a motivation and hence boost the confidence of the teammates. A positive sign of the coefficient is thus expected. Nevertheless, it is also necessary to consider the number of opportunities a player is given. Therefore, we use the efficacy of scoring as a goals-to-shots ratio to help efficient footballers and devalue the less economical ones.

Proceeding, another offensive factor is the number of assists. Creating a chance for a different player to attain a goal is sometimes more demanding than the actual scoring, which may result in the players having the most assists being valued higher than the scorers (Sinčák, 2020). Another key factor, highly connected to assists, is the percentage of successful passes, which, in our opinion, could be a core variable for midfielders. Again, a positive effect is anticipated.

Further, as players are getting experienced and physically and mentally stronger, it ought to influence their performance in a positive way and enhance their values. We believe, however, that at some point, it changes and, on average, they will not be competitive due to the declining body condition. This might lead to the employers' conviction that they are no longer deemed perspective and their younger colleagues would be preferred. To capture this supposed non-linear relationship, we also include age^2 beside age in the regression.

A similar non-performance characteristic is height. In general, it could be the case that the taller one is, the better for him when in a head tackle. On the contrary, small players tend to have better stability thanks to the centre of gravity being lower. They usually play as centre-midfielders, where their agility is highly required. It will be thus interesting to find out which effect will outweigh the other.

Consequently, the more time a player is on the field in the domestic league, the more he is visible for the scouts, managers and fans. Thus, including this effect in the equation might control for it. We have collected information on the matches each observed footballer appeared in and expect that a positive sign of the coefficient will occur.

Completing a tackle illegally has an ambivalent effect. On the one hand, it means that a player appeared in the duel too late and gives the opponent an opportunity to score from a free-kick. On the other hand, team managers appeal to players to stop some dangerously looking counter-attacks by "applying the parking brake", meaning that they want their team players to rather foul than letting the other team score. Fouls are, therefore, an interesting variable to take into consideration. For defenders and midfielders, we suppose a positive coefficient, while forwards are expected to have a negative one.

If it happens, on the other hand, that a tackle is successful and a player steals a ball from the opponent, it may be game deciding because it enables fast breaks. For this reason, we include a variable interceptions and believe it positively explains the market value. We expect midfielders to have the most significant coefficient due to their crucial position in the middle of the pitch.

Inspired by the previous researchers, we use some other variables, such as the final ranking of the team in the domestic league, to capture the team's performance throughout the whole season. Next, as Majewski (2016) came up with a mixed variable team market value/nationality ranking and showed its positive significance, we will proceed in the same way and create it. It will be denoted as TN Ratio. The nationality ranking is as of June 14, 2019, and has Belgium in the first place and then decreases unit-by-unit down to San Marino (211). Thus, this new variable will be higher for rich teams employing players from the top nations.

Apart from these characteristics, we have also collected performance data on the players in the Champions League (CL) and Europa League (EL). The former is the most prestigious club competition in the world, involving 32 top teams each year. It is broadly watched by media around the world and hence, a single appearance in it may surge up a player's market value. The second competition is intended for the clubs in the second wave, as there are other 48 teams. Nonetheless, it also encompasses football scouts and thereby attracts millions of fans. Playing in either of these brings an enormous amount of money for the clubs as well.

We possess data on the number of matches played in both of these competitions. For those players not competing in them, we just use 0 in these variables. We believe that these coefficients will be positive and thus, there will be a positive correlation between each of them and the market value.

The following Table 3 provides the summary statistics of these numerical variables together with the dependent one.

Table 3: Summary Statistics - Numerical Variables

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Market Value (€Th.)	1,565	13,247.160	19,707.840	100	2,500	15,000	200,000
Age	1,565	25.567	4.210	15	22	29	39
Height (m)	1,565	1.815	0.063	1.620	1.770	1.860	2.000
Goals	1,565	2.454	3.908	0	0	3	36
Assists	1,565	1.763	2.353	0	0	2.2	16
Matches	1,565	23.121	9.982	1	16	32	38
Goals/Shots (%)	1,565	0.083	0.115	0.000	0.000	0.130	1.000
Passes Completed (%)	1,565	77.445	8.483	12.500	72.500	83.400	100.000
Interceptions	1,565	15.706	13.083	0	5	23.5	76
Fouls	1,565	22.574	15.362	0	11	32	89
CL Matches	1,565	1.077	2.535	0	0	0	13
EL Matches	1,565	0.861	2.327	0	0	0	15
Team Market Value (€Th.)	1,565	329,207.800	299,230.300	54,400	113,250	385,825	1,203,450
Team Rank	1,565	10.165	5.706	1	5	15	20
Nationality Rank	1,565	17.989	23.343	1	4	17	176
TN Ratio	1,565	66,313.460	124,746.900	383.807	8,631.818	65,625.000	1,203,450.000

Source: Author's computations

Field Positions Comparison

Furthermore, we will now go deeper and examine the statistical differences among the three categories. At first glance, it can be noticed that forwards enjoy the highest market values while defenders are at the bottom. Together with the large standard deviation, this may support some theories of overpaying the strikers.

Surprisingly enough, the position having the most assists is not midfielders but forwards. It could be explained by the latter almost always being somewhere around when a goal is scored. Oppositely, there are usually five midfielders in a game per one side and they cannot have that many assists on average since they cover a wider part of the pitch.

Lastly, it can also be found interesting that the percentage of completed passes is the highest for defensive players. A possible explanation may lay in the fact that most of their passes are safe, usually among themselves. However, the other two positions are forced to initiate more risky/creative passes and thus tend to fail more often. A more detailed comparison is contained in Table 4.

Table 4: Field Positions Comparison

	Defender - M	Defender - SD	Midfielder - M	Midfielder - SD	Forward - M	Forward - SD
Market Value (€Th.)	10,619.540	13,738.700	13,252.190	18,792.740	16,945.260	26,289.290
Age	26.089	4.033	25.309	4.219	25.146	4.375
Height (m)	1.833	0.060	1.800	0.061	1.807	0.064
Goals	0.878	1.233	1.802	2.339	5.273	5.704
Assists	1.110	1.673	1.901	2.450	2.370	2.611
Goals/Shots (%)	0.076	0.134	0.065	0.090	0.114	0.107
Passes Completed (%)	80.020	7.244	79.342	7.582	71.514	8.268
Fouls	20.268	12.237	24.811	16.778	21.551	16.296
Interceptions	20.985	13.245	16.032	12.502	6.764	6.566

Note: M = Mean; SD = Standard Deviation
Source: Author's computations

4 Methods

This section examines the methodology for the field players. Goalkeepers' model will be analysed in section 6 for continuity purposes.

In the literature, researchers mostly used the OLS method to deal with this question. The ones that conducted their analyses with the help of more advanced models did so due to their validity fears in the case of transfer fees or because they wanted to tackle the issue from a different perspective. They, however, reached very similar conclusions regarding the influence of the variables. Our data frame consists of several variables obtained at a specific point of time that were measured during the whole season. We, therefore, possess a cross-sectional sample (Wooldridge, 2016). To address our research questions, we decided to use the advantages of the OLS method since it is the most suitable approach regarding our data. Other procedures, for example, a more complicated LASSO, could also perform solidly, but we see no necessity of deciding to go with them, as they are more data scientific than econometric.

There are a few assumptions needed to be discussed in order to have unbiased and consistent estimates (Wooldridge, 2016).

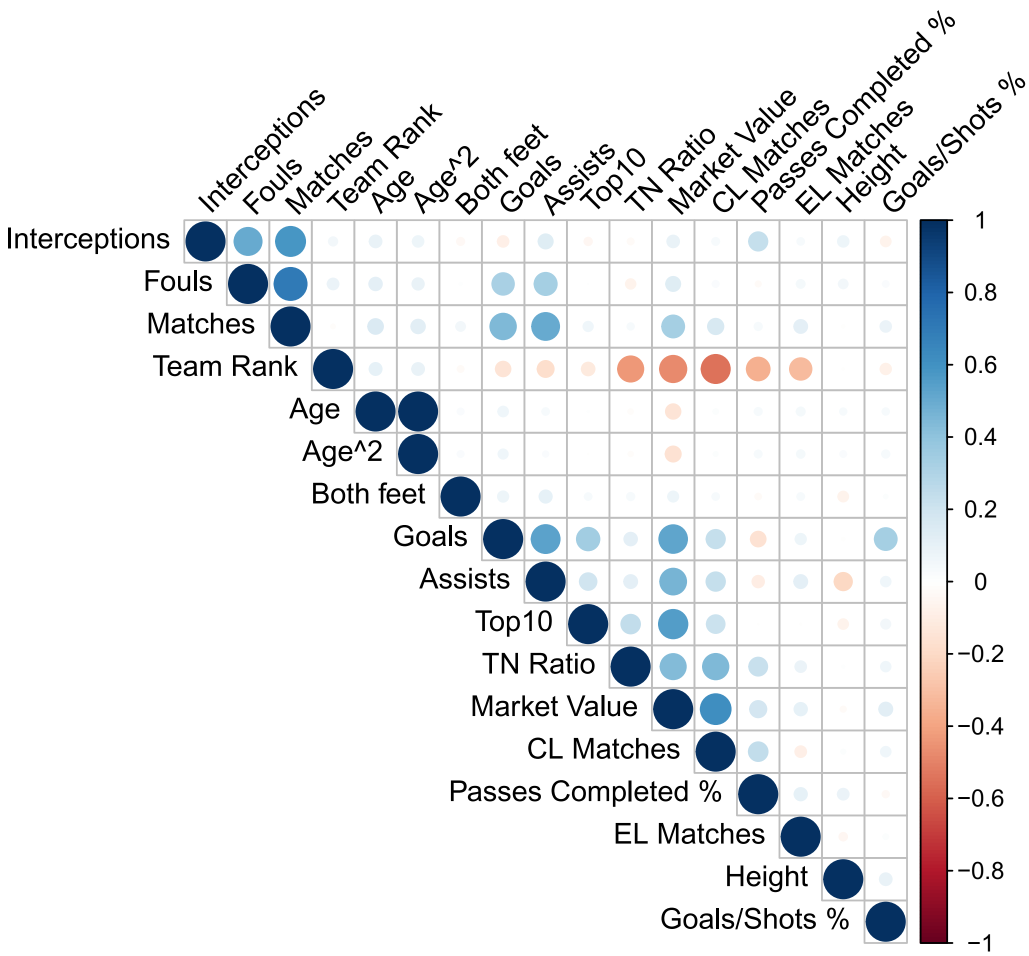
The first one is linearity in parameters. Since we expect this kind of relationship between our regressand and regressors, we can state that this assumption is satisfied.

The second one is the condition to have random sampling. This one is satisfied too because we randomly selected our data that matched our filter, starting with all the players with their market values estimated.

Thirdly, we are interested in whether (1) every regressor is constant and (2) there exist perfect linear relationships among the independent variables. The answer to the former issue is negative, as it can be noticed from Table 3 that the values vary. To verify the latter, we need to look at Figure 2, located on the next page, where the relationships of the variables used in the regression are presented in a correlogram. It is obvious that no single pair is equal to 1 with the exception of age and age^2, which are, however, correlated by definition. Subsequently, the Variance Inflation Factor (VIF), used to test possible multicollinearity, yielded the very same results since no other value was greater than 10, a threshold suggested by Wooldridge (2016). Therefore, we can conclude that this assumption holds.

We believe that all the variables are exogenous in the model, but this assumption is difficult to verify. As discussed in section 3, where the data were introduced, we suppose that all independent variables influence the dependent one and not the other way round and thus do not enter the model as endogenous. Hence, all the first 4 MLRs hold, implying that our regression ought to produce unbiased estimates (Wooldridge, 2016).

Figure 2: Correlogram of the Variables - Field Players



Source: Author's computations

4.1 Model Specification - Field Players

As a result, we want to estimate the following equation using the OLS. We will run this process four times and always on a different sample. Starting on defenders, followed by midfielders and forwards and finishing on all these three positions combined for comparison purposes. With respect to the Breusch-Pagan test, all the models suffer from heteroscedasticity. Thus, we use robust standard errors. The following equation represents the model we want to estimate.

$$
\begin{aligned}
log(Market\,Value_i) =& \beta_0 + \beta_1 Age_i + \beta_2 Age_i^2 + \beta_3 Height_i + \beta_4 Goals_i + \\
& \beta_5 Goals/Shots_i + \beta_6 Assists_i + \beta_7 Passes\,Completed_i + \\
& \beta_8 Fouls_i + \beta_9 Top10_i + \beta_{10} Interceptions_i + \beta_{11} CL\ Matches_i + \\
& \beta_{12} EL\ Matches_i + \beta_{13} Matches_i + \beta_{14} Team\ Rank_i + \beta_{15} TN\ Ratio_i + \\
& \beta_{16} Both\ Feet_i + \beta_{17} French\ League_i + \beta_{18} German\ League_i + \\
& \beta_{19} Italian\ League_i + \beta_{20} Spanish\ League_i + u_i
\end{aligned}
$$

Coefficient β_0 is not of our interest too much since we are more curious about the partial effects of each variable. Following the logic in subsection 3.3, coefficients β_1 and β_2 should represent an inverse U-shape, also known as $-x^2$. Therefore, we expect the former to be positive and the latter negative. Coefficients β_3 to β_7 are supposed to be positive. In contrast, β_8 ought to negatively explain the dependent variable, while β_9 to β_{13} should be positive in our eyes. The sign of β_{14} is anticipated to be negative and β_{15} positive.

Regarding the last four coefficients, we expect all of them to be negative because the Premier League is known for employing the most valuable players.

Regarding each position on the pitch, we do not expect any contradictory findings in terms of the coefficients' signs, with the exception of fouls (3.3). Since we use a logarithmic transformation of the response variable, the interpretation of the coefficients is following. A unit increase in the k-th independent variable corresponds to the $\beta_k * 100$ percentage change of the dependent one.

5 Results

Table 5: OLS - Field Players

	Dependent variable:			
	log(Market Value)			
	Defenders	Midfielders	Forwards	All
Age	0.577***	0.529***	0.518***	0.545***
	(0.085)	(0.079)	(0.077)	(0.047)
Age^2	−0.013***	−0.012***	−0.012***	−0.012***
	(0.002)	(0.001)	(0.001)	(0.001)
Height	1.306***	0.814*	−0.655	0.460*
	(0.447)	(0.466)	(0.584)	(0.271)
Goals	0.073***	0.075***	0.054***	0.059***
	(0.023)	(0.012)	(0.009)	(0.006)
Goals/Shots	0.156	0.286	0.301	0.249
	(0.196)	(0.360)	(0.479)	(0.168)
Assists	0.036***	0.059***	0.062***	0.053***
	(0.014)	(0.011)	(0.012)	(0.007)
Passes Completed	0.013***	0.020***	0.013***	0.017***
	(0.004)	(0.004)	(0.005)	(0.002)
Fouls	0.002	−0.001	0.003	0.002
	(0.003)	(0.002)	(0.002)	(0.001)
Top10		1.381***	−0.277	−0.007
		(0.201)	(0.242)	(0.291)
Interceptions	0.003	0.009***	0.005	0.005***
	(0.003)	(0.003)	(0.005)	(0.002)
CL Matches	0.070***	0.059***	0.057***	0.060***
	(0.012)	(0.014)	(0.015)	(0.008)
EL Matches	0.013	0.009	0.009	0.010
	(0.009)	(0.011)	(0.012)	(0.006)
Matches	0.040***	0.035***	0.036***	0.038***
	(0.005)	(0.005)	(0.006)	(0.003)
Team Rank	−0.085***	−0.079***	−0.082***	−0.082***
	(0.006)	(0.008)	(0.008)	(0.004)
TN Ratio	0.00000*	0.00000**	0.00000	0.00000***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Both Feet	−0.194	−0.132	0.066	−0.032
	(0.176)	(0.190)	(0.103)	(0.098)
French League	−0.970***	−0.999***	−1.042***	−1.009***
	(0.075)	(0.082)	(0.087)	(0.046)
German League	−0.726***	−0.834***	−0.764***	−0.766***
	(0.075)	(0.080)	(0.090)	(0.046)
Italian League	−0.790***	−0.699***	−0.662***	−0.719***
	(0.070)	(0.088)	(0.107)	(0.050)
Spanish League	−0.612***	−0.554***	−0.646***	−0.609***
	(0.076)	(0.081)	(0.100)	(0.048)
Constant	−0.574	0.392	3.485**	0.996
	(1.355)	(1.359)	(1.613)	(0.784)
Observations	609	524	432	1,565
R^2	0.783	0.801	0.818	0.795
Adjusted R^2	0.776	0.793	0.810	0.792
Residual Std. Error	0.587 (df = 589)	0.619 (df = 503)	0.619 (df = 411)	0.608 (df = 1544)
F Statistic	112.049*** (df = 19; 589)	101.015*** (df = 20; 503)	92.633*** (df = 20; 411)	299.265*** (df = 20; 1544)

*N*ote: *p<0.1; **p<0.05; ***p<0.01
*N*ote: Robust standard errors in parentheses
*S*ource: Author's computations

5.1 Discussion

Based on the results, it can be noticed that all the R-squared values as well as the adjusted R-squared values are similar. Most of them are centred around 0.8, suggesting a solid performance of the models. The highest score belongs to forwards, while the lowest one was achieved by defenders. This finding supports the theories from the literature review that it may be more straightforward to explain the market values of strikers compared to other positions, although the R-squared measures differ only marginally. The number of observations is varying, but for each regression it is sufficient since the least one is 432, which is deemed enough.

Starting with age, it is apparent that the normal and quadratic terms are always significant. We can compute the peaks from an expression $\frac{Age}{-2Age^2}$. The quotients can be viewed in Table 6.

Table 6: Peak Age (Years)

Defenders	Midfielders	Forwards	All
22.1	21.7	22.4	22.0

Source: Author's computations

The peak for all the field players is 22 years. *Ceteris paribus*, the market value rises up to this age and when reaching it, it starts decreasing. Thus, it exhibits a U-shape, which is in line with our expectations. Regarding the differences among the positions, midfielders seem to mature as the first ones. Forwards, on average, may enjoy more than an 8-month delay. It could be explained by the tasks these positions are given. Strikers usually do not need to return to the defence and are thereby subject to more lenient stamina requirements than midfielders or defenders. Therefore, age is a lesser hindrance for them.

The third variable in the regression is height. Except for forwards, the coefficient is significant. It is positive and appears to be more determining for defenders than midfielders, as the former have a higher slope and the level of significance. The units are metres; hence if they were 1 centimetre taller, their value would increase by 1.3% and 0.8%, respectively. Again, we maintain that it is logical since players of both these positions find themselves very often in head tackles, whose failure could have devastating effects. For strikers, this is not essential because the game tactics are usually set based on that specific type of the striker. Some are shorter forcing their team to play low and quick passes while having tall forwards enables kicking the ball towards them in the air. Therefore, the height is not such an obstacle for them.

Arguably, the factor that first comes up in everyone's mind is goals scored. For all the columns, a significant and positive coefficient has occurred. A surprising fact might be that defenders and midfielders have a higher gradient than forwards. This may be caused by the latter scoring more often than the former. Therefore, if a midfielder or defender score, it is a bonus they may gain from, not a necessity. On the other hand, strikers are expected to score and due to their usually high number of goals, an additional one does not have the effect to enhance the value that much. Since the ratio of goals versus shots appeared to be insignificant, we can state that, other things being equal, scoring a goal increases the market value by 7.3%, 7.5% and 5.4%, respectively.

A similar conclusion can be drawn about assists, i.e. significant and positive in all regressions. Forwards enjoy the steepest slope, closely followed by midfielders. For the first group, it is worth noticing that the coefficient is higher than in the case of goals. On the contrary, midfielders' slope is flatter than the one for scoring. Both these findings are against our initial beliefs. One possible explanation could be that forwards score relatively easy goals that midfielders or defenders hardly prepare. Midfielders' goals, however, may be more demanding, as they are inclined to shoot from a long distance, and strikers are considered altruistic when assisting to a goal. Speaking in particular terms, an assist for a defender increases his value by 3.6%, for a midfielder, it is 5.9% and finally, a forward receives the highest credit for his value is 6.2%.

The next determinant is the proportion of completed passes. All the coefficients are significant and positive. If forwards and defenders were to have one more percentage point, their value would equally grow by 1.3%. Midfielders' coefficient, however, is more significant, contributing with 2%. This implies that this variable is more explicative for midfielders. It goes in line with the logic in the data description where passes were deemed to have the highest effect on midfielders.

Another regressor in the table is the number of fouls. We supposed that it would explain forwards in a negative way, whereas the two other positions would be influenced positively. Surprisingly, it was found insignificant in all cases. Hence, according to our findings, it does not impact the values.

Further, the dummy variable Top10 was expected to enhance the market values. It only worked for midfielders, who are thus the only ones affected. It more than doubles their value. Among the top 10 players, we cannot find any defender. For this reason, the variable is not considered in their regression. Forwards seem to remain untouched by this explanator for the insignificance of the coefficient. Consequently, finding oneself at the top is only beneficial for midfielders. The concealed factors in this variable, such as media popularity or goodwill, play a role only for them.

Proceeding to the next driver, we encounter interceptions. Again, midfielders possess a significant coefficient, while defenders and forwards do not. This goes in line with our expectations from section 3. The reason may be the midfielders' crucial position on the pitch. They mainly operate in the middle and once they intercept a ball, they may go into attack. Thus, it is more of a focal point in the spectators' eyes than for the other two levels. One additional interception increases their explained variable by almost 1 per cent.

Moving along to the number of matches, we can spot that the Europa League is insignificant in all models. For the top 5 leagues it is still deemed a bit inferior, but for less developed leagues, it may serve as a good way for the footballers to get into the range of the prestigious clubs and agents around the world. On the contrary, the Champions League appears to be a powerful predictor. Interestingly enough, defenders earn the steepest gradient, whilst midfielders and forwards experience a similar but flatter one. The very same results are in the case of the domestic league matches. In other words, defenders obtain the highest coefficient once again, whereas the rest get similarly lower ones. A conceivable explanation would be that for defenders, it is ample to just appear on the field and play their standard. Forwards and midfielders, on the other hand, are supposed to perform uniquely, to stand out compared to the others. Therefore, a sole appearance is a necessary but insufficient condition for them. As it could be anticipated, the UEFA Champions League impacts the market values with a bigger weight than the home ones. In general, the former improves the market values by 6% and the latter by 3.8%.

Staying for a moment within the domestic leagues, another variable is team rank. In all the models, it was negative and significant. A worsening in the final table by one place implies a decrease by approximately 8% in the dependent variable. The sign was expected beforehand to favour the clubs whose season was successful. However, 8% is indisputably a tremendous value, which was not anticipated for a single one-place movement.

The TN Ratio was supposed to be positive and significant. We can see that it is the case, although with only a marginal effect. We are, however, more curious about the significance. Defenders and midfielders meet this condition, while forwards, albeit close, do not. Altogether, the coefficient is significant and positive. We can therefore conclude that we arrived at similar findings and that the TN Ratio has definitely something to do with the regressand. For this reason, employing rich players from the best countries increases the market value. Nevertheless, due to the partial significance the results do not hold for forwards. Moreover, we would probably need a greater significance in the other cases as well to produce strong claims.

Furthermore, we can observe the insignificance of the variable both feet. This outcome is a bit surprising, as we believed it would yield better effects since two-footed footballers are without a shadow of a doubt an admirable acquisition for the clubs.

The final pieces of information concern the differences throughout the five leagues. As the English Premier League is the base, we only include the four others in the models. With respect to the results, we can state that the English league contains the most valuable players since all the coefficients are both significant and negative. In terms of the market values, the French league seems to employ the worst players, no matter the position. The most eminent jump is in the case of forwards with a slope -1.04. In contrast, Spanish La Liga always ranked second place but with a decent distance behind the base. The last couple of leagues, which is the German and Italian one, attained similar outcomes. However, the former has overall a steeper gradient, suggesting that its footballers have, in general, lower values. These findings are interesting because it has been controlled for other factors in the regressions, yet such huge differences have still occurred. This indicates a dominant position of the Premier League, which may attract cheaper players from foreign competitions. The results align with Felipe et al. (2020) and our theories from section 3.

Lastly, the constant term is significant only in the model concerning forwards. As it was informed in the methodology section, this coefficient is not of our interest so much since we seek to observe the individual effects of each variable.

5.2 Robustness Check

We also aim to verify our models by checking their robustness. After plotting the densities of the residuals (Figure A.1 - Appendix), we see that the last model behaves best in terms of normality. The first three do not perform poorly either, but two of the graphs clearly do not look like equal distributions at first glance.

We, therefore, decided to rerun the models on a narrower dataset by eliminating the outliers. These were detected by computing the interquartile range (IQR) and finding the market values' first and third quartiles. Any observation lying above the third quartile + 1.5*IQR or below the first quartile - 1.5*IQR was dropped. After disregarding 167 observations altogether, we were left with 1398. This step has altered the summary statistics of the market values, as the mean value has plunged from €13 247 160 to €7 854 934. The regression table can be viewed in Table A.1 in the appendix.

First of all, the models' performance experienced a decline. Both the R-squared and the adjusted R-squared measures diminished in all models by roughly 8%, indicating that the initial ones worked better. Single coefficients are, however, similar. Worth commenting is the position of midfielders. They became most rewarded for scoring and assisting among the three levels. Further, the variable Top10 already included a defender and appeared to be significant and positive, while it lost its significance for midfielders.

Regarding the matches, these adjusted models behaved in a similar manner with the exception of the Europa League games for defenders. It was shown significant and positive, yet enhancing the value less than the domestic leagues. The rest did not differ dramatically, only the English Premier League strengthened its dominant position compared to the other competitions and the German Bundesliga became slightly better than the Italian Serie A.

To sum up the robustness check, we plotted the residuals and reran the models without outliers. The former revealed that the error terms are not perfectly distributed, but at the same time, it did not confirm that the models would be surely erroneously specified. The latter showed that subsetting the dataset did not utterly change the coefficients, meaning that the implications hold for players in both sets. This suggests that the effects of performance characteristics of the outliers did not distort the results presented in subsection 5.1.

5.3 Leagues Comparison

The English and French leagues seemed to be the most different pair among the five ones. In England, a mean market value is €20 400 000, while in France, it is €8 400 000. Therefore, this short section will be devoted to their regression analysis to determine how the drivers differ for both competitions.

The two models are specified in the same way as the previous ones, with two exceptions. First, we will now have a variable top 5 instead of top 10. We believe that including the effect of only the top 5 players is more representative for such decreased samples. Second, we will not use the categorical variables leagues for apparent reasons. The numbers of observations are 352 and 310, which makes both models still valid.

Lastly, we use the robust standard errors due to the occurrence of heteroscedasticity in both models with respect to the Breusch-Pagan test.

Results

The regression results are situated in Table 7 at the end of this section. The R-squared measures are comparable to the ones for the field positions, that is around 0.8 or a little lower. The Premier League, however, possesses a greater value, which makes it a better fit.

Moving to the analysis of the variables, we may analogously find the peak age. It is equal to 22.4 years for the English league and 22.0 for the French one. Hence, the former appreciates a bit older players.

Height, as a next factor, appeared to be a good predictor only for the French league for its significance. It increases the market value by 1.3% if 1 centimetre taller. In England, this variable is not considered relevant.

Goals, however, seem to be an appreciable explanator. In both leagues, they are found significant and positive. In France, they elevate the dependent variable by 4.7% and in the UK, it is 5.8%.

A relative measure goals to shots was shown insignificant. The shooting efficiency, therefore, does not play a role. In contrast, assists are vastly welcome in both competitions. In France, the influence on the regressand is way more, as it rises by more than 11%. In the Premier League, players enjoy only a 4.3 per cent growth.

Passes completion was found both positive and significant in the two models. The effect is approximately 2% for both competitions.

Fouls, on the contrary, do not explain the market values in either of the leagues. This finding is in line with the ones for the positions.

The variable top 5 appeared to be significant only for the players in the Ligue 1. Thanks to its remarkable slope, it almost doubles the market value. It implies that belonging to the top 5 footballers in France vastly enhances the market value, whilst in England, it is not the case. This might suggest that superstars in France are far above the average players in this league. In contrast, top athletes in England do not attain that exaggerated market values in comparison to the mean footballers of this competition.

Interceptions appeared insignificant in both regressions. Therefore, they do not seem to be a good predictor.

Matches turned out to be extremely interesting. While in England, a single game in the Champions League enhances the market value by only 3.7%, in France it is twice this value. It fosters our reasoning from subsection 5.1 since the French Ligue 1, as the poorer one in this case, emphasises more on the participation in the most renowned international competition than the Premier League does. Although the difference is not huge, it still suggests a presence of such a relationship.

The Europa League is relevant only to the Premier League, where it mounts up the dependent variable by 1.7%.

Stepping on the field and playing a game in domestic competitions is significant and positive in both columns. In France, the slope is a bit steeper than in England.

Team rank was found negative and significant in both leagues. A drop by a single position would cause a decline of 6.6% and 5.5% for England and France, respectively.

The TN Ratio appeared to be significant and positive for both countries. Hence, being in a club that employs great footballers and coming from a successful country increases the market value.

Ultimately, the ability to control the ball by both feet does not seem to be informative, in regard to the market values in either of the leagues.

Table 7: OLS - English and French Leagues

	Dependent variable:	
	log(Market.Value)	
	English	French
Age	0.620***	0.531***
	(0.093)	(0.107)
Age^2	−0.014***	−0.012***
	(0.002)	(0.002)
Height	0.546	1.255**
	(0.418)	(0.616)
Goals	0.058***	0.047***
	(0.009)	(0.015)
Goals/Shots	0.055	0.121
	(0.287)	(0.477)
Assists	0.043***	0.113***
	(0.011)	(0.016)
Passes Completed	0.019***	0.017***
	(0.005)	(0.005)
Fouls	0.0005	0.004
	(0.002)	(0.003)
Top5	0.198	0.878**
	(0.317)	(0.348)
Interceptions	0.003	0.003
	(0.002)	(0.003)
CL Matches	0.037***	0.074***
	(0.010)	(0.023)
EL Matches	0.017*	0.025
	(0.009)	(0.018)
Matches	0.027***	0.032***
	(0.005)	(0.007)
Team Rank	−0.066***	−0.055***
	(0.007)	(0.009)
TN Ratio	0.00000***	0.00000***
	(0.00000)	(0.00000)
Both Feet	0.070	−0.157
	(0.099)	(0.296)
Constant	−0.029	−1.550
	(1.455)	(1.721)
Observations	352	310
R^2	0.801	0.775
Adjusted R^2	0.791	0.762
Residual Std. Error	0.492 (df = 335)	0.585 (df = 293)
F Statistic	84.017*** (df = 16; 335)	62.962*** (df = 16; 293)

*N*ote: *p<0.1; **p<0.05; ***p<0.01
*N*ote: Robust standard errors in parentheses
*S*ource: Author's computations

6 Goalkeepers

This section will be devoted to the goalkeepers' analysis, making us the pioneers who delve into the explanation of their market values. Firstly, we will provide data statistics of the variables, then move to the model description and finally comment on the regression results.

6.1 Data Characteristics - Goalkeepers

In this section, we aim to introduce our variables that could potentially explain the market value of goalkeepers. Many of them are identical to the field players. There are, however, some new ones too. The data obtaining and cleaning process was similar to the one in the previous section.

Age is anticipated to be less constraining for goalkeepers since their body decline does not dramatically limit them to save the shots. Experience is, on the other hand, more welcome. We, therefore, expect the peak in the regression to be above the one estimated for the field players.

Moving to height, taller keepers are prone to perform better when catching high crosses, corner kicks and jumping towards shots. In contrast, it can be argued that being smaller might improve quick reactions and reflexes. We believe that the first advantages will outbalance the other and the coefficient will be above zero.

Another explanator could be the matches played. Being visible for the football world and getting experience is intrinsic for all players.

The most determining factor is probably the number of goals received. Of course, it is not self-explaining, as there are other aspects like the strength of the defence that play a role, but for a football club, it is considered to be essential. We, however, use goals to 90 minutes ratio (denoted as Goals/90) to capture a more precise value.

Saves, as a further statistic, are in some eyes equally crucial as the two precedent factors. Nevertheless, we include the relative measurement of saves to shots, which could be more informative in our opinion.

For a few years now, there has been a trend that goalies are required to possess a skill to play with their feet confidently since they are becoming a greater part of the game. Hence, to include this phenomenon, we use a relative variable passes percentage (without kick-offs) to favour the keepers whose ability to control the ball is high.

Similar to the field players, we also gathered information about the matches in the two European leagues. We suppose a positive relationship with the market value.

In the same way as for the field players, we also include the team rank. The data are presented in Table 8, together with the market value.

Table 8: Summary Statistics - Goalkeepers

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Market Value (€Th.)	145	8,826.897	16,530.230	150	800	10,000	100,000
Age	145	28.097	4.678	19	25	31	40
Height (m)	145	1.895	0.045	1.800	1.860	1.930	2.010
Matches	145	18.912	14.185	1	4	34	38
Minutes	145	1,686.883	1,279.025	50	352.1	3,060	3,420
Goals Received	145	25.282	19.424	0	5	42	78
Goals/90	145	1.462	0.728	0.000	1.060	1.750	6.000
Saves	145	54.374	43.182	0	10	96	147
Saves (%)	145	68.643	13.190	0.000	65.000	75.000	100.000
Passes Completed (%)	145	48.525	17.360	2	36.8	58.6	96
CL Matches	145	0.966	2.556	0	0	0	13
EL Matches	145	0.469	1.659	0	0	0	12
Team Rank	145	9.910	5.621	1	5	15	20

*S*ource: Author's computations

Comparison

We also provide a simple comparison between goalkeepers and field players. Regarding the market value, the first group has substantially less than the second one. The standard deviations of both of them are very high from our point of view. The reason may rest in the fact that there are huge differences among the five leagues and also because the top players tend to be overpaid.

Further, one can see that the mean ages are also different by more than two years. This vaguely approves our thought that keepers mature later. Height will also be subject to a closer look in the regression because for the goalkeepers, the average is greater by 8 centimetres. Lastly, the variable passes percentage is not directly comparable since field players are more often on the ball. In spite of that, we maintain that it could be quite interesting to see how the numbers differ. A more detailed comparison can be observed in the following Table 9.

Table 9: Position Comparison

	Field Players - M	Field Players - SD	Goalkeepers - M	Goalkeepers - SD
Market Value (€Th.)	13,247.160	19,707.840	8,826.897	16,530.230
Age	25.567	4.210	28.097	4.678
Height (m)	1.815	0.063	1.895	0.045
Passes Completed (%)	77.445	8.483	48.525	17.360

Note: M = Mean; SD = Standard Deviation

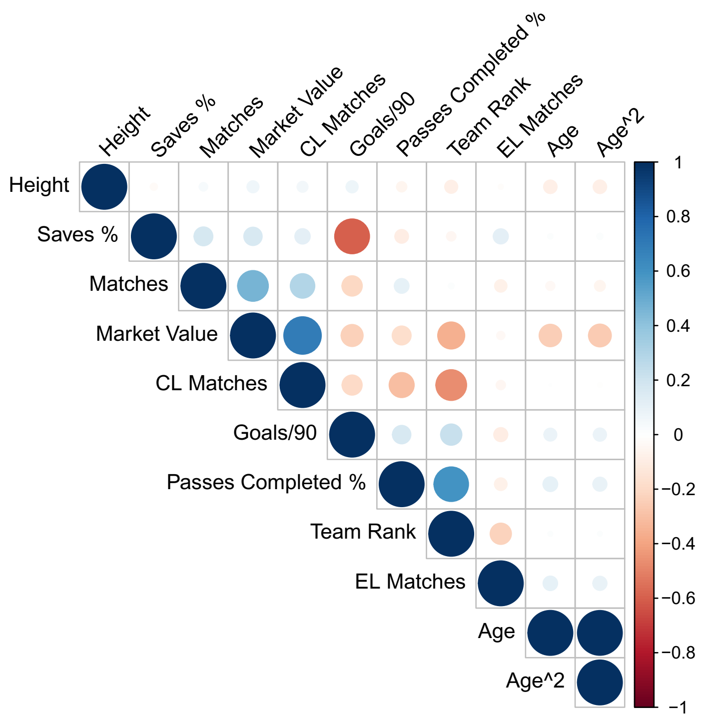
Source: Author's computations

6.2 Methodology

Since goalkeepers have not been practically analysed at all, we cannot be inspired by any model choices from the previous studies. We, however, drew a conclusion to use the OLS approach once again in that we believe it is an appropriate one.

Our variables were subjected to the same choice procedure as the field players. The 4 MLRs, therefore, hold from identical principles. We also provide a correlogram of the variables situated in Figure 3. Some factors introduced in the summary statistics were not used to avoid multicollinearity or for econometric reasons.

Figure 3: Correlogram of the Variables - Goalkeepers



Source: Author's computations

In that case, our estimates are unbiased. The only concern at the moment is the possibility of heteroscedasticity. After running the Breusch-Pagan test, the p-value turned out to be equal to 0.08, making the rejection of the null hypothesis dependent on the level of significance we choose. Nonetheless, we decided to present rather the robust standard errors to control for the possible occurrence of such behaviour. We want to estimate the following model.

$$\begin{aligned}
log(Market\,Value_i) = &\beta_0 + \beta_1 Age_i + \beta_2 Age_i^2 + \beta_3 Height_i + \beta_4 Saves(\%)_i + \\
&\beta_5 Goals/90_i + \beta_6 CL\ Matches_i + \beta_7 EL\ Matches_i + \\
&\beta_8 Matches_i + \beta_9 Team\ Rank_i + \beta_{10} Passes\,Completed_i + \\
&\beta_{11} French\ League_i + \beta_{12} German\ League_i + \\
&\beta_{13} Italian\ League_i + \beta_{14} Spanish\ League_i + u_i
\end{aligned}$$

Again, we are more interested in the coefficients 1 to 14 to find out the partial effects than in the constant β_0. Based on the discussion in subsection 6.1, we expect β_1 and β_2 to behave in a similar manner as in the case of the field players. However, the peak is anticipated to be higher for goalkeepers. β_3 is supposed to be positive with a lower p-value than for other positions. Perhaps the most significant coefficient for goalkeepers is expected to be the Goals/90 ratio with a negative slope. The last coefficients are predicted to have the same signs as for defenders, midfielders and forwards combined.

6.3 Results

Table 10: OLS - Goalkeepers

	Dependent variable:
	log(Market Value)
Age	0.420^{***}
	(0.155)
Age^2	-0.009^{***}
	(0.003)
Height	1.892
	(1.467)
Saves(%)	0.002
	(0.005)
Goals/90	-0.183^{*}
	(0.096)
CL Matches	0.054^{**}
	(0.023)
EL Matches	0.018
	(0.026)
Matches	0.065^{***}
	(0.004)
Team Rank	-0.078^{***}
	(0.014)
Passes Completed	-0.002
	(0.004)
German League	-0.946^{***}
	(0.183)
Spanish League	-0.465^{**}
	(0.179)
French League	-1.206^{***}
	(0.183)
Italian League	-0.764^{***}
	(0.187)
Constant	0.544
	(3.740)
Observations	145
R^2	0.828
Adjusted R^2	0.809
Residual Std. Error	0.668 (df = 130)
F Statistic	44.685^{***} (df = 14; 130)

Note: *p<0.1; **p<0.05; ***p<0.01
Note: Robust standard errors in parentheses
Source: Author's computations

Discussion

First of all, the R-squared together with the adjusted R-squared are higher than 0.8, making this model similar or slightly better than the one for field players. Although the number of observations is significantly lower than in previous cases, 145 is still enough to obtain valid results and make inferences.

The first two variables concern age. Both are significant and we can find the peak the same way as before. Surprisingly, we have arrived at only 22.4 years, which does not differ from forwards. When we take into account our initial expectations, supported by the summary statistics, it was expected to be way more - somewhere around 24-25 years.

Height appeared insignificant by a small margin. This does not foster the theory from the data description, where it was believed that additional height could raise the market value. It could be explained by the fact that goalkeepers are supposed to be tall, which makes their height mean above standard, but at the same time, they do not benefit from being a bit taller than average.

A succeeding variable is the percentage of successful saves. It was expected to influence the regressand positively, but that was not the case, as it did not turn out to be significant.

On the contrary, the Goals/90 ratio was found significant and negative. If the statistics were lowered by 1 point, the dependent variable would plummet by almost 20 per cent. A logical conclusion would imply that their market values are more related to the overall conceded goals, caused by more players in the team, than to their single performance in terms of their saves proportion. This suggests that the market values of these footballers are more determined by the team results than their own performance during the games.

Moving along to the number of matches, we failed to find significance for the Europa League, while the Champions League and the domestic league seem to be good predictors. They both possess a positive sign and increase the market value by 5.4% and 6.5%. Two interesting conclusions can be drawn from this. The first one is that the Champions League contributes with a milder effect than the home one, which is the opposite to the field players. Therefore, it seems that domestic leagues are preferable in creating higher market values for these top 5 competitions. Second, a single game in the home league heightens the regressand way more for goalkeepers than for the other three positions

(6.5% to 3.8%). An ensuing suggestion is that for goalkeepers, it is more intrinsic to play and to acquire the experience, whilst the other positions, especially midfielders and forwards, are expected to be visible and to have solid game statistics.

An upcoming variable is the team rank in the final standings. Similarly to the field players, it is significant and negative. A decrease by one position would imply a fall of the market value by 7.8%, making both of them almost identical (-8.2%). Although the sign was anticipated beforehand, its gradient is again found surprisingly steep.

Next in line is the passes accuracy. Following the logic in subsection 6.1, we believed that there is a trend nowadays that goalkeepers are required to partially simulate another defender in the game. Unfortunately, this was not proven in the regression, as the coefficient is insignificant.

Finally, we will comment on the differences among the five leagues. Again, England is set as the base. The model yielded very similar results as the ones for the field players. The Premier League is all over at the top, followed by the Spanish La Liga. The Italian Serie A ranked third place, then there is the German Bundesliga, and at the bottom, we find the French Ligue 1. Even though it seems that there was no dramatic change, it is still noteworthy how ones' market values strikingly vary across the countries after it is controlled for other drivers.

The last piece of information encompasses the constant term. In the very same way, it is not so informative for us since we seek the marginal effects of the explanatory variables, especially in the case of the log-level model.

Robustness Check

The robustness check procedure is conducted analogously to the field players. First of all, we plotted the density function of the residuals from the model. The figure can be observed in the appendix (Figure A.2). It is apparent that there is a tendency towards normality, but at the same time, the distribution is not perfect at first glance. Compared to the field players, it looks a bit less gratifying.

Hence, we also reduced the dataset by dismissing 11 outliers altogether and estimated the model again. We were left with 134 observations, which is still an ample number for validity. The results can be seen in the appendix (Table A.2). To compare, the performance measures, such as both the R-squared values, declined by approximately 7%. This indicates that the initial model produced better results.

In particular words, the coefficients as well as the errors are almost indistinguishable. The only 2 news worth pointing out is that the peak age grew from 22.4 to 22.8 and that the Champions League narrowly lost its significance. Otherwise, the model yielded similar outcomes.

To conclude, the residuals analysis did not reject the model's good specification and dropping the outliers did not substantially alter the results. Therefore, it seems that the concealed effects of the superstars did not bias the results in Table 10.

7 Conclusion

This thesis aims to investigate the determinants of football players' market value using the ordinary least squares method. On a sample from the top 5 European leagues in season 2018/19, it delves into the comparison of these drivers across positions. Building on the existing knowledge, it uses both the personal characteristics and performance statistics from domestic and international competitions. Moreover, it uniquely deep dives into the analysis of goalkeepers, who have been vastly overlooked in the literature.

Regarding the field players, we find the significance of age, goals, assists, passes accuracy, European and domestic matches, team rank and the effects of each league for all of them. Further, height seems to be relevant only to the defenders and midfielders, raising their value. For the latter group, we also reveal a positive influence of interceptions and belonging to the top 10 players. After analysing the 5 countries, the biggest difference in terms of the market values was found between the French and English leagues. In particular words, an appearance in the Champions League is way more appreciated for footballers in the former country. For goalkeepers, we discover significant variables to consist of age, goals to 90 minutes ratio, European and home matches, team rank and the single impacts of countries they play in. Surprisingly, height, percentage of successful saves and percentage of completed passes fail to be located in the set of the significant drivers. All the positions' models were subject to the robustness checks carried out by plotting the density functions and rerunning the estimations without outliers. Neither of those, however, rejected their good specification, indicating they are robust.

Presumably, the biggest contribution of this work lies in two findings. Firstly, defenders receive higher credit for mere international and domestic caps than midfielders and forwards, who could thereby be more expected to deliver added value on the pitch. Secondly, to the best of our knowledge, this thesis is the pioneer in analysing goalkeepers' market values. After compiling the data based on their evaluation criteria, we estimated the model and found two interesting outcomes. The peak age appears to be 22.4 years, which is similar to the field players. Nevertheless, we anticipated that goalkeepers would reach their turning point later; therefore, this information is surprising. Further, goalkeepers' market values seem to be driven to a greater extent by the overall team performance than by the statistics directly oriented on them. This reasoning is based on the insignificance of the successful saves and passes percentages and significance of the goals to 90 minutes ratio, team rank and the rest of the variables. All of these contributions could be used by managers, teams and fans keen on football. It may also inspire other sports branches to conduct similar research. Thus, academicians might benefit from this work as well.

Possible extensions of this thesis lie in examining the effects of the COVID-19 pandemic. Quansah et al. (2020) find an approximate 20% fall in the market values of the English Premier League football players. Therefore, it would be interesting to investigate whether or not the drivers somehow changed for each position and league. Another great study might be carried out on extending the goalkeepers' analysis since their world is still quite unexplored. This could be done by including more variables, such as media popularity, youth international caps and more detailed game statistics.

References

Ante, L. (2019). Determinants of Transfers Fees: Evidence from the Five Major European Football Leagues. Research Gate Publication, 1-25.

Baan, S. (2016). The wisdom of crowds and football players. Master Thesis, Erasmus University Rotterdam.

Blaha, J. (2017). An Empirical Investigation of Wage Discrimination in Professional Football. Bachelor thesis, Charles University

Bryson, A., Frick, B., Simmons, R. (2013). The returns to scarce talent: Footedness and player remuneration in European soccer. Journal of Sports Economics, 14(6), 606-628.

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, pp. 125–50.

Carmichael, F., Thomas, D. (1993). Bargaining in the transfer market: theory and evidence. Applied Economics, 25(12), 1467-1476.

Dobson, S., Gerrard, B. (1999). The determination of player transfer fees in English professional soccer. Journal of Sport Management, 13(4), 259-279.

Dobson, S., Gerrard, B., Howe, S. (2000). The determination of transfer fees in English nonleague football. Applied Economics, 32(9), 1145-1152.

Dobson, S., Goddard, J. A., Dobson, S. (2001). The economics of football (Vol. 10). Cambridge: Cambridge University Press.

Erkmen, N., Taşkin, H., Sanioğlu, A., Kaplan, T., Baştürk, D. (2010). Relationships between balance and functional performance in football players. Journal of Human Kinetics, 26, 21-29.

Felipe, J. L., Fernandez-Luna, A., Burillo, P., de la Riva, L. E., Sanchez-Sanchez, J., Garcia-Unanue, J. (2020). Money talks: team variables and player positions that most influence the market value of professional male footballers in Europe. Sustainability, 12(9), 3709.

Frenger, M., Follert, F., Richau, L., Emrich, E. (2019). Follow me...on the relationship between social media activities and market values in the German Bundesliga. Working Papers of the European Institute for Socioeconomics No. 32.

Frick, B. (2007). The football players' labor market: empirical evidence from the major European leagues. Scottish Journal of Political Economy, 54: 422–46.

Fry, T. R., Galanos, G., Posso, A. (2014). Let's get Messi? Top-scorer productivity in the European Champions League. Scottish Journal of Political Economy, 61(3), 261-279.

Garcia-del-Barrio, P., Pujol, F. (2007). Hidden monopsony rents in winner-take-all markets—sport and economic contribution of Spanish soccer players. Managerial and Decision Economics, 28(1), 57-70.

He, M., Cachucho, R., Knobbe, A. (2015). Football player's performance and market value. In Proceedings of the 2nd workshop of sports analytics, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)., 87–95

Herm, S., Callsen-Bracker, H. M., Kreis, H. (2014). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. Sport Management Review, 17(4), 484-492.

Korzynski, P., Paniagua, J. (2016). Score a tweet and post a goal: Social media recipes for sports stars. Business Horizons, 59(2), 185-192.

Majewski, S. (2016). Identification of factors determining market value of the most valuable football players. Journal of Management and Business Administration. Central Europe, 24(3), 91-104.

Medcalfe, S. (2008). English league transfer prices: is there a racial dimension? A re-examination with new data. Applied Economics Letters, 15(11), 865-867.

Müller, O., Simons, A., Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. European Journal of Operational Research, 263(2), 611-624.

Peeters, T. (2018). Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. International Journal of Forecasting, 34(1), 17-29.

Quansah, T., Frick, B., Lang, M., Maguire, K. (2020). The Impact of the Coronavirus Outbreak (COVID-19) on Player Salaries, Transfer Fees, and Net Transfer Expenses in the English Premier League. Transfer Fees, and Net Transfer Expenses in the English Premier League (September 9, 2020). Working Paper.

Reilly, B., Witt, R. (1995). English league transfer prices: is there a racial dimension?. Applied Economics Letters, 2(7), 220-222.

Richau, L., Follert, F., Frenger, M., Emrich, E. (2019). Performance indicators in football: The importance of actual performance for the market value of football players. Sciamus–Sport und Management, 10, 41-67.

Ruijg, J., van Ophem, H. (2015). Determinants of football transfers. Applied Economics Letters, 22(1), 12-19.

Sinčák, J. (2020). Determinants of Football Players' Market Value in the Czech Football League. Bachelor thesis, Charles University

Speight, A., Thomas, D. (1997). Football league transfers: a comparison of negotiated fees with arbitration settlements. Applied Economics Letters, 4(1), 41-44.

Surowiecki, J. (2005). The wisdom of crowds (First Edition). New York, Anchor Books

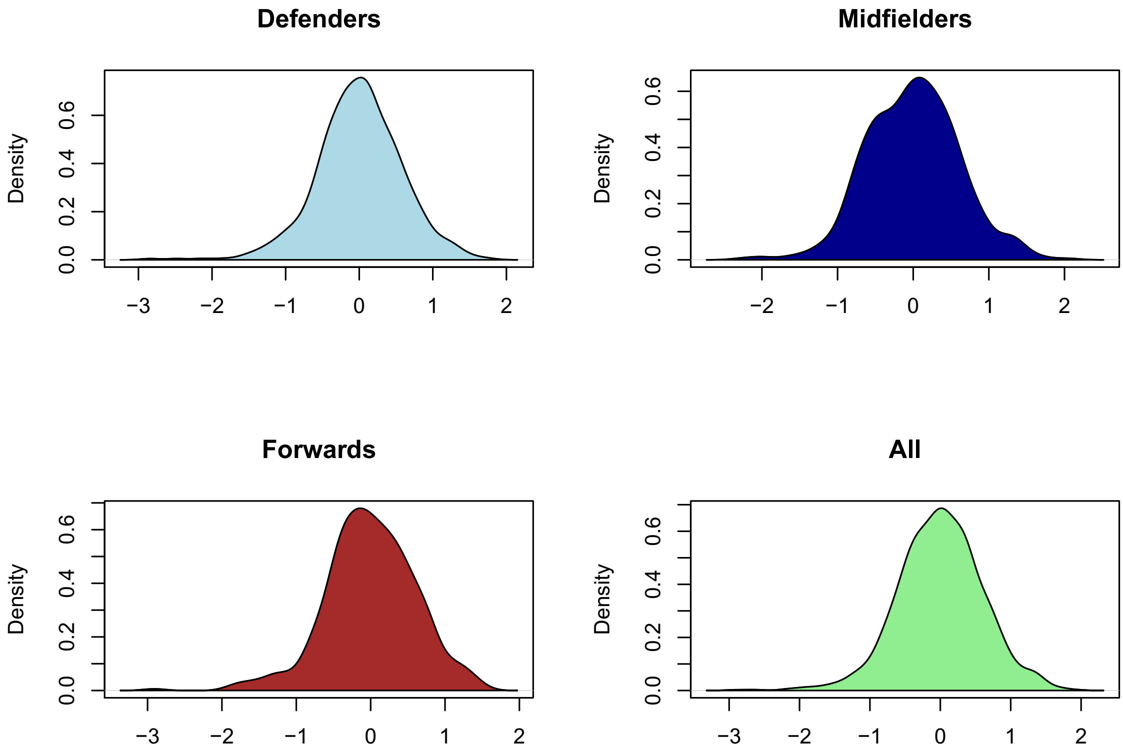
Torgler, B., Schmidt, S. L. (2007). What shapes player performance in soccer? Empirical findings from a panel analysis. Applied Economics, 39(18), 2355-2369.

Wooldridge, J. (2016). Introductory econometrics: A modern approach (Sixth Edition). USA, Cengage Learning.

Appendix

Appendix A: Robustness Check - Field Players

Figure A.1: Densities of the Residuals - Field Players



Source: Author's computations

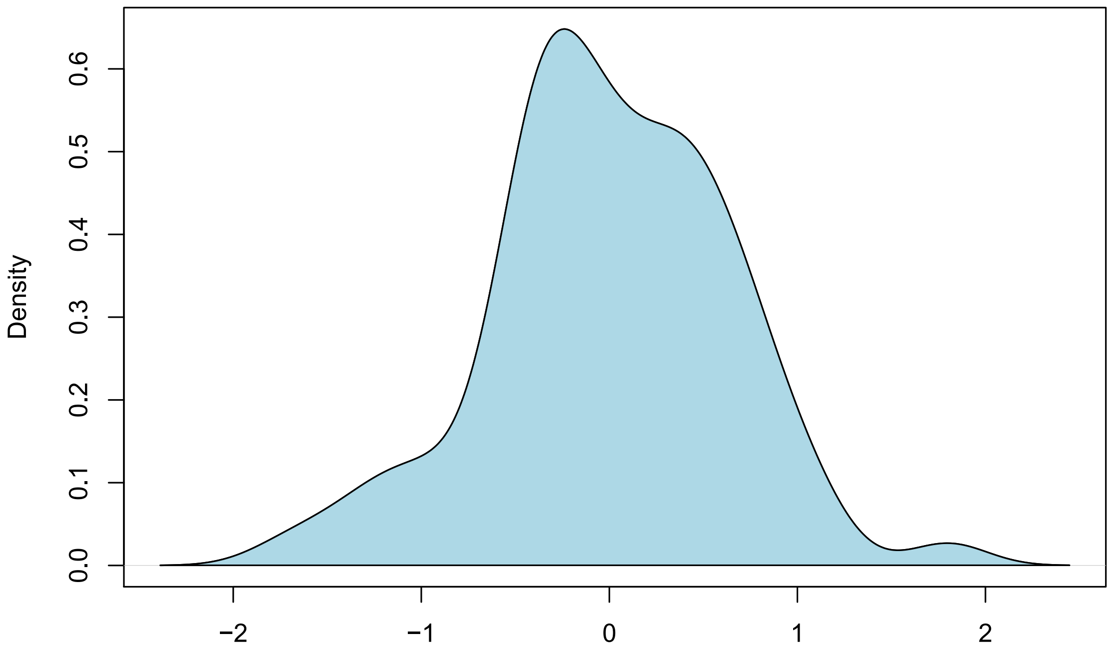
Table A.1: OLS Field Players - Without Outliers

	Dependent variable:			
	log(Market Value)			
	Defenders	Midfielders	Forwards	All
Age	0.575***	0.496***	0.546***	0.546***
	(0.087)	(0.083)	(0.081)	(0.049)
Age^2	−0.013***	−0.011***	−0.012***	−0.012***
	(0.002)	(0.002)	(0.001)	(0.001)
Height	1.172**	1.149**	−0.707	0.459
	(0.465)	(0.486)	(0.607)	(0.285)
Goals	0.067***	0.098***	0.051***	0.062***
	(0.023)	(0.018)	(0.012)	(0.007)
Goals/Shots	0.201	0.030	0.470	0.245
	(0.193)	(0.362)	(0.519)	(0.172)
Assists	0.047***	0.064***	0.054***	0.057***
	(0.018)	(0.013)	(0.017)	(0.009)
Passes Completed	0.013***	0.021***	0.013**	0.017***
	(0.004)	(0.004)	(0.005)	(0.002)
Fouls	0.003	−0.003	0.004	0.002
	(0.003)	(0.002)	(0.003)	(0.001)
Top10	0.446***	0.193	0.266	0.352***
	(0.168)	(0.153)	(0.188)	(0.096)
Interceptions	0.003	0.010***	0.006	0.005***
	(0.003)	(0.003)	(0.005)	(0.002)
CL Matches	0.072***	0.045***	0.040*	0.055***
	(0.016)	(0.017)	(0.021)	(0.010)
EL Matches	0.021**	0.008	0.009	0.013*
	(0.009)	(0.014)	(0.013)	(0.007)
Matches	0.037***	0.035***	0.033***	0.035***
	(0.005)	(0.005)	(0.006)	(0.003)
Team Rank	−0.079***	−0.074***	−0.074***	−0.075***
	(0.007)	(0.008)	(0.008)	(0.004)
TN Ratio	0.00000	0.00000*	0.00000	0.00000**
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Both Feet	−0.162	−0.087	0.101	0.018
	(0.180)	(0.223)	(0.108)	(0.107)
French League	−0.980***	−1.029***	−1.034***	−1.021***
	(0.077)	(0.086)	(0.095)	(0.048)
German League	−0.771***	−0.805***	−0.745***	−0.779***
	(0.079)	(0.087)	(0.099)	(0.050)
Italian League	−0.825***	−0.735***	−0.794***	−0.784***
	(0.072)	(0.092)	(0.118)	(0.052)
Spanish League	−0.663***	−0.601***	−0.738***	−0.671***
	(0.077)	(0.089)	(0.112)	(0.051)
Constant	−0.406	0.046	3.004*	0.878
	(1.367)	(1.418)	(1.709)	(0.818)
Observations	564	472	362	1,398
R^2	0.736	0.752	0.727	0.733
Adjusted R^2	0.727	0.741	0.711	0.729
Residual Std. Error	0.582 (df = 543)	0.614 (df = 451)	0.617 (df = 341)	0.601 (df = 1377)
F Statistic	75.828*** (df = 20; 543)	68.474*** (df = 20; 451)	45.463*** (df = 20; 341)	188.795*** (df = 20; 1377)

Note: *p<0.1; **p<0.05; ***p<0.01
Note: Robust standard errors in parentheses
Source: Author's computations

Appendix B: Robustness Check - Goalkeepers

Figure A.2: Density of the Residuals - Goalkeepers



Source: Author's computations

Table A.2: OLS Goalkeepers - Without Outliers

	Dependent variable:
	log(Market Value)
Age	0.444***
	(0.162)
Age^2	−0.010***
	(0.003)
Height	1.369
	(1.522)
Saves	0.001
	(0.005)
Goals/90	−0.202**
	(0.095)
CL Matches	0.049
	(0.041)
EL Matches	0.027
	(0.027)
Matches	0.064***
	(0.004)
Team Rank	−0.072***
	(0.015)
Passes Completed	−0.003
	(0.005)
German League	−0.970***
	(0.216)
Spanish League	−0.602***
	(0.209)
French League	−1.279***
	(0.211)
Italian League	−0.846***
	(0.206)
Constant	1.309
	(3.996)
Observations	134
R^2	0.767
Adjusted R^2	0.740
Residual Std. Error	0.677 (df = 119)
F Statistic	27.974*** (df = 14; 119)

Note: *p<0.1; **p<0.05; ***p<0.01
Note: Robust standard errors in parentheses
Source: Author's computations