CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Determinants of football players' market value during the season: Discrepancies between TOP 5 and the rest of the leagues

Bachelor's thesis

Author: Vojtěch Chalupa

Study program: Economics and Finance

Supervisor: Mgr. Josef Kurka

Year of defense: 2024

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

During the preparation of this thesis, the author used ChatGPT in order to acquire assistance, particularly for coding in R and LaTeX, and for minor modifications in academic language. After using this tool/service, the author reviewed and edited the content as necessary and takes full responsibility for the content of the publication.

Prague, April 30, 2024

Vojtech Chalupa

Abstract

Football is a global phenomenon, primarily driven by clubs acquiring players who perform on their behalf on the pitch. The valuation of footballers during transfer negotiations is influenced by numerous factors. This thesis expands upon existing research by examining the discrepancies in the mechanisms of player market value formation between the top 5 European leagues and other competitions, using values from Transfermarkt.com together with the data from the 2022/23 season. Employing the Ordinary Least Squares method, we discover that the substantial enhancement in the total price tags of players in the elite leagues is accompanied by variations in the factors affecting these values. Notably, the analysis reveals significant disparities in the influence of ball possession-related statistics such as successful passes, dribbles, and the xG indicator, which tend to positively affect player values only in less prestigious leagues. Common significant factors with strong effects in both groups include pre-season player's worth, number of minutes played, goals scored, and statistics reflecting team success. Contrary to previous studies, this research identified that age consistently impacts all players negatively. Further analysis also explores differences between individual positions on the pitch and within the top 5 leagues themselves.

Keywords Football, Transfer market, Market value, Determinants, Differences

Title Determinants of football players' market value during the season: Discrepancies between TOP 5 and the rest of the leagues

Author's e-mail 79321775@fsv.cuni.cz
Supervisor's e-mail josef.kurka@fsv.cuni.cz

Abstrakt

Fotbal je celosvětový fenomén, jehož hlavním motorem jsou kluby kupující hráče, kteří je reprezentují na hřišti. Oceňování fotbalistů během přestupových jednání je ovlivněno mnoha faktory. Tato práce rozšiřuje stávající výzkum o vhled do rozdílů v mechanismech formování tržní hodnoty hráčů mezi pěti nejlepšími evropskými ligami a ostatními soutěžemi, přičemž využívá hodnoty z Transfermarkt.com spolu s daty ze sezóny 2022/23. Použitím metody nejmenších čtverců zjistila, že výrazně vyšší celkové cenovky hráčů v elitních ligách jsou doprovázeny odlišnostmi ve faktorech ovlivňujících tyto hodnoty. Analýza zvláště odhaluje významné rozdíly ve vlivu statistik souvisejících s držením míče, jako jsou úspěšné přihrávky, driblingy a indikátor xG, které mají tendenci pozitivně ovlivňovat hodnoty hráčů pouze v méně prestižních ligách. Společně významnými faktory se silným vlivem v obou skupinách jsou předsezónní hodnota hráče, množství odehraných minut, vstřelené góly a statistiky odrážející úspěch týmu. V rozporu s předchozími studiemi tento výzkum zjistil, že věk má konzistentně negativní dopad na všechny hráče. Další analýza také zkoumá rozdíly mezi jednotlivými pozicemi na hřišti a v rámci samotných pěti nejlepších lig.

Klíčová slova Fotbal, Přestupový trh, Tržní hodnota, Determinanty, Rozdíly

Název práce Determinanty tržní ceny fotbalistů během sezony: rozdíly mezi TOP 5 a zbylými ligami

E-mail autora `79321775@fsv.cuni.cz`
E-mail vedoucího práce `josef.kurka@fsv.cuni.cz`

Acknowledgments

I am profoundly grateful especially to Mgr. Josef Kurka for supervising this thesis and for his consistent assistance throughout the writing process, always offering advice and tips for its improvement. Additionally, I would like to thank my family and close friends for their support and encouragement during the whole period of my studies.

Typeset in LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Chalupa, Vojtech: *Determinants of football players' market value during the season: Discrepancies between TOP 5 and the rest of the leagues.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2024, pages 54. Advisor: Mgr. Josef Kurka

Contents

List of Tables

List of Figures

Acronyms

NBA National Basketball Association

NHL National Hockey League

MLB Major League Baseball

FIFA Fédération Internationale de Football Association

UEFA Union of European Football Associations

Chapter 1

Introduction

Football, recognized as the world's number one sport, transcends geographical, cultural, and linguistic differences, attracting an extensive network of players and billions of fans. The popularity of the game is evidenced by the viewership figures showing that Manchester City's 2023 Champions League victory was watched by around 450 million fans worldwide and the 2022 FIFA World Cup final in Qatar attracted even up to 1.5 billion viewers (Reidy, 2023).

This huge global audience brings more intensive attention to player transfers between clubs, particularly their financial aspects. The valuation of footballers during transfer negotiations presents an important issue for clubs, as well as for the fans. This economic question came to the surface even more during the instability caused by the COVID-19 crisis. Contrary to expectations that the economic downturn would curb spending, the transfer market, notably within the English Premier League, has witnessed spending that defies the global economic challenges. Despite the worldwide recession, the 2020 summer transfer window saw Premier League clubs spending slightly under €1.5 billion on new acquisitions. This expenditure mirrors the investment levels of the preceding two years. Although a marginal reduction in spending the subsequent summer, the years 2022 and 2023 saw a significant surge, with expenditures nearly doubling the pre-pandemic figures and approaching the €3 billion mark, as reported by Transfermarkt.com. However, these increasing sums spent on reinforcements by clubs from the leading European leagues seemingly widen the gap between these competitions and their less prominent counterparts.

Several papers have already been written on the topic of player valuation, trying to uncover the underlying pricing mechanisms. Their findings indicate that a player's age and basic game statistics, such as goals and assists, exert a significant influence on their market value (Carmichael *et al.*, 1999; Majewski, 2016). Other offensive and defensive statistics, including successful passes or interceptions, are also determinative (Franck & Nüesch, 2012). Furthermore, the popularity of the player and the influence of the team they represent appear to be important factors (Herm *et al.*, 2014).

Nevertheless, many of the existing works focus on only one of the elite European leagues, such as He *et al.* (2015) with the Spanish LaLiga, or compare the commonly perceived top 5 competitions, as discussed by Ante (2019). Despite these insights, a lack of attention has been paid to the divergence in valuation between the most prestigious leagues and the rest, which should be even more pronounced. The similar stands for the distinction between player positions, as some authors (Majewski, 2016) prefer to focus their analysis on only one position, in order to avoid complications.

Therefore, this thesis is dedicated to an in-depth investigation of mechanisms that set apart the top 5 European leagues from the others, which is an underexplored area. The objective is to identify the core factors that contribute to the significant gaps in market valuation. Further, the study will examine the variance across different playing positions and finally within the elite leagues themselves. This exploration is crucial for a better understanding of the aspects causing discrepancies in player valuation. Such insights might help explain the substantially higher valuation of players in the most prestigious leagues. At the same time, the results could uncover the significant disparities arising due to the unique responsibilities and roles assigned to each playing position on the pitch. This leads to, for example, the fact there is not a single defender present among the 20 most expensive players overall. The whole thesis could serve as a strategic guide for clubs and stakeholders within the football market, enhancing their comprehension of player valuation dynamics and the underlying processes by which these sums are established.

In order to carry out this research, the Ordinary Least Squares method is employed along with LASSO regression for the purpose of regularization. The empirical results suggest there occur significant discrepancies between leagues and among individual positions. The largest of these are primarily associated with the effect of game statistics related to ball possession, but also with other personal or team performance indicators.

The remainder of this thesis is structured as follows. Initially, a Literature review summarizes the findings from previous research, Chapter 3 presents the data and variables employed in the empirical analysis. Subsequent to this, Chapter 4 elaborates on the chosen model and its construction, Chapter 5 presents a synthesis of the results derived from the analysis, including the implications of these findings. Finally, the Conclusion part summarizes the outcomes of this work.

Chapter 2

Literature review

In this section, we summarize the relevant literature from several perspectives. Firstly, we review the valuation of athletes generally, then we also focus specifically on the valuation of football players. As the main goal is to investigate the influence of league affiliation and also the position played, papers giving more attention to these specific effects represent another point of interest. In the end, we shortly elaborate on the way the athletes' market values are predicted and on the precision of such estimates.

2.1 Player valuation in professional sport

Nowadays sports stretch beyond the results of the games played as club owners, managers, fans, and many others are interested in pricing athletes indicating their worth on their respective markets. The system of assigning certain values to players varies substantially from sport to sport. Arguably, the most transparent method is the use of market values representing the amount for which a player should be exchanged in the event of a transfer to a new club. However, such type of valuation is specific to football.

In individual sports like tennis, golf, or darts, the measurable amount of money comes with the prize pool won during the season. The situation also differs in popular team sports such as ice hockey or basketball. For example, in the Czech environment, ice hockey clubs must follow certain table prices based on the player's age, experience, and playing statistics (Český svaz ledního hokeje z.s., 2022).

Even major sports leagues in the USA do not employ those sums we can see in the world of football. Various swaps and trades are negotiated between the clubs involved, utilizing a system in which players are valued by salaries guaranteed in their contracts. However, the worth of these deals and the factors influencing them serve as researchers' points of analysis the same way. In the NHL, the extent of a player's contract is based on his statistics such as goals and assists, as well as his experience and popularity factor (Peck, 2012). Pacák (2021) presents a similar verdict for NBA contracts, while Berri (2006) concludes that points scored are by far the strongest wage determinant in the NBA. Other game statistics came out as much less significant.

The "traditional" factors influencing players' value can be characterized as those arising from the concepts of human capital theory, and they might be divided into 4 basic areas of interest (Deutscher, 2018). Firstly, athletes' value is positively affected by experience. However, experience is naturally positively correlated with athletes' age, and athletic abilities inevitably reduce with increasing age. Hence a certain form of inverse U-shaped relationship should be expected between experience and athlete's value. The second factor is, unsurprisingly, past sports performance. A good past performance is perceived to indicate a high level of performance in the future and therefore increases the valuation of an athlete, e.g. Frick (2011) suggests that past (and recent) performance is among the two leading indicators of athletes' value.

Naturally, talent must also be included in the list of factors affecting the market values of athletes. Talent is not directly observable though, thus researchers must use proxy or instrumental variables while controlling for talent, e.g., the draft position, youth competitions, and youth national team appearances. The last factor claimed to have a substantial effect is players' popularity, which can be proxied by participation in All-Star games, frequency of magazine articles, clicks on the internet, or number of followers on social networks. Although this factor does not relate directly to sports performance, it is documented to be among the most influential predictors of athletes' valuation (Frick, 2011; Franck & Nüesch, 2012).

Nevertheless, with the development of new technologies, the trend of valuation in professional sports shifted more and more to the analysis of collected data. Hakes & Sauer (2006) examined the "Moneyball"[1] hypothesis from the economic point of view on MLB data and confirmed that certain player abilities are systematically undervalued as originally stated in Lewis (2003). The use of data analysis has been starting to play a major role when assessing player val-

[1]By analyzing data, baseball teams can find undervalued players that will help them win more games.

ues and capabilities in football after lagging behind other sports for a long time (Müller *et al.*, 2017). It is mainly caused by the development of big sports-data companies like Opta (www.optasports.com), allowing its users to deconstruct the whole game into pieces of statistical data.

2.2 Player valuation in football

In the case of football, it is straightforward to determine the value of a player as transparent transfer fees are often negotiated between the two parties involved when a player is sold. The amount of money paid by the buying club usually becomes publicly known, sometimes even before the deal gets completely closed between the two sides. Since players normally do not change clubs every transfer window, it would be inappropriate to keep track of their valuation only in this way. To obtain an idea about the possible transfer fee magnitude, football fans, as well as journalists, frequently refer to footballers' market values from Transfermarkt.com. Researchers often relate to Transfermarkt values as the dependent variable while building models attempting to determine player values (Herm *et al.*, 2014; Majewski, 2016). Other possible measures of player value include actual fees paid (Carmichael *et al.*, 1999; Ante, 2019) or salaries earned by players (Brandes & Franck, 2012; Bryson *et al.*, 2013)[2]. Most of the already existing works using Transfermarkt values do not pay enough attention to the fact that performance statistics from one season lead to a change in the players' worth, rather than its establishment (Herm *et al.*, 2014; Majewski, 2016). The authors only focus on the market value at one particular point in time and take this amount as an explained variable in their model without considering any importance of the previous player's worth. However, some papers used the market value from the pre-season season period as an explanatory variable together with the value from the end of the season to account for the undeniable impact of the former valuation. We want to follow a similar approach and take both these numbers into account (Müller *et al.*, 2017).

Among the results of previous research, many different factors were already tested in terms of the level of influence they have on the market value (or transfer fee). We divide them into categories concerning the kind of information about the player they yield. The first type of important player value determinants consists of basic observable characteristics. Age is substantially

[2]The determinants of Transfermarkt values, actual transfer fees paid, or salaries paid should only be slightly different (Müller *et al.*, 2017)

dominant among such factors, as authors unanimously confirm its key role (Franck & Nüesch, 2012; Majewski, 2016; Müller *et al.*, 2017). Age serves as a proxy of either the potential or experience of the player, however, it displays the expected inverse U-shape as the athletic abilities diminish with increasing age. Its positive effect has a decreasing rate until the peak around the age of 24, then the effect becomes negative (Herm *et al.*, 2014; Carmichael *et al.*, 1999), thus, it should be involved in a quadratic form. Apart from age, another important determinant of player value is nationality. Teams prefer players coming from countries with a history of producing high-quality football players. Therefore, a player's national team's FIFA ranking displays a negative effect on the player's value (Majewski, 2016), and it has also been found that footballers from South America are valued higher than players from other continents. Additionally, physical attributes like height or footedness (ability to use both feet) are also significant drivers of value (Ante, 2019). This can be attributed mainly to the increased goal threat taller players bring and the versatility given to the manager by having "two-footed" players who have no problem with playing on any side of the pitch.

The second category of metrics delves into player performance and game statistics. Carmichael *et al.* (1999) assert that the number of matches played throughout a player's career and contribution to the team particularly in terms of goals scored, significantly impact their value. This study posits that being part of a national team squad also exerts an influence on the transfer amount. Similar conclusions are drawn by Majewski (2016) in his examination of the 150 most valuable attackers, where he additionally considers the impact of assists provided by the player. These conclusions regarding the importance of goals, assists, and other game-related metrics are echoed in various studies (Franck & Nüesch, 2012; Herm *et al.*, 2014; Ante, 2019). Apart from goals and assists, there are also other less eminent in-game factors such as dribbles made, successful passes, tackles, or interceptions (Franck & Nüesch, 2012) contributing positively to a player's value. On the other hand, the yellow cards received negatively affect the player's value, albeit not particularly strongly (Müller *et al.*, 2017).

A certain level of importance is also attributed to external factors, such as average grades given by football experts and the number of Google search hits (Herm *et al.*, 2014). The effect of popularity on the internet and social media is a common finding of many models considering popularity through metrics like the frequency of Google searches, the number of social media followers, YouTube videos, or performance non-related press citations (Garcia-del Barrio & Pujol, 2007; Franck & Nüesch, 2012; Müller *et al.*, 2017; Ante, 2019).

Since talent is a hardly observable kind of criterion, researchers usually do not include any specific proxies to measure it in their models. It is perceived to be reflected in the footballer's performance on the pitch. More often variables dealing with players' contracts are incorporated. These stem from elements like clauses or the length of the contract (Franceschi *et al.*, 2023). Also, the addition of factors concerning the team, e.g. the overall market value, emerges as significant (Herm *et al.*, 2014; Majewski, 2016). The rationale behind this observation suggests that more expensive squads are typically associated with top teams, creating a favorable perception of a player belonging to such a club.

2.3 Differences between positions and leagues

Given the diverse nature of football positions, all players cannot be evaluated based on the same statistics. Various positions on the pitch require distinct traits and abilities, leading researchers to focus on multiple subsamples to distinguish between roles during the game. A similar principle applies to switching between leagues as players producing comparable statistics in competitions of different qualities cannot be evaluated identically. Consequently, certain disparity in the valuation of players with matching game statistics may be expected when one plays in a top-ranked league (e.g. English Premier League), and the other competes in a one outside the top 5, such as the Scottish Premiership.

Constructing an effective model for the valuation of footballers overall poses challenges, because substantial differences exist between the four traditional positions - goalkeepers, defenders, midfielders, and strikers - when assigning market values (He *et al.*, 2015). This discrepancy arises, for example, as a result of strikers garnering far more attention than other team members, primarily due to their role as the main goal contributors. However, the goalkeeper stands out as the most distinct, attributed to the non-universality of these players. The significance of this position is underscored by completely different statistical indicators than those applicable to outfield players, prompting an examination of metrics such as the saves-to-shots ratio (Franck & Nüesch, 2012). Another notable divergence among on-pitch positions lies in movement requirements, with variations in the distance covered and intensities of specific runs depending on the player's role (Di Salvo *et al.*, 2006).

He *et al.* (2015), propose that statistics such as goals, assists, shots on goal, or successful dribbling are pivotal for strikers, while Carmichael *et al.* (1999) expand the importance of these variables also to midfielders. On the contrary, the quality of defenders needs to be assessed based on different metrics associated with defensive responsibilities. Consequently, there occurs a more pronounced effect of defensive statistics on the valuation of defenders, while the influence of offensive statistics on attackers is less apparent (Ante, 2019). Additionally, these positional disparities may extend to basic statistics and characteristics, such as age or the number of games played (Cvrček, 2021). Given these considerations, Majewski (2016) and He *et al.* (2015) opted to construct a model specifically for forwards.

Comparisons between leagues present another contentious point, due to the varying nature of domestic competitions across Europe. This divergence in drivers forming the imaginary price tag can be reported even between the most famous leagues, the so-called top 5 (English Premier League, Spanish LaLiga, German Bundesliga, Italian Serie A, and French Ligue 1). Not only there occurs an additional "bonus" in the value associated with participation in the Premier League compared to the other four leagues, but factors deemed significant for players in one of them may not necessarily hold the same importance for footballers in the rest (Ante, 2019). To mitigate this league bias, many studies focus on one competition exclusively. For instance, Herm *et al.* (2014) examine the Bundesliga, while He *et al.* (2015) concentrate on LaLiga.

Examination of discrepancies between the top 5 and out-of-top 5 leagues is an area that has not been given that much attention in the previous research. It is therefore useful to have a closer look into this issue. To analyze this, we will distinguish between models for separated subsamples of players performing in the top 5 leagues and the other leagues. Additionally, in line with the approaches of Ante (2019) or Cvrček (2021), also models for different positions and each of the top 5 leagues will be the focus of interest.

2.4 Transfermarkt market values and transfer fees

As previously noted, to explore the influence of various factors on the valuation of football players both market values and actual transfer fees are often used. Since market values serve the purpose of predicting the amount paid as accurately as possible, they act as proxies in this context. However, examining true relationships can be more effectively undertaken using these estimates, because transfer fees are frequently subject to the unique circumstances of both clubs and various unobservable factors, such as injuries, transfers out, and rivalry, which are challenging to incorporate into a model (Müller *et al.*, 2017). The values provided on Transfermarkt.com result from crowd-sourced evaluations of footballers. The underlying concept is rooted in the notion that the collective predictions of numerous football fans can match or even surpass the quality of predictions made by a few experts, the so-called "wisdom of crowds" (Surowiecki, 2005). Herm *et al.* (2014) illustrated the decision-making process regarding Transfermarkt market values using Brunswik's lens model, as depicted in Figure 2.1.

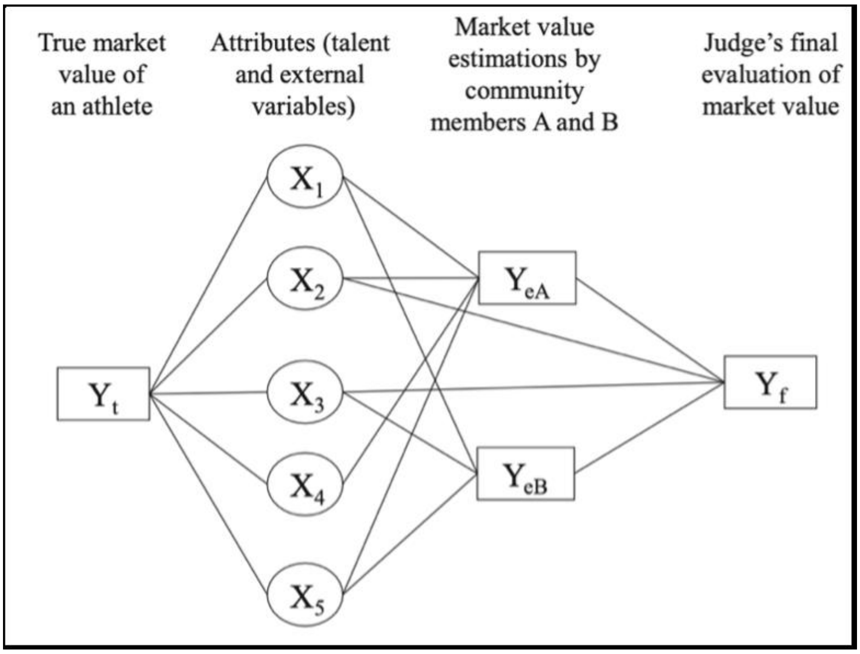

Figure 2.1: Brunswik's lens model as illustrated by Herm et al. (2014).

Because of the disparities occurring between the expected and exact fees paid in a case of transfer, researchers have already turned their attention also on the reliability of these predictions. There is a large degree of consensus that these crowdsourced estimates are relatively accurate and potentially superior to other methods of predicting transfer values. Moreover, they are precise indicators that can be used to estimate other variables like players' salaries (Prockl *et al.*, 2018), and when combined with the interpretation of expert judges, they explain most of the variance in transfer fees (Herm *et al.*, 2014). Furthermore, values from Transfermarkt might even serve as effective predictors of football match results, outperforming conventional methods such as FIFA ranking or ELO ratings. Peeters (2018) also fails to find any evidence of "wishful thinking bias" which would lead to a higher valuation of favorite players. A counterargument was presented by Coates & Parshakov (2022) who argue that values on Transfermarkt.com are biased estimates of actual transfer fees, primarily due to the crowdsourced valuation methodology that causes them to be frequently underestimated. Overall, the empirical results suggest that the Tranfermarkt values present trustworthy, although not perfect, estimates of the player's worth. The precision of these crowdsourced predictions increases for higher levels of the value hierarchy (Müller *et al.*, 2017).

Chapter 3

Data

Reviewing the current state of knowledge helps us better understand the core matters of this topic, thereby facilitating the extension of previous research. This section describes the process of collecting the data used to estimate the desired models. We focus on player statistics and market values from the 2022/23 season, in order to employ the most recent records available. Data procurement spanned across the leading 10 European football leagues based on the current UEFA coefficient rankings, and the same measure is taken into account when determining the top 5 leagues.[1]

3.1 Data collection

Compiling the data from various sources was necessary to construct the required dataset, given the absence of a unified data repository comprising all essential attributes. The basis of our dataset, which includes team sheets and basic player characteristics such as age and height, was taken from the Football Manager 23 video game database.[2]

Market valuations for each player were obtained from Transfermarkt.com, the leading online platform for this subject. For match statistics, FBref.com was used due to its comprehensive football statistics database. To address data gaps primarily concerning the Scottish Premiership and Turkish Super Lig,

[1] The top 5 leagues include the English Premier League, Spanish LaLiga, German Bundesliga, Italian Serie A, and French Ligue 1. The other 5 leagues in the sample consist of Dutch Eredivisie, Portuguese Liga Portugal, Belgian Jupiler Pro League, Turkish Super Lig and Scottish Premiership.

[2] Football Manager is renowned not only for its entertainment value but also as a reliable resource for football scouts and professionals, due to its comprehensive player database.

FotMob.com was employed. Additionally, Livesport.cz served as an auxiliary resource for verification purposes in instances of data ambiguity.

3.2 Data filtering

We apply several filters to reduce noise in the data and decrease the bias of the estimates. Firstly, in line with the previous research (Bryson *et al.*, 2013; Ante, 2019), all of the models are estimated excluding goalkeepers. It is very difficult to develop a metric that would price goalkeepers and the rest of the players simultaneously, as the tasks of goalkeepers during games and metrics indicating the level of their performance are very different from the rest of the squad.

Secondly, players who were on loan at another club were not incorporated in the dataset. The value of a loaned player may be dramatically influenced by their affiliation with another club which could bias the main focus of this thesis, i.e. revealing the discrepancies in value determination between the top 5 and the rest of the leagues. Furthermore, we exclude players who were transferred to a new club either during the summer or winter trading window. The primary concern is that market values are typically updated at the beginning and after the end of transfer periods. In the case of transfers, the player's value during the updates may be significantly influenced by the transfer fee which can result in a substantial increase in the player's market value without sufficient performance-based justification.[3]

Lastly, players were required to play for at least 90 minutes during the season to be included in the sample. This is important for two reasons. Firstly, there were often missing statistics for players who played for a very short amount of time. Secondly, since most of the in-game statistics took the per 90 minutes form in our model, it should be statistically more appropriate only to include players who participated for at least this amount of time.

Following the data filtering, the resulting sample comprises 1952 players. Table 3.1 displays their frequencies by both league affiliation and position. The Premier League is the most represented league, with a total of 269 observations, the least represented one is the Scottish Premiership. Defenders are the most frequent position in the sample with 799 observations, followed by midfielders with 653 observations, while attackers encompass 500 observations.

[3]Recent examples are, e.g., the transfers of Marc Cucurella to Chelsea or Jakub Kiwior to Arsenal, where the transfer fees greatly exceeded their previous market values.

Table 3.1: Frequencies by league and position

	ENG	SPA	GER	ITA	FRA	NED	POR	BEL	TUR	SCO	Total
Defenders	108	106	84	84	77	77	81	72	62	48	799
Midfielders	86	88	75	71	76	50	54	60	46	47	653
Attackers	75	68	58	51	52	36	47	40	41	32	500
Total	269	262	217	206	205	163	182	172	149	127	1952

3.3 Variables

In the subsequent section, we list the variables that will be incorporated into our model. Beginning with the explained variable, we introduce each of them, along with presenting the descriptive statistics.

3.3.1 Market values

As discussed in Chapter 2, closing market values at the end of the season obtained from Transfermarkt serve as our dependent variable, as they present a well-performing proxy of the actual player's worth. However, since the objective is to explain the valuation shift within the run of the season, it was essential to control for the pre-season market value to demonstrate its non-negligible impact (Müller *et al.*, 2017). Moreover, it also performs as a good proxy for a player's quality.

Table 3.2 presents the descriptive statistics for all numerical variables. The findings support our hypothesis concerning the interrelation between the two market values, as detected through their descriptive metrics. The post-season valuation only marginally surpasses the pre-season figures across most metrics, evidenced by average values exceeding 9,000,000 euros and a consistent median of 3,000,000 euros for both instances.

3.3.2 Numerical variables

Our model mainly consists of numerical variables related to both the individual players and the teams they play for. For enhanced comprehensibility, we categorize the numerical explanatory variables into three different groups.

Table 3.2: Descriptive statistics of variables

Variable	Measurement type	Mean	Median	Min	Max	St. Dev.
Dependent variable						
MV season end	€	9 646 657.27	3 000 000.00	50 000.00	180 000 000.00	16 376 914.38
Independent variables						
MV season start	€	9 069 582.48	3 000 000.00	25 000.00	160 000 000.00	14 892 840.53
Age	years	26.23	26.00	16.00	40.00	4.27
Height	cm	181.19	181.00	162.00	201.00	6.39
Minutes played	season total	1 685.37	1 738.00	90.00	3 409.00	853.82
Goals	season total	2.48	1.00	0.00	30.00	3.65
Assists	season total	1.79	1.00	0.00	21.00	2.29
Yellow cards	season total	3.78	3.00	0.00	18.00	2.87
Red cards	season total	0.20	0.00	0.00	3.00	0.46
Fouls	per 90 min	1.24	1.14	0.00	5.71	0.61
Successful passes	per 90 min	35.78	34.70	5.70	113.10	14.79
Successful dribbles	per 90 min	0.81	0.60	0.00	6.71	0.73
Shots	per 90 min	1.24	0.94	0.00	5.56	0.95
xG	per 90 min	0.14	0.08	0.00	1.52	0.15
Aerial duels won	per 90 min	1.39	1.08	0.00	10.30	1.13
Tackles won	per 90 min	1.02	0.97	0.00	5.38	0.53
Interceptions	per 90 min	0.95	0.92	0.00	3.33	0.57
National ranking	FIFA ranking	22.04	10.00	1.00	181.00	26.99
Team goals scored	regular league season	52.44	50.00	19.00	114.00	16.82
Team goals conceded	regular league season	50.52	49.00	20.00	83.00	13.19

Note: n=1952

Basic observable characteristics

The first group refers to variables encapsulating general player information. In line with the previous research (Carmichael *et al.*, 1999; Franck & Nüesch, 2012; Majewski, 2016), our model includes the age variable. This metric functions dually as an indicator of the potential for younger players and as a proxy for gained experience for those in the latter stages of their careers. In anticipation of an inverted U-shaped relationship, we also incorporate the quadratic term of age (Age.sq). This should cover both the initial accumulation of experience in the early seasons as well as the gradual deterioration, especially of athletic ability, due to aging. The age at the start of the season is considered for the purposes of the empirical estimation.

Another player-level attribute is the height (Hght), measured in centimeters. The impact of a player's stature is twofold. While smaller players may possess advantages in areas like dribbling, acceleration, etc., taller players have an undeniable superiority in duels, especially aerial ones. This endows taller players with a key advantage during set pieces, which frequently determine match outcomes in modern football. Therefore, we expect a positive relationship between height and market value. Finally, we account for the effect of player's nationality, included in the form of National Ranking (NR) as published by FIFA. It stands to reason that players hailing from countries with superior NRs are sought-after, hence displaying higher average market valuations.

Individual match statistics

Anticipating that shifts in market valuation are primarily influenced by players' performances, the majority of variables incorporated into our model relate directly to individual match statistics from domestic league campaigns throughout the 2022/23 season.

For the metrics that are easily quantifiable and commonly benchmarked, such as minutes played (Mins), goals scored (Gls), and assists provided (Ast), season totals are used. The total minutes played are favored over the number of games due to the latter's insensitivity to substitutions, both on and off, during the match, offering a more accurate representation of a player's actual time on the pitch.

To account for discipline and on-field behavior, our model includes three variables. The total number of yellow (YC) and red cards (RC) received throughout the season serves as indicators of more serious rule violations. These infractions are expected to exert a negative influence, as they not only weaken the team's immediate competitive capability but may also lead to further player suspensions. This aspect is further elaborated by incorporating the average number of fouls committed per 90 minutes (Fls). The direction of the relationship between fouls and market value is ambiguous since some fouls are generally a negative phenomenon for a team, but some of them may be perceived as tactically appropriate and hence appraised.

For better clarity and comparability across players with varying playing time, all additional individual statistics are presented in a per-90-minute form. Attributes reflecting a player's contribution to chance creation and overall ball possession, such as the number of successful passes (ScPass) and successful dribbles (ScDrib), are expected to positively influence market value. Successful passes signify broad game involvement, whereas successful dribbles indicate proficiency in one-on-one situations. Furthermore, metrics such as the number of shots taken (Sht) and expected goals (xG) are included to capture involvement in finishing off the team's offensive efforts. These variables are presumed to positively affect market value, particularly for players in offensive roles.

Since judging footballers is not limited to the scenarios in which their team has the ball, our model also considers metrics related to defensive performance. These include the number of aerial duels won (AerW), defensive tackles won (TckW), and interceptions (Int), all of which play a pivotal role in maintaining or regaining ball possession. Such accomplishments are expected to be regarded favorable in the context of overall market valuation.

Team statistics

The last category regards team performance, rather than individual achievements. It stems from the understanding that a player's valuation is not solely dependent upon their personal contributions but is also influenced by the overall efficacy of their team. Consequently, the number of goals scored (GS) and goals conceded (GC) by the team are incorporated. The rationale underpinning this is twofold. Firstly, a higher tally of goals scored is indicative of superior offensive performance, and secondly, a greater number of goals conceded suggests deficiencies in defensive capabilities. Furthermore, these statistics exhibit a strong correlation with a team's standing within the league, thereby serving as proxies for the collective success or failure experienced throughout the season.

League quality

To address the nuances distinguishing players within the top 5 European leagues from their counterparts in other leagues, an initial analysis comparing these two subgroups based on the compiled data is suitable. Observations from Table 3.3, showing averages for individual variables, reveal minimum statistical divergence across most of the selected metrics. Nonetheless, major discrepancies occur between the averages of the two market values, underscoring a systemic tendency towards higher valuations of players competing in the more esteemed leagues. This discrepancy is indicative of an inherent expectation of superior quality associated with participation in these competitions. It therefore suggests that the valuation mechanism for players will significantly differ between those within the top 5 leagues and those outside.

Table 3.3: TOP 5/other 5 leagues variables comparison

Variable	**Measurement type**	**Top 5** (Mean)	**Other 5** (Mean)
Dependent variable			
MV season end	€	14 244 564,28	2 926 639,34
Independent variables			
MV season start	€	13 718 766,18	2 274 621,69
Age	years	26,42	25,95
Height	cm	181,55	180,67
Minutes played	season total	1 718,97	1 636,27
Goals	season total	2,55	2,38
Assists	season total	1,82	1,76
Yellow cards	season total	3,84	3,69
Red cards	season total	0,18	0,23
Fouls	per 90 min	1,24	1,23
Successful passes	per 90 min	36,68	34,46
Successful dribbles	per 90 min	0,83	0,76
Shots	per 90 min	1,25	1,22
xG	per 90 min	0,14	0,13
Aerial duels won	per 90 min	1,39	1,40
Tackles won	per 90 min	0,99	1,05
Interceptions	per 90 min	0,90	1,04
National ranking	FIFA ranking	17,91	28,09
Team goals scored	regular league season	52,75	51,98
Team goals conceded	regular league season	50,76	50,16

3.3.3 Dummy variables

To fulfill the aims of this thesis, the employment of various dummy variables was essential for exploring distinctions among various groups of footballers. The first variable, denoted as 'top 5', attributes to players affiliated with clubs within the top 5 European leagues.[4] Secondly, the individual players are sorted by positions following the traditional division into defenders, midfielders, and attackers, represented by the dummy variables M (Midfielders) and A (Attackers), with defenders being the benchmark group.

The variable 'EurC' refers to the player's team participating at least in the group stage of one of the following three European competitions: the UEFA Champions League, the UEFA Europa League, or the UEFA Europa Conference League during the autumn part of the season. The primary intent here is

[4] Premier League, LaLiga, Bundesliga, Serie A and Ligue 1

to demonstrate the impact of providing players from these teams with the opportunity to showcase their abilities on the international stage, contending with clubs from other leagues. This aspect bears benefits, particularly for representatives of lower-quality domestic competitions.

The relative frequencies of each group are shown in Table 3.4.

Table 3.4: Dummy variables relative frequencies

	top 5	**M**	**A**	**EurC**
Frequency	0.594	0.335	0.256	0.341

Chapter 4

Methodology

4.1 Methodology

Considering the attributes of our dataset and its interpretative feasibility, the Ordinary Least Squares (OLS) method is identified as appropriate for our requirements (Wooldridge, 2012). OLS is widely recognized and utilized within the realm of such analyses, as evidenced by its adoption by numerous other authors (Herm *et al.*, 2014; Majewski, 2016; Franck & Nüesch, 2012) aiming to test factors influencing the price of footballers. However, the incorporation of more advanced methodologies, including the Least Absolute Shrinkage and Selection Operator (LASSO), is deemed advantageous for enhancing precision in certain analytical contexts (He *et al.*, 2015). Due to the high number of variables, LASSO will be applied also in this thesis. It will serve as a mechanism to verify the robustness of our model, facilitating careful variable selection and regularization. LASSO achieves this through the incorporation of an L1 penalty on the coefficients, which encourages some coefficients to shrink to zero, effectively reducing the number of variables in the model. This process is governed by a lambda parameter[1] that balances between model complexity and fit (Hastie *et al.*, 2009).

Before the application of the OLS model and the subsequent interpretation of our analytical findings derived therefrom, it is necessary to ensure the adherence to basic assumptions of this method. Our study uses a random sample of a sufficient size, which contains all observations that align with the selection criteria. We assume a linear relationship between the explanatory and explained

[1]In order to perform a LASSO regression, we obtain the ideal value of the lambda parameter using cross-validation.

variables. With regard to the exogeneity assumption, it is evident that MVA is contingent upon unobservable characteristics, such as talent, which may give rise to endogeneity, but we control for it by incorporating the price at the beginning of the season. Thus, our primary focus is on the assurance of the absence of perfect multicollinearity among variables and the homoscedasticity of residuals. The former condition is verifiable by consulting Table 4.1, which illustrates the correlations between numerical variables alongside their Variance Inflation Factor (VIF), serving as a diagnostic of the multicollinearity within the model. It was observed that solely the variables Age and Age^2 exhibit VIF readings substantially surpassing the threshold of 10, a demarcation recommended by Wooldridge (2012). However, this is attributed to their inherent correlation, which is given by definition. Consequently, it does not compromise the integrity of our model. The latter condition can be inspected, for example, by employing the Breusch-Pagan test. Its results confirmed the presence of heteroscedasticity within our model. To maintain the validity of the statistical inference, heteroscedasticity-robust standard errors will be calculated and incorporated into the analysis.

4.2 Model

With these considerations in place, the OLS method is employed to estimate the end-of-season market value by integrating all variables specified in the 'Data' section, resulting in the following regression model

$$\log(\mathrm{MVA}_i) = \beta \mathrm{X}_i + \epsilon_i,$$

where $\mathrm{X}_i = \{X_{1i}, X_{2i}, ..., X_{ki}\}$ is the vector of explanatory variables for individual i, and k represents the number of explanatory variables. The full content of the model, along with its gradually expanding variations for the joint sample of all players, is described in the Appendix.

To preserve model linearity, logarithmic transformation is applied to both the pre-season (MVB) and post-season (MVA) market values, owing to their strongly right-skewed distribution. This approach is consistent with methodologies used in previous studies (Franck & Nüesch, 2012; Müller *et al.*, 2017). Given this, our analysis primarily leverages the log-lin specification, interpreted as semi-elasticity, when assessing the results. This implies that a one-unit change in an independent variable results in a $\beta_k \times 100\%$ modification in MVA.

The interrelation between the two market values adopts a log-log specification, indicative of a percentage variation in MVA resultant from a 1% fluctuation in MVB.

Table 4.1: Correlation matrix

	MVB	Age	Age^2	Hght	Mins	Gls	Ast	YC	RC	Fls	ScPass	ScDrib	Sht	xG	AerW	TckW	Int	NR	GS	GC	VIF
MV season start	1.000	−0.113	−0.126	0.026	0.219	0.337	0.307	0.064	−0.049	−0.089	0.199	0.190	0.199	0.222	−0.074	−0.095	−0.154	−0.137	0.373	−0.256	3.42
Age		1.000	0.995	0.036	0.031	0.010	−0.027	0.058	0.057	−0.056	0.079	−0.260	−0.074	−0.015	0.117	−0.103	0.001	−0.005	−0.047	0.019	130.21
Age squared			1.000	0.035	0.016	0.004	−0.032	0.046	0.054	−0.055	0.080	−0.254	−0.070	−0.009	0.115	−0.104	−0.002	−0.014	−0.044	0.018	131.70
Height				1.000	0.084	−0.030	−0.179	0.074	0.024	−0.019	0.071	−0.319	−0.119	−0.014	0.523	−0.120	0.118	−0.007	−0.003	0.006	1.53
Minutes played					1.000	0.345	0.411	0.546	0.138	−0.209	0.159	−0.048	−0.075	−0.035	−0.039	−0.053	0.083	−0.065	0.046	−0.062	2.58
Goals						1.000	0.508	0.076	−0.006	−0.038	−0.277	0.266	0.610	0.713	−0.035	−0.293	−0.385	−0.009	0.236	−0.110	3.19
Assists							1.000	0.103	−0.024	−0.097	−0.029	0.317	0.357	0.300	−0.252	−0.106	−0.235	−0.045	0.287	−0.193	1.86
Yellow cards								1.000	0.284	0.245	0.115	−0.077	−0.125	−0.126	0.012	0.135	0.162	−0.046	−0.095	−0.017	2.02
Red cards									1.000	0.075	0.052	−0.039	−0.063	−0.065	0.038	0.053	0.107	−0.012	−0.074	0.022	1.11
Fouls										1.000	−0.195	0.067	0.166	0.128	0.127	0.179	−0.056	0.008	−0.086	0.002	1.60
Successful passes											1.000	−0.177	−0.442	−0.442	−0.084	0.227	0.343	−0.130	0.428	−0.349	2.57
Successful dribbles												1.000	0.468	0.308	−0.330	0.036	−0.305	0.003	0.122	−0.088	1.75
Shots													1.000	0.841	−0.005	−0.295	−0.548	0.025	0.167	−0.098	4.95
xG														1.000	0.092	−0.365	−0.514	0.017	0.197	−0.103	5.50
Aerial duels won															1.000	−0.143	0.081	0.054	−0.056	0.041	1.70
Tackles won																1.000	0.412	−0.036	−0.031	0.056	1.51
Interceptions																	1.000	0.008	−0.069	0.039	1.78
National ranking																		1.000	−0.049	0.098	1.08
Team goals scored																			1.000	−0.568	2.43
Team goals conceded																				1.000	1.72

Chapter 5

Results

5.1 Joint regression results

Initially, all players are analyzed together. The first regression only included MVB and basic observable characteristics. Subsequently, match statistics are incorporated, followed by team-related data and dummy variables. Table 5.1 presents all the results obtained by this procedure.

The performance of the models can be assessed using adjusted R-squared. This coefficient shows high values for all three regressions, ranging from slightly over 0.87 for the simplest model to almost 0.93 for the most complex one. It suggests that the model's quality increases as more factors are added. This statement is supported by the fact that some estimates changed remarkably, particularly in the transition between the second and third models. The high F-statistic values also confirm the significant overall impact of all the included explanatory variables on MVA.

The most significant driver, in terms of strength and statistical significance, is the market value at the beginning of the season. This confirms our assumption that including it in the model represents a crucial part of explaining player value based on statistics from a single season. The estimate of 0.61 indicates a strong dependence of MVA on the previous valuation. A 1% increase in MVB results in a 0.61% higher price at the end of the season ceteris paribus. This connection is not surprising given the strong linear relationship between the two market values.

The coefficients representing Age and Age^2 substantially differ from prior

Table 5.1: Regressions with all players

	Dependent variable: log(MVA)		
	Model 1	Model 2	Model 3
Intercept	6.488***	7.022***	7.761***
	(0.720)	(0.696)	(0.617)
log(MVB)	0.881***	0.794***	0.610***
	(0.009)	(0.011)	(0.016)
Age	-0.338***	-0.331***	-0.178***
	(0.044)	(0.041)	(0.034)
Age^2	0.005***	0.005***	0.002**
	(0.001)	(0.001)	(0.001)
Hght	0.005***	0.004**	0.003*
	(0.002)	(0.002)	(0.002)
NR	-0.001**	-0.001**	0.0001
	(0.000)	(0.000)	(0.000)
Mins		0.0003***	0.0003***
		(0.000)	(0.000)
Gls		0.023***	0.025***
		(0.005)	(0.005)
Ast		0.015**	0.010*
		(0.006)	(0.006)
YC		0.003	0.008
		(0.005)	(0.005)
RC		-0.068***	-0.038**
		(0.022)	(0.019)
Fls		0.026	-0.028
		(0.025)	(0.022)
ScPass		0.008***	0.003***
		(0.001)	(0.001)
ScDrib		0.084***	0.069***
		(0.021)	(0.019)
Sht		-0.006	-0.002
		(0.025)	(0.023)
xG		0.324*	0.288*
		(0.177)	(0.159)
AerW		0.021*	0.017
		(0.012)	(0.011)
TckW		-0.036	-0.012
		(0.028)	(0.026)
Int		0.005	0.036
		(0.027)	(0.024)
GS			0.007***
			(0.001)
GC			-0.009***
			(0.001)
EurC			0.076***
			(0.028)
top5			0.610***
			(0.035)
M			0.075***
			(0.028)
A			0.035
			(0.042)
n	1952	1952	1952
R^2	0.8726	0.9048	0.927
Adjusted R^2	0.8723	0.9039	0.9261
Resid. Std. Error	0.5619 (df = 1946)	0.4874 (df = 1933)	0.4272 (df = 1927)
F Statistic	2665*** (df = 5; 1946)	1020*** (df = 18; 1933)	1020*** (df = 24; 1927)

Note: $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Robust SE in parentheses

studies and our anticipations. Despite yielding statistically significant results[1], they contradict the expected inverted U-shape across all three models. However, the minimum point of the parabola delineating our emergent relationship appears to be well beyond the age of 40 when calculated based on the estimates from model 3, which we perceive as an irrelevant turning point for our dataset with a maximum of 40. Consequently, within this analysis, age exhibits a solely negative impact with diminishing intensity. Regarding the rest of the basic observable characteristics (Hght and NR), their significance was lost upon the addition of further terms to the regression. Only height retained it in the final modification at the 0.1 level, with each additional centimeter of height translating to an approximate 0.3% increase in MVA.

Upon examining the estimates for variables representing player's match statistics, we find a significant and pronounced positive influence of minutes spent on the pitch and the number of goals scored on MVA, corroborating expectations and consistency with preceding studies. This equates to a 2.7% enhancement in MVA for every extra full game played, equivalent to 90 minutes. Each additional goal scored invokes an approximately 2.5% increase. In terms of strong significance even at the 0.01 level, they are supplemented by two more variables related to playmaking, namely the number of successful passes and dribbles. Predictably, both lead to an upswing in MVA, with ScDrib being quite substantial at 6.9% for every additional one.

The sole further significant metric is red cards exerting an expected and considerably adverse effect, reducing the player's value by 3.8%. This can be linked to the weakening of the team consequent to such penalty. Among the remaining personal statistics, only assists and xG turned out informative at least at the 0.1 level. However, these effects remain modest in light of their estimates and average values. Other in-game factors culminate as statistically insignificant.

Variables related to the player's team emerged as significant, illustrating that team achievements in the domestic league, denoted by GS and GC, alongside participation in the main stages of European cups, substantially influence the dependent variable. Each goal scored adds approximately 0.7%, whereas each goal conceded subtracts 0.9%. The coefficient of the top 5 dummy provides further evidence of higher player prices in these leagues, specifically by as

[1]Unless indicated otherwise within this text, the designation "statistically significant" refers to statistical significance at the 5% level.

much as 61%. The evidence also suggests increased valuation for the other two positions relative to defenders, but statistically significant only for midfielders.

5.2 LASSO

Before delving into the distinctions between the subsamples, it is reasonable to perform a robustness check to evaluate the integrity of the model that will be employed repeatedly. Due to the extensive number of variables in the model, LASSO regression was identified as a suitable method for examination. This technique enables us to thoroughly regularize and select variables by diminishing the magnitude of certain coefficients and nullifying others (Hastie *et al.*, 2009).

In this context, the LASSO regression method was used to provide the advisability of excluding specific factors from the model. Initially, the LASSO regression was executed to obtain the individual coefficients. The analysis indicates that the model's generalizability and interpretability could be enhanced by eliminating variables Sht, NR, and Age^2. Subsequent validation of this proposition was conducted by rerunning the reduced model and scrutinizing the new estimates alongside the Adjusted R^2 and the Akaike Information Criterion (AIC), which revealed that the main improvement in results occurs by omitting Sht and NR. This can be attributed to LASSO regression's tendency to penalize variables for their high correlation with others, which would justify the exclusion of Age^2 from our model. Moreover, taking into account the inverted U-shaped effects of age observed in previous research, it was decided to keep Age^2 in the model.

Thus, the refined model for subsequent analysis omits only Sht and NR. The equation and outcomes of the resulting reduced regression for the full dataset are documented in the Appendix. The examination of fit measures and the vast majority of coefficients alongside their statistical significance demonstrated minimal distinctions from the original model. The only notable divergence occurred in the variable xG, which exhibited an increase in predictive relevance following the omission of Sht and was statistically significant at the 0.05 level. This adjustment is rationalized by the heightened correlation between the omitted and retained variables, diminishing their collective efficacy in the original model.

Based on the findings collected through the LASSO regression, it can be affirmed that our findings from the joint model remain valid. However, the primary reason for the robustness check is the concern of overfitting the model used, due to the high number of variables involved. Following the conclusions, dimensionality reduction will be conducted to make the findings more generalized for any other data. Therefore, in the remainder of the regressions, where the dataset gets segmented into smaller subsamples for comparative evaluation, the potential improvements arising from the robustness check are applied. Consequently, we exclude the variables NR and Sht from further analysis.

5.3 Discrepancies between TOP 5 and the rest of the leagues

Comparing the top 5 European football leagues and the remaining competitions using descriptive statistics showed that the primary differential attribute was the market valuation of players. This suggests the presence of differences in pricing mechanisms. Given that the essence of this thesis is to examine the variables influencing this valuation, it becomes necessary to confront the two subsamples through analysis of the two separated regressions.

Before delving into the inspection of the coefficients of the two OLS models, a Chow test was conducted to confirm the anticipation of disparities between them. The test outcomes indicated the existence of a structural break based on the top 5 dummy variable, with the p-value for rejecting the null hypothesis[2] being negligibly small. Therefore, a closer examination of the two regression results presented in Table 5.2 shall unveil differences in the relevance of certain factors. Observations from the adjusted R^2 values suggest a marginally reduced explanatory capacity compared to the unified model. However, the capability of both models to explain 90 and 89 percent, respectively, of the variance in the dependent variable, underscores a good fit.

Mirroring the joint model, the previous market value keeps its significance. However, this influence is more pronounced within the top 5 leagues, where a 1% enhancement leads to a 0.63% increase in MVA, as opposed to 0.56% in the other leagues. This indicates that the influence of other factors should be stronger in the less prestigious leagues. Discrepancies also occur in other fundamental attributes. Notably, age is a more pivotal factor in the top 5 leagues, with both Age and Age^2 demonstrating increased significance compared to the lesser leagues, where neither Age nor Age^2 reaches the 0.05 significance thresh-

[2]The null hypothesis assumes equal coefficients for both models.

Table 5.2: Regressions TOP 5/other 5 leagues

	Dependent variable: log(MVA)	
	Top 5	Out of top 5
Intercept	8.916***	7.145***
	(0.765)	(1.003)
log(MVB)	0.635***	0.563***
	(0.020)	(0.028)
Age	-0.224***	-0.106*
	(0.041)	(0.055)
Age^2	0.002***	0.0002
	(0.001)	(0.001)
Hght	0.001	0.005*
	(0.002)	(0.003)
Mins	0.0003***	0.0003***
	(0.000)	(0.000)
Gls	0.023***	0.027***
	(0.006)	(0.008)
Ast	0.011	0.008
	(0.007)	(0.009)
YC	0.009	0.004
	(0.005)	(0.008)
RC	-0.033	-0.030
	(0.021)	(0.036)
Fls	-0.025	-0.030
	(0.028)	(0.032)
ScPass	0.002	0.005***
	(0.001)	(0.002)
ScDrib	0.040*	0.107***
	(0.023)	(0.027)
xG	0.066	0.539***
	(0.177)	(0.179)
AerW	0.035**	-0.009
	(0.015)	(0.015)
TckW	-0.052	0.032
	(0.032)	(0.039)
Int	-0.021	0.091**
	(0.029)	(0.039)
GS	0.007***	0.007***
	(0.001)	(0.002)
GC	-0.005***	-0.013***
	(0.001)	(0.002)
EurC	0.098***	0.110**
	(0.034)	(0.051)
M	0.097***	0.053
	(0.032)	(0.045)
A	0.037	0.037
	(0.051)	(0.062)
n	1159	793
R^2	0.9031	0.8894
Adjusted R^2	0.9013	0.8864
Resid. Std. Error	0.41 (df = 1137)	0.4385 (df = 771)
F Statistic	504.5*** (df = 21; 1137)	295.2*** (df = 21; 771)

Note: *p<0.1; **p<0.05; ***p<0.01
Robust SE in parentheses

old. Nevertheless, the age-related effect stays negative. In contrast, height presents modest explanatory utility at 10% significance exclusively outside the top 5 leagues.

Estimates for the minutes played and goals yielded outcomes similar to the aggregate model but with slightly stronger effects outside the top 5 leagues. Furthermore, no major differences were detected between the two subsamples regarding assists and disciplinary metrics, with all these variables losing significance across both models.

The divergence between the models appears predominantly in statistics related to ball possession. Evidently, in the leagues of lower prestige, this gameplay aspect is deemed more important, as indicated by the positive and significant impact of successful passes, dribbles, and the expected goals metric. Each additional successful pass in a match corresponds to an increase in MVA by approximately 0.5 percent, a successful take-on nearly 11%, and an expected goal by almost 54%. It is important to note that these variables must be contextualized within the range of their average fluctuation. In the top 5 leagues, only successful dribbles retained their informativeness at the 0.1 level, with a generally diminished impact across these metrics.

Further distinctions were observed within the domains of physical play and defensive contributions. In the elite leagues, victorious aerial duels are appreciated more, catalyzing a 3.5% MVA rise, while in other leagues, interceptions notably elevate player valuation by approximately 9%. Tackles won do not hold statistical significance in any group.

Regarding the variables concerning player's team success, a consensus in their perceived importance was evident across both subsamples. The magnitude of the positive effects of GS and EurC are very similar, albeit goals conceded exert nearly thrice as potent negative influence on players outside the top 5 leagues. As for positions on the field, the only group of players valued higher compared to the rest are midfielders in the elite leagues, by nearly 10 percent.

To summarize the main differences, it can be stated that there is a stronger influence of previous market value and age on the post-season valuation of a player within the elite leagues. This implies that a player's value in other leagues is subject to greater fluctuations based on performance, which is corroborated by the considerably more pronounced impact of ball possession statistics. We also find some discrepancies in defensive metrics and team-related variables. Finally, it seems that midfielders are the most valued position in the top 5 leagues.

5.4 Market value by positions

Another division used to analyze disparities was based on individual playing positions. The results of the regressions corresponding to these subdivisions, as reported in Table 5.3, indicate a marginally enhanced model fit with adjusted R^2 values surpassing 92% across the board. The first part of the coefficients revealed no notable variances between the positions nor any deviation from prior results. Nonetheless, the impact of the previous market value appears to be weaker for midfielders compared to the rest. The analysis also reaffirms that age exerts a uniformly negative impact on player valuation, with the quadratic term of age demonstrating significance only for forwards. Height's lack of relevance persists across all playing positions.

The analysis further underscores playing time as an integral player value determinant across all positions, though with a varying impact. Defenders and midfielders are awarded with approximately 3% rise in MVA for a full match played, while in the case of forwards, this influence gets reduced to less than 2 percent for each additional 90 minutes. Consistent with expectations, goals scored are particularly pivotal for offensive roles, thus for midfielders and forwards, each accorded a 3.4 percent valuation increase per goal. Additionally, assists bear significance for attackers, contributing an added valuation of 2.5%, while discipline during the match appears not to hold significance in any position.

At least at the 0.05 significance level, the sole other game statistics of note are successful passes for defenders, raising MVA by 0.3% for each, and successful dribbles for midfielders and forwards, yielding enhancement by 13.4 and 6.3 percent, respectively. This aligns with the presumption, as modern defenders are typically responsible for a high volume of passes, playing a vital role in orchestrating team build-up play. On the other hand, midfielders and attackers are praised for their ability to create advantageous situations, such as through successful take-ons, which turned out significant for them. It is noteworthy that more position-specific statistics, such as xG for forwards or defensive performance indicators in the case of defenders, failed to reach statistical significance.

In relation to team-associated variables, the coefficients largely reflect those observed in preceding models, with the exception of forwards. Here, the signif-

Table 5.3: Regressions by position

	Dependent variable: log(MVA)		
	Defenders	Midfielders	Attackers
Intercept	7.968***	7.609***	8.337***
	(1.055)	(1.091)	(0.963)
log(MVB)	0.628***	0.572***	0.633***
	(0.026)	(0.029)	(0.028)
Age	-0.171***	-0.131**	-0.220***
	(0.058)	(0.057)	(0.054)
Age^2	0.001	0.001	0.002***
	(0.001)	(0.001)	(0.001)
Hght	0.001	0.003	0.002
	(0.003)	(0.003)	(0.004)
Mins	0.0003***	0.0003***	0.0002***
	(0.000)	(0.000)	(0.000)
Gls	0.021	0.034***	0.034***
	(0.015)	(0.012)	(0.007)
Ast	0.005	0.002	0.025**
	(0.011)	(0.009)	(0.010)
YC	0.005	0.008	0.003
	(0.007)	(0.007)	(0.010)
RC	-0.033	-0.029	-0.070
	(0.024)	(0.038)	(0.047)
Fls	-0.015	-0.013	-0.025
	(0.043)	(0.032)	(0.041)
ScPass	0.003**	0.0004	0.006*
	(0.001)	(0.002)	(0.003)
ScDrib	0.014	0.134***	0.063**
	(0.041)	(0.031)	(0.029)
xG	0.337	0.045	0.266
	(0.407)	(0.270)	(0.183)
AerW	0.010	0.001	0.021
	(0.020)	(0.029)	(0.017)
TckW	-0.033	0.021	-0.040
	(0.036)	(0.044)	(0.052)
Int	0.033	0.044	0.083
	(0.033)	(0.046)	(0.071)
GS	0.007***	0.009***	0.004*
	(0.002)	(0.002)	(0.002)
GC	-0.009***	-0.011***	-0.008***
	(0.002)	(0.002)	(0.002)
EurC	0.083**	0.127**	0.016
	(0.040)	(0.053)	(0.056)
top5	0.562***	0.671***	0.579***
	(0.051)	(0.061)	(0.072)
n	799	653	500
R^2	0.929	0.9241	0.9322
Adjusted R^2	0.9272	0.9217	0.9294
Resid. Std. Error	0.4088 (df = 778)	0.4468 (df = 632)	0.4252 (df = 479)
F Statistic	509.1*** (df = 20; 778)	384.6*** (df = 20; 632)	329.4*** (df = 20; 479)

Note: *p<0.1; **p<0.05; ***p<0.01
Robust SE in parentheses

icance of the team's goal-scoring unexpectedly decreased, and even the effect of participation in European competitions does not emerge as determinative. A comparison of the other two positions reveals stronger team-related effects for midfielders. Affiliation with the top 5 leagues continues to significantly elevate player valuations across all positions, most notably for midfielders, who see up to a 67% increase in their market valuation.

Therefore, the main differences between the positions start with the number of minutes spent on the pitch, where this influence weakens for forwards. Conversely, attackers, together with midfielders, are valued for scoring goals and successful take-ons. For defenders, only successful passes are determinative in this respect. Lastly, team success along with top 5 league affiliation exert the strongest influence on midfielders.

5.5 Differences within the TOP 5 leagues

To support the hypothesis that analogous to the distinctions detected between players in the top 5 leagues and those outside, variations within the top 5 leagues themselves are also observable, the final comparative regression analysis was specifically aimed at examining this assertion, as suggested by Ante (2019). The outcomes of the five models, each representing an individual league, are listed in Table 5.4.

A reduction in the magnitude of MVB can be observed in the results, with the influence declining below 43% for Ligue 1 players. This may indicate an increased impact of other variables. Regarding the effect of age, particularly within the Premier League and Ligue 1, an anticipated inverted U-shape was observed, with age exerting a positive impact on players up to 21.5 years in Ligue 1, as denoted by the coefficient values. However, the overall significance of age diminishes, and the identified turning points in these instances are situated at lower thresholds than the minimum age present in our dataset, indicating that, alongside the other three leagues exhibiting classical U-shape dynamics, age predominantly has a negative influence across the vast majority of players. Height does not seem to be a noteworthy determinant in any of the top 5 leagues.

The quantity of played minutes remains a significant driver in all leagues, with its impact oscillating from approximately 2.2% per 90 minutes in the English Premier League to over 3.5% in Serie A. Goals turned out to be positively significant in Spain (4.7%), Germany (3.5%), and France (4.9%). In Ligue 1,

Table 5.4: Regressions for each of the TOP 5 leagues

	Dependent variable: log(MVA)				
	Premier League	LaLiga	Bundesliga	Serie A	Ligue 1
Intercept	7.114***	11.732***	8.384***	10.499***	8.035***
	(1.382)	(1.442)	(1.122)	(1.646)	(1.456)
log(MVB)	0.500***	0.471***	0.546***	0.542***	0.428***
	(0.041)	(0.041)	(0.033)	(0.036)	(0.062)
Age	0.086	-0.197**	-0.152**	-0.336***	0.004
	(0.104)	(0.090)	(0.061)	(0.064)	(0.059)
Age^2	-0.004**	0.002	0.001	0.004***	-0.002**
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Hght	0.005	-0.003	0.003	0.005	0.002
	(0.004)	(0.004)	(0.004)	(0.006)	(0.004)
Mins	0.0002***	0.0004***	0.0003***	0.0004***	0.0003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Gls	0.006	0.047***	0.035***	0.016	0.049***
	(0.007)	(0.014)	(0.013)	(0.011)	(0.015)
Ast	0.011	0.035*	-0.001	0.009	0.036**
	(0.011)	(0.018)	(0.012)	(0.016)	(0.015)
YC	0.015	-0.003	-0.019*	0.020*	0.012
	(0.011)	(0.008)	(0.011)	(0.012)	(0.013)
RC	-0.123	-0.007	0.119***	-0.086*	0.046
	(0.105)	(0.031)	(0.043)	(0.044)	(0.041)
Fls	-0.048	-0.015	0.014	-0.022	0.034
	(0.045)	(0.034)	(0.047)	(0.060)	(0.058)
ScPass	0.009***	0.004	0.003	-0.001	0.007*
	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)
ScDrib	0.039	0.067	0.037	-0.003	0.084**
	(0.041)	(0.050)	(0.041)	(0.050)	(0.037)
xG	0.554*	0.219	-0.041	0.693*	-0.494
	(0.287)	(0.336)	(0.316)	(0.354)	(0.396)
AerW	0.010	0.050*	0.042*	-0.030	0.061**
	(0.030)	(0.030)	(0.024)	(0.039)	(0.027)
TckW	-0.002	-0.106*	-0.073	0.086	-0.197***
	(0.056)	(0.056)	(0.049)	(0.058)	(0.068)
Int	0.003	0.075	0.007	-0.036	0.003
	(0.051)	(0.061)	(0.055)	(0.050)	(0.068)
GS	0.002	0.004	0.014***	0.015***	0.016***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)
GC	-0.008***	-0.009***	-0.002	-0.002	-0.004*
	(0.002)	(0.003)	(0.005)	(0.004)	(0.002)
EurC	0.112*	0.223***	0.079	0.233***	0.329***
	(0.062)	(0.065)	(0.066)	(0.072)	(0.083)
M	0.139**	0.077	0.093	0.160**	0.166**
	(0.064)	(0.059)	(0.062)	(0.066)	(0.064)
A	0.276**	-0.031	0.008	0.0003	0.055
	(0.109)	(0.090)	(0.094)	(0.119)	(0.099)
n	269	262	217	206	205
R^2	0.9107	0.9086	0.9394	0.9368	0.9313
Adjusted R^2	0.9031	0.9006	0.9329	0.9296	0.9234
Resid. Std. Error	0.3384 (df = 247)	0.379 (df = 240)	0.3272 (df = 195)	0.3447 (df = 184)	0.3686 (df = 183)
F Statistic	119.9*** (df = 21; 247)	113.6*** (df = 21; 240)	144.1*** (df = 21; 195)	129.8*** (df = 21; 184)	118.1*** (df = 21; 183)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Robust SE in parentheses

assists also enhance the market value, contributing with a 3.6% increase for each additional one.

Concerning compliance with the rules on the pitch, the significance was solely attributed to red cards within the Bundesliga, where, against the anticipation, each instance leads to an 11.9% increase in market value. The rest of the game statistics did not demonstrate notable significance across the individual leagues. Exceptions include the positive effects of successful passes for Premier League players, followed by successful dribbles and victorious aerial duels for players in Ligue 1, with an unexpectedly negative impact observed for tackles won within the same league.

Team statistics exhibit a division between leagues where only scored goals hold significance—namely, the Bundesliga, Serie A, and Ligue 1—and those where goals conceded are determinative, specifically the Premier League and LaLiga. Participation in European competitions is taken into account in Spain, Italy, and France, where this variable induces a substantial market value enhancement of 32.9%. Relative to defenders, midfielders frequently exhibit a significantly elevated valuation across all leagues, with the exception of LaLiga and the Bundesliga. Notably, the most pronounced premium in market valuation is observed for Premier League attackers, where the differential escalates to as much as 27.6% in comparison to defenders.

Chapter 6

Conclusion

The objective of this thesis is to compare the mechanisms underpinning the market valuation of football players within the top 5 European leagues with their lower ranked counterparts as measured by the UEFA coefficient. To this end, an analytical framework using the Ordinary Least Squares method was employed, scrutinizing a dataset comprising 1952 footballers along with their performance and market data from the 2022/23 season. This analysis aims to uncover both the shared and distinct factors influencing their valuations across these two groups. In the subsequent comparisons, these differences are also examined between individual positions on the pitch and within the top 5 leagues.

The main contribution of this paper lies in complementing research on the analysis of footballers' market value drivers with deeper insight into previously unexplored disparities between elite European leagues and other competitions. This is especially relevant given the notably higher valuations observed in the top 5 leagues. Therefore, this paper primarily addresses the assumed discrepancies in the mechanisms that cause this gap in player prices. This presumption is confirmed by the observed differences between the main factors driving market values. The outcomes hold considerable value for stakeholders engaged in transfer negotiations, especially when clubs from elite leagues acquire players from less prestigious competitions. The same stands for football fans keen on comprehending the transfer market situation.

In the joint model containing players from all leagues and positions, the players worth before the season stood out as the dominant factor in deciding their final market valuation, with age surprisingly exhibiting only a negative impact. Individual player statistics such as minutes played, goals scored, successful passes, and dribbles were identified as statistically significant in determining player value. From a team perspective, positive contributions arose from goals scored and participation in European competitions, whereas goals conceded bear negative effect. A higher valuation of midfielders relative to other positions was observed, accompanied by a notable premium on players affiliated with clubs in the top 5 leagues.

The core of this paper, focusing on the discrepancies in valuation mechanisms between the top 5 leagues and the other competitions, unveiled substantial divergences not only in total valuation figures but also in the underlying determinants. The analysis demonstrated a more pronounced effect of the preceding market value observed within the top-tier leagues. Age maintains its significance only within these elite leagues, although still connected to negative impact. Metrics such as minutes played and goals scored exhibited a positive influence on the player's valuation for both groups, slightly stronger outside the top 5 leagues. Significant variances were observed in game statistics related to ball possession, with successful passes, dribbles, and the xG indicator gaining relevance only in the leagues of lesser prestige. Defensive metrics also displayed league-specific importance, with interceptions having an impact in the less prestigious leagues and aerial duels being valued in the top 5 leagues only. The positive effects of team goals scored and participation in European competitions are almost identical. However, the negative effect of goals conceded is much stronger for players outside the top 5 leagues. Lastly, in the elite competitions, midfielders were found to be valued higher.

Regarding the differences between positions, further analysis suggested variously strong influences of the previous market valuation and minutes played across the regressions. Surprisingly, in the case of individual positions on the pitch, except for successful passes to defenders and dribbles for the other two positions, none of the other defensive or offensive statistics achieved significance. Yet, the relevance of team performance metrics and a pronounced valuation premium for players from top-tier leagues were reaffirmed. The only exceptions are attackers, for whom goals scored by the team and participation in European competitions are irrelevant. Among the top 5 leagues, Ligue 1 exhibits the most distinct valuation mechanism. This is because, unlike the rest, more game statistics show significance there. Specifically, successful takeons combined with victorious aerial duels and tackles. Other differences appear rather only in the changing magnitude of the influence of minutes played and in the varying significance of goals scored and team-related variables. Midfielders appear to be generally perceived as a higher valued position.

It must be acknowledged that this work is subject to certain limitations. These mainly relate to the impossibility of collecting data for all matches played during the season, including cup or play-off matches, due to the inclusion of less prestigious competitions. The unobserved effect of the FIFA World Cup, played exceptionally in the middle of the season, could also play a minor role.

Possible future research on this issue could attempt to further differentiate various positions on the pitch, thus focusing even more on the distinctiveness of each role. Additionally, an analysis relating to the valuation of players transferring or being loaned out within a given season, who have been omitted from this thesis, could yield interesting results. Such an investigation might unveil the actual impact of club changes on players' worth, thereby enriching our understanding of football's market dynamics.

Bibliography

ANTE, L. (2019): "Determinants of transfers fees: Evidence from the five major european football leagues." *University of Hamburg: Hamburg, Germany* .

GARCIA-DEL BARRIO, P. & F. PUJOL (2007): "Hidden monopsony rents in winner-take-all markets—sport and economic contribution of spanish soccer players." *Managerial and Decision Economics* **28(1)**: pp. 57–70.

BERRI, D. J. (2006): "Economics and the national basketball association: Surveying the literature at the tip-off." In J. FIZEL (editor), "The handbook of sports economics research," pp. 21–48. Armonk, NY: M.E. Sharpe, Inc.

BRANDES, L. & E. FRANCK (2012): "Social preferences or personal career concerns? field evidence on positive and negative reciprocity in the workplace." *Journal of Economic Psychology* **33(5)**: pp. 925–939.

BRYSON, A., B. FRICK, & R. SIMMONS (2013): "The returns to scarce talent: Footedness and player remuneration in european soccer." *Journal of Sports Economics* **14(6)**: pp. 606–628.

CARMICHAEL, F., D. FORREST, & R. SIMMONS (1999): "The labour market in association football: who gets transferred and for how much?" *Bulletin of Economic Research* **51(2)**: pp. 125–150.

COATES, D. & P. PARSHAKOV (2022): "The wisdom of crowds and transfer market values." *European Journal of Operational Research* **301(2)**: pp. 523–534.

CVRČEK, J. (2021): *Determinants of Football Players' Market Value.* Bachelor thesis (bc.), Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis Supervisor Mgr. Roman Kalabiška.

Deutscher, C. (2018): "Determining the drivers of player valuation and compensation in professional sport: traditional economic approaches and emerging advances." In "Personnel Economics in Sports," pp. 73–87. Edward Elgar Publishing.

Di Salvo, V., R. Baron, H. Tschan, F. C. Montero, N. Bachl, & F. Pigozzi (2006): "Performance characteristics according to playing position in elite soccer." *International journal of sports medicine* pp. 222–227.

Franceschi, M., J.-F. Brocard, F. Follert, & J.-J. Gouguet (2023): "Determinants of football players' valuation: A systematic review." *Journal of Economic Surveys* .

Franck, E. & S. Nüesch (2012): "Talent and/or popularity: what does it take to be a superstar?" *Economic Inquiry* **50(1)**: pp. 202–216.

Frick, B. (2011): "Performance, salaries, and contract length: Empirical evidence from german soccer." *International Journal of Sport Finance* **6(2)**: p. 87.

Hakes, J. K. & R. D. Sauer (2006): "An economic evaluation of the moneyball hypothesis." *Journal of Economic Perspectives* **20(3)**: pp. 173–185.

Hastie, T., R. Tibshirani, J. H. Friedman, & J. H. Friedman (2009): *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

He, M., R. Cachucho, & A. J. Knobbe (2015): "Football player's performance and market value." In "Mlsa@ pkdd/ecml," pp. 87–95.

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

Lewis, M. (2003): *Moneyball: The Art of Winning an Unfair Game.* W. W. Norton Company.

Majewski, S. (2016): "Identification of factors determining market value of the most valuable football players." *Central European Management Journal* **24(3)**: pp. 91–104.

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

Pacák, V. (2021): *The Salary Discrimination in NBA.* Bachelor thesis, Charles University, Faculty of Social Sciences, Institut of Economic Studies.

Peck, K. (2012): *Salary Determination in the National Hockey League: Restricted, Unrestricted, Forwards, and Defensemen.* Honors thesis, Western Michigan University.

Peeters, T. (2018): "Testing the wisdom of crowds in the field: Transfermarkt valuations and international soccer results." *International Journal of Forecasting* **34(1)**: pp. 17–29.

Prockl, F., B. Frick *et al.* (2018): "Information precision in online communities: player valuations on www. transfermarkt. de." *International Journal of Sport Finance* **13(4)**: pp. 319–335.

Reidy, P. (2023): "Tv ratings: how many people watched the 2023 champions league final?" Accessed: 19 April 2024.

Surowiecki, J. (2005): *The wisdom of crowds.* Anchor.

Wooldridge, J. M. (2012): *Introductory Econometrics: A Modern Approach.* Mason, Ohio: South-Western Cengage Learning, 5 edition.

Český svaz ledního hokeje z.s. (2022): "Přestupní řád pro kluby extraligy a i. ligy." Online. Dostupné z: `https://www.ceskyhokej.cz/data/document/file/2022-06-18-prestupni-rad-pro-kluby-extraligy-a-i.-ligy.pdf`.

Appendix A

Joint model

A.1 Model 1

$$\log(\mathrm{MVA}_i) = \beta_0 + \beta_1 \log(\mathrm{MVB}_i) + \beta_2 \mathrm{Age}_i + \beta_3 \mathrm{Age^2}_i + \beta_4 \mathrm{Hght}_i + \beta_5 \mathrm{NR}_i + \epsilon_i$$

A.2 Model 2

$$\begin{aligned}\log(\mathrm{MVA}_i) = {} & \beta_0 + \beta_1 \log(\mathrm{MVB}_i) + \beta_2 \mathrm{Age}_i + \beta_3 \mathrm{Age^2}_i + \beta_4 \mathrm{Hght}_i + \beta_5 \mathrm{NR}_i \\ & + \beta_6 \mathrm{Mins}_i + \beta_7 \mathrm{Gls}_i + \beta_8 \mathrm{Ast}_i + \beta_9 \mathrm{YC}_i + \beta_{10} \mathrm{RC}_i \\ & + \beta_{11} \mathrm{Fls}_i + \beta_{12} \mathrm{ScPass}_i + \beta_{13} \mathrm{ScDrib}_i + \beta_{14} \mathrm{Sht}_i + \beta_{15} \mathrm{xG}_i \\ & + \beta_{16} \mathrm{AerW}_i + \beta_{17} \mathrm{TckW}_i + \beta_{18} \mathrm{Int}_i + \epsilon_i\end{aligned}$$

A.3 Model 3

$$\begin{aligned}\log(\mathrm{MVA}_i) = {} & \beta_0 + \beta_1 \log(\mathrm{MVB}_i) + \beta_2 \mathrm{Age}_i + \beta_3 \mathrm{Age^2}_i + \beta_4 \mathrm{Hght}_i + \beta_5 \mathrm{NR}_i \\ & + \beta_6 \mathrm{Mins}_i + \beta_7 \mathrm{Gls}_i + \beta_8 \mathrm{Ast}_i + \beta_9 \mathrm{YC}_i + \beta_{10} \mathrm{RC}_i \\ & + \beta_{11} \mathrm{Fls}_i + \beta_{12} \mathrm{ScPass}_i + \beta_{13} \mathrm{ScDrib}_i + \beta_{14} \mathrm{Sht}_i + \beta_{15} \mathrm{xG}_i \\ & + \beta_{16} \mathrm{AerW}_i + \beta_{17} \mathrm{TckW}_i + \beta_{18} \mathrm{Int}_i + \beta_{19} \mathrm{GS}_i + \beta_{20} \mathrm{GC}_i \\ & + \beta_{21} \mathrm{EurC}_i + \beta_{22} \mathrm{top5}_i + \beta_{23} \mathrm{M}_i + \beta_{24} \mathrm{A}_i + \epsilon_i\end{aligned}$$

Appendix B

Reduced model from LASSO

The equation representing the reduced model resulting from the LASSO regression has the following form

$$
\begin{aligned}
\log(\mathrm{MVA}_i) = \beta_0 &+ \beta_1 \log(\mathrm{MVB}_i) + \beta_2 \mathrm{Age}_i + \beta_3 \mathrm{Age^2}_i + \beta_4 \mathrm{Hght}_i \\
&+ \beta_5 \mathrm{Mins}_i + \beta_6 \mathrm{Gls}_i + \beta_7 \mathrm{Ast}_i + \beta_8 \mathrm{YC}_i + \beta_9 \mathrm{RC}_i \\
&+ \beta_{10} \mathrm{Fls}_i + \beta_{11} \mathrm{ScPass}_i + \beta_{12} \mathrm{ScDrib}_i + \beta_{13} \mathrm{xG}_i \\
&+ \beta_{14} \mathrm{AerW}_i + \beta_{15} \mathrm{TckW}_i + \beta_{16} \mathrm{Int}_i + \beta_{17} \mathrm{GS}_i + \beta_{18} \mathrm{GC}_i \\
&+ \beta_{19} \mathrm{EurC}_i + \beta_{20} \mathrm{top5}_i + \beta_{21} \mathrm{M}_i + \beta_{22} \mathrm{A}_i + \epsilon_i.
\end{aligned}
$$

This model is executed multiple times to test for differences between various subsamples. This involves the creation of models addressing the main research question of this thesis, i.e. comparison between the top 5 leagues and the rest, as well as individual positions on the pitch, and the analysis of discrepancies within the top 5 leagues. Each variation will stem from the same foundational equation, only with the exclusion of the dummy variable delineating the respective subgroup under investigation. Thus, in the distinctions concerning league affiliation, the "top5" dummy is omitted, while in the analysis of the individual positions, dummies M and A are excluded. Therefrom, we will receive two regressions distinguishing players from the top 5 leagues and the rest, three regressions for playing positions, and finally five regressions for individual elite leagues.

Table B.1: Reduced Model 3 results

	Dependent variable: log(MVA)
Intercept	7.757***
	(0.617)
log(MVB)	0.610***
	(0.016)
Age	−0.178***
	(0.034)
Age^2	0.002**
	(0.001)
Hght	0.003*
	(0.002)
Mins	0.0003***
	(0.000)
Gls	0.025***
	(0.005)
Ast	0.010*
	(0.005)
YC	0.008
	(0.005)
RC	−0.038**
	(0.019)
Fls	−0.029
	(0.022)
ScPass	0.003***
	(0.001)
ScDrib	0.069***
	(0.019)
xG	0.280**
	(0.135)
AerW	0.017
	(0.011)
TckW	−0.013
	(0.026)
Int	0.036
	(0.024)
GS	0.007***
	(0.001)
GC	−0.009***
	(0.001)
EurC	0.076***
	(0.028)
top5	0.610***
	(0.035)
M	0.074***
	(0.027)
A	0.034
	(0.040)
Observations	1952
R^2	0.927
Adjusted R^2	0.9262
Residual Std. Error	0.427 (df = 1929)
F Statistic	1114*** (df = 22; 1929)

Note: *p<0.1; **p<0.05***p<0.01
Robust SE in parentheses.