CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Transfers and success of football players

Bachelor's Thesis

Author of the Thesis: Kryštof Zatloukal

Study programme: Economic studies

Supervisor: Mgr. Petr Polák, M.Sc., Ph.D.

Year of the defence: 2023

Declaration

1. I hereby declare that I have compiled this thesis using the listed literature and resources only.
2. I hereby declare that my thesis has not been used to gain any other academic title.
3. I fully agree to my work being used for study and scientific purposes.

In Prague on May 3, 2023

Kryštof Zatloukal

Abstract

This thesis aims to investigate how successful are players' transfers to the English Premier League given their performance, characteristics and the characteristics of the transfers. The thesis defines the transfer success as the change in estimated market value of player. Collected data includes information about 101 players transferring into the EPL in the 2020 summer transfer window and is then analysed using OLS regression. The result of the regression showed that the better a player is performing before the transfer, the more successful his transfer is going to be. Furthermore, performance of the club, from which a player is transferring, has also shown to be a determinant of transfer success, with higher ranked clubs on the Kickalgor.com club ranking providing more successful transfers. Besides a player's individual performance and the quality of his previous club, his market value and age at the time of the transfer also play a significant role, with younger players with a lower estimated market value having more successful transfers. On the other hand, player's country of origin showed no significant effect, suggesting that nationality of a player isn't a determinant of successful transfer in the English Premier League.

Keywords	Football, Market value, Transfers, English Premier League, OLS
Title	Transfer and success of football players

Author's e-mail	83389552@fsv.cuni.cz
Supervisor's e-mail	polakpet@gmail.com

Abstrakt

Cílem této práce je zjistit, jak úspěšné jsou přestupy hráčů do anglické Premier League vzhledem k jejich výkonnosti, charakteristikám a vlastnostem přestupů. Práce definuje úspěšnost přestupu jako změnu odhadované tržní hodnoty hráče. Shromážděná data zahrnují informace o 101 hráčích přestupujících do anglické Premier League v letním přestupovém období 2020 a následně jsou analyzována pomocí OLS regrese. Výsledek regrese ukázal, že čím vyšší je výkonnost hráče před přestupem, tím úspěšnější bude jeho přestup. Kromě toho se ukázalo, že výkonnost klubu, z něhož hráč přestupuje, je také určujícím faktorem úspěšnosti přestupu, přičemž kluby s vyšším postavením na žebříčku klubů Kickalgor.com zajišťují úspěšnější přestupy. Kromě individuální výkonnosti hráče a kvality jeho předchozího klubu hraje významnou roli také jeho tržní hodnota a věk v době přestupu, přičemž úspěšnější přestupy mají mladší hráči s nižší odhadovanou tržní hodnotou. Na druhou stranu země původu hráče nevykazovala žádný významný vliv, což naznačuje, že státní příslušnost hráče není určujícím faktorem úspěšného přestupu v anglické Premier League.

Klíčová slova	Fotbal, Tržní hodnota, Market value, Přestupy, Anglická Premier League, OLS
Název práce	Přestupy a úspěchy fotbalových hráčů

E-mail autora	83389552@fsv.cuni.cz
E-mail vedoucího práce	polakpet@gmail.com

Acknowledgement

I would like to express my gratitude to Mgr. Petr Polák, MSc. Ph.D, who agreed to supervise this thesis and provided me with much useful insights and advice. I would also like to thank my girlfriend Sharon for her unconditional support and help along the whole process.

Bibliographic Record

Zatloukal, Kryštof: *Transfers and success of football players*. Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2023, pages 43. Supervisor: Mgr. Petr Polák, MSc. Ph.D.

Table of Contents

List of tables

List of figures

List of equations

1. Introduction

In 2022, 3.4 million people attended the Football World Cup in Qatar in person. Additionally, 1.5 billion spectators all over the world tuned in to watch the final match remotely, as reported by (FIFA, 2023) This is only one example of the size and reach of the sport of football (meaning the sport of soccer and not the American football), which belongs to one of the most popular and played sports all around the world. Football also stands out commercially, with the bigger leagues generating billions of dollars in revenue, becoming an attractive investment for investors all over the world - Middle East, China, etc. (Huth, 2020). One league in particular stands out.

The English Premier League (EPL) is the top tier of professional football in England and one of the most watched, competitive and sought-after leagues by not only fans and spectators, but also players in Europe and in the world (Barros & Leach, 2006). It is also very attractive to investors and sponsors, as it generates more revenue, than any other football league in the world, as reported in annual Deloitte Football Money League 2023. In the EPL, there are 20 teams that compete throughout 38 rounds, with each team playing every other team twice in one season (Premier League, 2022). Being one of the most popular and competitive football leagues in the world, there are big opportunities for significant financial investments from clubs and sponsors. Twice every season, during the winter and summer transfer windows, millions of GBP are spent in the football transfer market bringing in new players, in hopes of improving the team and with that the team's chances of success, but also at the same time millions of GBP are earned by selling players (Burdekin & Franklin, 2015).

Considering this, transfers to the EPL are sought after by football clubs all around the world, looking to sell their top talents to a league, which pays among the highest transfer fees in the world (up to over 100 million GPB - Enzo Fernandez to Chelsea in 2022, according to transfermarkt.co.uk) with the goal of generating profit. It is, furthermore, considered to be one of the most prestigious moves for any football player to transfer and play in the English Premier League, as through that move they can acquire not only fame and prestige, but also high salaries, with the best players earning over 20 million GBP per year (Quansah et al., 2021).

Selling and buying players has become an essential part of modern European football, which has been commercialized in the 20th century (Morrow, 1999), not only because of the contribution of the transfers to the sport itself, such as increased attractiveness and quality of the game because of the participation of high-profile players. Additionally, the transfers also play a major role in the total revenue of the clubs: For example, Chelsea football club generated 568.3 million EUR in revenue in the 2021/2022 season (*Deloitte Football Money League 2023*, 2023). The following season 2022/2023, Chelsea FC spent 611.5 million EUR on signing new players for the club (this amount includes only reported transfer fee) and sold players for 67.8 million EUR, according to transfermarkt.co.uk. In this case, the club spent almost its entire revenue generated in the previous season or signing new players. This shows, just how big a role transfers play in the world of football and how big a business the football player transfer market actually is.

It is therefore crucial for the clubs to have a clear understanding of a football player's value that he can bring to the club. Hence this will determine how likely it is that the transfer will be successful, resulting in positive return on the investment, whether they plan that short term or long term. Additionally, with the EPL being one of the most competitive and high-profile football leagues in the world, the risk of unsuccessful transfers creates a risk of potentially costing owners or investors of the club tens, sometimes even hundreds of millions of GBP.

The thesis builds on findings from previous studies such as (He et al., 2015), (Majewski, 2016) or (Newman, 2016). These studies are focused on topics, such as researching the future transfer values of certain players, key determinants in the transfer decision making, prediction of football team's success, the relationship between performance of a player and his market value or transfer fee. The aim of this thesis is to identify and then also estimate the determinants of a successful transfer to the EPL and the likelihood of a player's success based on his performance and characteristics.

This work focuses on the question, whether there is a relationship between player's demographic characteristics (e.g., age or nationality), his market value and his performance (e.g. goals scored or yellow cards) prior to his transfer, the characteristics of the transfer (e.g. the club and league, where the player was sold from or transfer fee) and between the success of the transfer. Collected and analysed data will include information about the transfer, such as the price, for which the player was sold, club and league, where the player was sold from and club and league, the player was sold to. It will also include

information about the player's performance before and after the transfer, such as number of scored goals/assists/clean sheets (based on the position of the player) and about performance of the clubs he represented before and after the transfer.

The thesis' structured is as follows: Firstly, relevant literature is reviewed and summarized, to provide background and groundwork for the topics of this work, such as football transfers, valuation of players or quantification of their success. Then the data, which is analysed, is described, followed by the description of the data-collection process, including data sources or data limitation. Further on, methodology used is described, to guide the reader through the creation and testing of the created linear regression models. After all the necessary tests are conducted and the final model is retrieved, results of this regression are interpreted and discussed in regard to the topic of this thesis. Limitations of the model and suggested direction of future research are included in the discussion part.

2. Literature review

As the topic of predicting a player's transfer success is open to interpretation, numerous studies can be leveraged and used as groundwork for this thesis. For example, studies looking into the relationship between a player's performance, such as goals scored, assists provided or pass completion rate and his market worth can provide useful insights into the creation of a model.

2.1 Transfer success, market value and performance in football

Although transfer success in football is not a widely researched topic, there have been attempts at defining the term and using it in optimizing football team building strategies.

Ćwiklinski et al. (2021) proposed three definitions of transfer success and implemented machine learning, trying to predict a player's successful integration into the team after his arrival at the club. One of the mentioned definitions of a successful transfer is a transfer, which affects high results in all of the player's aspects. Transferring player's performance was assessed at the end of the season and compared against a pre-defined threshold. If the assessment score was higher than the threshold, the transfer was considered to be successful. While this thesis differs in methods used, this definition of success provides a useful base, which will be utilized for defining the transfer success in this research.

Despite market value of a player not being the main researched variable, due to definition of transfer success used in this thesis[1], it is a relevant term directly related to the topic of this work. Previous research into determinants of market value of a player will therefore be useful in identifying the factor influencing the success of a player's transfer.

Considering the game mechanics and main objectives (scoring as many goals as possible, conceding as little goals as possible, avoiding breaking the rules and ultimately, winning the games) it doesn't come as a surprise that in four Europe's top leagues there is a positive correlation between market value and goals, assists and pass completion rate as well as a negative correlation between market value and received yellow and red cards, as found by He et al., (2015).

[1] transfer success will be defined and described in further chapters.

Furthermore, when looking at the top 100 football most valuable players, Majewski, 2016 suggests that the key factors influencing a player‘s market value, besides his game-related specifics, such as position played and his performance, indicated by goals scored, are also his demographic characteristics like age and nationality, indicating that a market value isn’t solely determined by a player’s on-field performance.

Assuming a strong positive correlation between a market value of a player’s and the transfer fee, for which that player is sold, as an approach to player's market valuation is an estimate of the amount for which a player can be sold (Herm et al., 2014), not all studies fully support Majewski’s claim. (Newman, 2016) found that the performance metrics, such as goals scored, assists and defensive playmaking were much better determinants of transfer fees than other factors like position or age.

As the transfer fee plays a detrimental role in not only valuating a player (omitting other factors influencing the transfer fee, such as negotiating power of the club, negotiating skills, etc., logically, the higher the transfer fee, the higher the player is valuated) but also determining the success of a transfer (e.g. from a player bought for 50 million EUR, a higher level of performance might be expected, than that of a player purchased for 500 thousand EUR by the same club), a relationship between transfer fee of a player and his performance provides additional assistance in creating a model for this thesis.

There is also evidence of the correlation between performance and success on team level. As concluded by (Gebbett, 2018) in the "Predictive Analysis of Success in the English Premier League", measures such as the goal differential (the amount by which the total number of goals a football team scores is greater than the total number of goals scored against it, used as a way of deciding the position of teams with the same number of points, Cambridge dictionary), ratio of total shots and shots on target and amount of significant opportunities created by a side were discovered to be the most significant predictors of a team’s success, with the probability that a team would place in the top six being over 75%.

Based on these papers, key determinants of a player’s quality (being it his market value or predicted transfer fee) can be categorized into following groups:

- performance determinants – in-game statistics or characteristics, which are direct reflection of a player’s on-field performance. These include, but are not limited to played position, goals scored, assists provided, yellow and red cards received, etc.

- demographic determinants – demographic and constant characteristics of a player, such as his age or nationality.

There were also other factors, which have been found to influence a player's value, including the physical attributes, such as speed, strength and endurance, as well as their technical skills, tactical awareness and mental toughness. Despite these attributes being harder to observe and quantify, there have been attempts to do so using data from the FIFA video game (Al-Asadi & Tasdemir, 2022).

In summary, a range of factors, which determine a player's market worth and can potentially also influence his transfer success has been identified by previous research. Some of these determinants are highly complex (such as physical attributes or market value) and connected with each other, therefore proper understanding of them, before conducting any analysis, is necessary in order to make correct assumptions and achieve conclusive outcomes. Those will hopefully bring useful insights, making predictions of transfer successes more accurate and thereby optimizing transfer strategies for football clubs.

2.2 Research methods

With the goal of investigating the effect of player's performance, his characteristics and the characteristics of the transfer on the success of his transfer, it is important to choose the correct method that will yield the most relevant and reliable results. And as past publications have shown, choosing the OLS method is a common approach, given nature of the data.

Preston (2012) used OLS to study the determinants of football transfer market value. In his work, he arrived at the conclusion that players plaid position has no significance in determining his market value, opposing findings of Majewski, (2016). His research not only suggests direction, which this thesis follows, focusing not only on a player's performance, but also on other characteristics and indices. Additionally, it provides a guideline for optimal methods used in investigating determinants of transfer success. Furthermore, the regression in Majewski's thesis yielded relevant results, despite the R squared obtained being below 0.5, as in football, there is a lot of factors influencing a player's market value, which are hard to observe and quantify.

2.3 Sources of information

As there is no central database of players or one official source of information about their performance, market values or transfers, researchers studying football related topics depend on alternative sources to collect relevant data, which is then analysed and used the outcomes are then utilized by members of the football community across the whole world.

One of the most commonly used sources is transfermarkt.co.uk. This website provides its users with information about teams' standings in professional leagues worldwide and about players' statistics, such as goals scored or number of appearances for a given club. It also contains the estimated market value of all professional football players, which is updated throughout the season based on a player's performance. The estimated market value is commonly used as a valuation tool, to quantify a player's worth on the transfer market. Bhilawa & Fahriansyah (2022) used market values from transfermarkt.co.uk to study the effect of performance, age, transfer fees and salaries on a player's market value. Similar study was conducted by Margareta & Malinda in 2022 using the same source of data for the value of footballers. While not being an official report of a player's worth and therefore not being 100% accurate, market values from transfermarkt.co.uk provide a reliable proximation on a football player's value.

Information about clubs' and leagues' performance is also provided by Kickalgor.com. The website also includes ranking of professional football clubs and leagues, which is then used to quantify these variables. Saravia (2022) used Kickalgor.com, when researching aggressive play in the sport of football, to quantify the quality of the league in question.

When looking at players' performance, it is difficult to accurately quantify it, considering how many observations and statistics can be measured during a football game. Ratings from Whoscored.com represents an accurate source of a players' performance ratings, utilizing over 200 individual statistics in their algorithms to award after-game rating on the scale from one to ten. Whoscored.com rating has been used in numerous studies (Carlos et al., 2019; Niederer et al., 2018; Phatak & Gruber, 2019), when observing and evaluating in-game performance.

3. Data and Methodology

3.1 Data collection

As there is no single, official and centralized source of statistics for all football players of all leagues or an official database of all players, which would have all the necessary data point for this thesis, therefore several sources are used and combined together.

Fifa.com: Official website of the international football association. In general, this website contains information about the world of football including news or current/upcoming events. It also includes the official ranking of both women's and men's national teams[2], which is updated 6 times a year. This thesis used the Fifa.com website to collect the ranking of the men's national teams in order to determine players' nationality.

Transfermarkt.co.uk: Football database, which had become one of the most used and cited databases in the world of football. Being updated daily, this website includes information about virtually every professional player, which is being leveraged in all spheres of this sport. In this thesis, transfermarkt.co.uk was used for collecting the information about the transfers:

- transfer window, in which the player transferred,
- transfer fee, for which the player was sold,
- estimated market value of the transferred player at the time of the transfer and at the end of the season, in which he arrived at the new club,
- his name,
- his age at the time of transfer,
- played position,
- club and league from which he transferred,
- club and league to which he transferred,
- his nationality,
- his performance in the season before the transfer.

[2] for detailed information about the algorithms used, please visit the FIFA website.

Kickalgor.com: Football website collecting and analysing data of football games in all of world's major leagues including all teams on a national and international level. Kickalgor.com uses complex algorithms to rate leagues, clubs and national teams, providing a proximation twice for every season, by which these leagues and teams can be ranked and quantified[3]. In this thesis, Kickalgor.com was used to collect data on the ranking of football clubs and national leagues. As the quantification of football leagues and clubs doesn't have one singular approach and as there aren't any official ranking of leagues and clubs, data collected from Kickalgor.com represents a one of many possible approaches to quantifying these variables. Additionally, as this source has been used in literature before (Saravia, 2022), it can be considered reliable.

Whoscored.com: Whoscored.com is a website dedicated to football analytics. Analysing data from both live and past football matches from the biggest leagues in the world, Whoscored.com uses its own algorithms, compromising of over 200 raw statistics, to provide its users with a player' ratings and characteristics[4]. For this thesis, average seasonal rating of transferred player before and after the transfer was collected, to provide a more sophisticated and quantifiable measure for a player's performance. Despite not being an official source of a football player's performance ratings, Whoscored.com represents a reliable source of a player's rating, being commonly used in studies across the field (Carlos et al., 2019; Niederer et al., 2018; Phatak & Gruber, 2019).

At first, all available transfers from one transfer window were manually collected from transfermarkt.co.uk.

For this thesis, transfers from the summer transfer window of the 2020/2021 were collected, in order to assure availability of the data before and after the transfer, while keeping the data relevant, as football is a highly dynamic and ever-changing environment. At the same time, comparing data from the seasons 2019/2020 and 2020/2021 at least partially cancels out the effect of the COVID-19 pandemic, which was affecting the whole world at the time, as both seasons were affected by it in some shape or form (Quansah et al., 2021).

[3] for detailed information about the algorithms used, please visit the Kickalgor.com website.
[4] for detailed information about the algorithms used, please visit the Whoscored.com website.

Following the collection of all transfers, only transfers from the EPL were filtered out, as those were the observations, which this thesis intends to study. The rest of the original data set could potentially be used in the future for replicating this study in different leagues.

After cleaning the data of missing or faulty values, a total of 101 players transferring to the EPL was collected, which will be used in the analysis. These transfers were then deep dived into, adding information about the performance of these players in the season 2019/2020 (season before the transfer).

All of these statistics were collected from transfermarkt.co.uk. Apart from individual performances of the players, information about the clubs (final placing in the EPL of all teams, participating in the 2020/2021), the players transferred to, was collected from this database.

After the collection of the data from transfermarkt.co.uk was completed, quantification of parameters such as nationality, left club (player's club from which he transferred), left league (the league, in which the player played the previous season) and joined club (club to which the player transferred) from Fifa.com and Kickalgor.com was necessary, in order for these variables to be measurable.

The last used source of data was Whoscored.com, which provided player rating based on detailed analysis of 200 statistics, providing further information about a player's performance.

3.2 Data Description

Prior to description of the collected data set, which will be used in the model, it is important to define a few key terms, in order to understand the logic behind the model, which will be created and interpreted and to ensure understandability and interpretability of this model.

First and probably the most important term is "transfer success". Success is a relatively abstract term on its own. The Cambridge dictionary defines success as "*something that achieves positive results",* which again could be interpreted in many ways on its own. For the sake of this thesis, the term success will be defined as "*being better off than before".* Similarly to Ćwiklinski et al. (2021), who used transferring player's assessment at the end of the season and compared it to a pre-defined threshold, transferring player's market value at the end of the season will be compared to his market value at the beginning of the season. More specifically, the transfer success will be

defined as the absolute change in the estimated market value of the player in millions EUR between the time of his arrival and the end of his first season at the new club. Negative transfer success will indicate unsuccessful transfers, while positive transfer success will imply successful transfers, as theoretically increase in market value is a favourable outcome for any club or player after a transfer. The absolute value in millions EUR will demonstrate the magnitude of the transfer success/failure.

It is important to keep in mind that in this case, observed success will be of short-term nature (change in estimated market value after one season) because of data availability and standardization (if more seasons would be included, more outside factors would come in, which couldn't be effectively accounted for) reasons, which does bring with it certain limitations. For example, the momentum of a player's career, which may be reflected by a longer adjustment period than one season in the new club, before he starts successfully performing (e.g., Aaron Ramsdale has been valued at 12 million EUR at the time of the transfer and at the end of first season his value stayed at the same level. At the end of the 21/22 season however, his estimated value increased to 28 million EUR and keeps increasing, as reported by transfermarkt.co.uk). A detailed summary of model limitations will be included in the following chapters.

Subsequently, an important term, which also needs to be properly defined, is the market value of a player. Market value is defined by the Cambridge dictionary as "*the price at which something can be sold, at a particular time in a particular place"*. In this thesis, the price at which something can be sold will be estimated by the market value provided by an unofficial source, transfermarkt.co.uk. Despite not being an official database, transfermarkt.co.uk has become a reliant benchmark for a player's valuations, being used not only in numerous published papers and studies (Locks et al., 2018; Scholz, 2011; Wand, 2022), but also in annual reports of top European clubs, at transfer negotiations by scouts and by football club officials (Keppel, 2020), making it a relevant source on this topic and these observations.

Despite previously mentioned websites not being sources directly affiliated with the international football association, they represent a reliable proximation, by which a player's performance or a club/league's performance can be estimated. All the aforementioned sources have been used in literature and are accepted as viable sources by the community, representing data with acceptable accuracy and reliability. Collected data from transfermarkt.co.uk included following statistics for the season, before the transfer happened:

- Market value at the time of the transfer (MV 19/20): estimated market value of the player in millions of EUR at the time of the transfer.
- Market value at the end of 2020/2021 season (MV 20/21): estimated market value of the player in millions of EUR at the end of the first season after the transfer.
- Market value change (Outcome MV): change in market value in millions of EUR over the period of the 2020/2021 season.
- Transfer success (successMV): success measured as the absolute change in market value in millions of EUR ($successMV = Outcome\ MV$).
- Transfer fee (fee): reported fee, for which a player was sold and transferred to a club in the EPL.
- Position played (position): in-game position played of the player, differencing between goalkeepers (GK), defenders (DEF), midfielders (MID) and forwards (ATT).
- Appearances (APP): number of times, when played actively participated in a football match (either as part of the starting eleven or as a substitute).
- Goals (G): number of goals scored for his team.
- Assists (A): number of assists provided for his teammates, which directly led to a goal.
- PPG: average number of points a player's team gained (3 points for winning a match, 1 point for drawing a match and 0 points for losing a match), when the player actively participated in the match (either as part of the starting eleven or as a substitute).
- Yellow cards (YC): number of yellow cards received.
- Second yellow cards (2YC): number of consecutive second yellow cards received, leading to receiving a red card and to suspension.
- Red cards (RC): number of red cards received directly (without a yellow card before).
- Number of suspensions (SUSP): number of times a player has received either a second yellow card (2YC) or a red card (RC), leading to his suspension from current and following match (calculated as: $2YC + RC$).
- Clean sheets (CS): collected only for goalkeepers for data availability reasons, number of matches a goalkeeper didn't receive a goal.

- Clean sheets + Goals (CSG): metric designed to standardize performance of players across all positions while minimizing the difference between performances of on-field players and goalkeepers (calculated as: $CS + G$).
- Minutes played (MP): number of minutes spent actively participating in the game.
- Minutes per appearance (MPA): on average, how long did a player spend on the field (calculated as: $\frac{MP}{APP}$).
- 19/20 team finish (20TF): placement of the team, to which the player transferred at the end of the 2019/2020 season (i.e., placement without the participation of the player).
- Joined club (JC): quantified variable of the team, which a player joined, in the form of the team's final placement in the 2019/2020 EPL season, currently providing the best estimate of team's rank and form ($JC = 20TF$).
- EPL last season (EngL): a dummy variable representing whether a player played in the Premier League the previous season.

Demographic information about the player, such as the age at the time of the transfer and nationality was also collected from transfermarkt.co.uk.

Following data was collected from FIFA, The Kick Algorithms and Whoscored.Com:

- Left club (LC): team, from which the player transferred to the EPL – to quantify this variable, ranking of football clubs from the June 30th, 2020 (rating the 2019/2020 season), released by Kickalgor.com, was used. Teams were assigned a value between 1 and 897, with each assigned number corresponding to a team's rank and with 1 being assigned to the best ranked team. The ranking included 895 teams at the time and so two teams not included in the rating were assigned the rank 896 and 897, as that was the current best estimate of their position and as this assumption didn't pose any significant risk to relevancy of this variable.
- Left league (LL): league, from which the player transferred to the EPL – to quantify this variable, ranking of football leagues from the June 15th, 2020 (rating the 2019/2020 season), released by Kickalgor.com, was used. Every league was assigned a value between 1 and 76, with each assigned number corresponding to a league's rank and with 1 being assigned to the best ranked team.

- Whoscored rating (WSrating): rating of a player's performance, evaluated based on over 200 statistics by sports analytics of Whoscored.com. A player is rated from 1 to 10, with a 10 representing a perfect performance (for more details about the ranking system and algorithms, please visit the Whoscored.com website). This rating includes data from only top 11 biggest leagues in the world and European international leagues, indirectly providing additional variable of a player's appearance in one of the top leagues, further implicating his value. Players without a WSrating have been valued at 0.
- National team appearance (NatT): a dummy variable representing whether a player has appeared in a game for the national team.

In order to quantify a player's nationality, official ranking of Men's national football teams from July 16th, 2020 (around the time of analysed transfers) released by Federation Internationale de Football Association (FIFA) was used. A correlation between a player's football ability and ability of nation's selected representatives is assumed. Each country was assigned a value between 1 and 124, with each assigned number corresponding to a nation's rank in said release from FIFA, with 1 being assigned to the best ranked team.

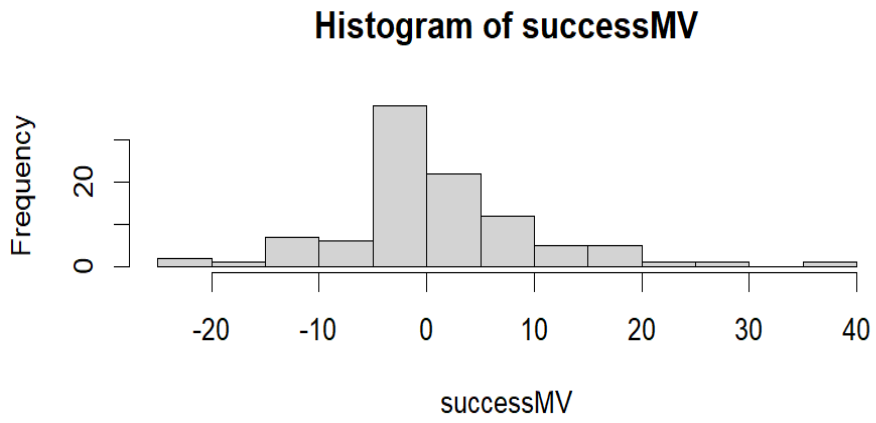
Before concluding on data description, with this thesis being concerned with identifying factors influencing the success of a player's transfer, it is critical to fully define and describe the independent variable. As defined above, success of a transfer is determined based on the change in estimated market value by transfermarkt.co.uk after the completion of the 2019/2020 season and quantified in millions EUR. Looking at Table 3.1, it can be concluded that based on the mean, the average transfer to the EPL in the 2020 summer transfer window is successful with a 1.7069 million EUR increase in estimated market value.

Table 3.1: Summary of the independent variable

Statistic	Min	Median	Mean	Max	St. Dev.
successMV	-22.0000	0.0000	1.7069	40.0000	9.3564

Keeping the size of max. and min. values of the independent variable in mind, the value of the median suggests that half of the observations is below zero – and are therefore unsuccessful transfers, and the other half is above zero, indicating increase in market value and therefore a successful transfer. Similar trend can be observed in Figure 3.1.

Figure 3.1: Histogram of the independent variable



Furthermore, from the Histogram in Figure 3.1 it can also be concluded that the max. successful transfer, which has seen 40 million EUR increase in market value, is by far the biggest increase and can potentially be an influential outlier skewing the picture. Deeper analysis of this observation is to follow.

3.3 Data Limitation

Despite the data collection being very thorough, it should be considered that the data was collected manually with over 1500 transfers in total and over 100 of them being deep dived into, hence there may be a possibility of small discrepancies. It is also important to point out that although data collected is expected to be relevant for the intended analysis, the data set has its limitations and inconsistencies.

For quantitative purposes, all appearances of the player for both seasons were considered. This means that there are not only appearances in the EPL matches included, but also appearances in national and international Cups, such as the F.A. Cup, League Cup, UEFA Champions League, UEFA Europa League, etc. Herewith, teams face potentially different opponents, resulting in different game standards and therefore different opportunities for the players. Furthermore, this can lead to varying approaches of the teams (e.g. to some teams, it can be more important to win an EPL game, than a Champions League game and vice versa), again resulting in different strategies (e.g. starting worse performing players in less important matches or giving them more playing time) and expectations, which are reflected in a player's performance (in the example stated above, a worse performing player's performance could be overestimated, as he got more playing time and opportunities in less important matches, than he would without them included).

On the other hand, Whoscored.com rating is included only for players, which played in one of the top leagues in the 2019/2020 season, as other leagues are not available on the website. Players without a WS-rating are assigned 0. This variable therefore represents two indices for a player. Firstly, his average performance during the 2019/2020 season, which considers virtually all in-game actions of the player, resulting in a highly representative rating. Secondly, this variable provides information on whether a player played in one of the best leagues, which represents a player's overall quality, on and off the field, which allows him to perform at this top level.

In the dataset, for sample size reasons, there are permanent transfers used and there are loan transfer used, which are not completely comparable in both quantifiable aspects, such as transfer fee (loan transfer fee to market value ratio is much lower (0.079 on average) than permanent transfer fee to market value ratio (1.356 on average) and qualitative, more un-quantifiable aspects, such as commitment to the club or performance expectations from the club (players brought in on a permanent contract are arguably expected to perform at a much higher standard and are given more opportunities, than players loaned out.

It could be argued that even this distinction has exceptions, as some players are loaned out with the ultimate goal to be purchased by the club, they are loaned out to. This can also be reflected in cases, where loan transfers are feed (e.g., from the dataset: Craig Dawson being loaned out to West Ham in the 20/21 season and being bought by West Ham in the 21/22 season).

At the same time, some players were in the 2019/2020 season loaned out to teams, by which they were the following 2020/2021 season bought (e.g., from the dataset: Grady Diangana was loaned out to West Brom, who then bought him in the 20/21 season). This can potentially overestimate a player's success, as the rank of the joined or left team plays a smaller role, because the player is more likely to already be assimilated. Luckily, this is not the case for vast majority of observation from the data set and for most of transfers in football in general and shouldn't therefore influence the reliability of the model in any significant way.

3.4 Methodology

To analyse the data collected and to estimate the relationship between a player's transfer success (increased estimated market value) and between his performance and characteristics before the transfer, the ordinary least squares method (OLS) was used, as it is one of the commonly used methods in the field of sports science (used by Gerhards & Mutz, 2017 or Milenovský, 2015). As the data was collected from one time-period it can be treated as cross-sectional type of data, for which the OLS method is suitable.

The initial linear regression model, estimating the impact of the following variable on a player's transfer success, is as follows:

Equation 3.1: Initial linear regression model

$$successMV_{\mathrm{i}} = \beta_0 + \beta_1 age_i + \beta_2 MV_i + \beta_3 NatT_i + \beta_4 JC_i + \beta_5 EngL_i + \beta_6 fee_i + \beta_7 LC_i + \beta_8 LL_i + \beta_9 PPG_i + \beta_{10} MPA_i + \beta_{11} SUSP_i + \beta_{12} WSrating_i + \beta_{13} CSG_i + \beta_{14} YC_i + \epsilon_i$$

The model assumes the following coefficients of estimators of dependent variables, described in Table 3.2:

Table 3.2: Summary of independent variables

Statistic	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
age	18	22	24	25	28	36
MV (mil. EUR)	1.00	5.00	10.00	15.43	20.00	81.00
NatT (dummy)	0.0000	0.0000	1.0000	0.6832	1.0000	1.0000
JC (rank)	1.00	6.00	12.00	11.25	18.00	18.00
EngL (dummy)	0.000	0.000	0.000	0.198	0.000	1.000
fee (mil. EUR)	0.00	1.00	8.25	13.89	20.90	80.00
LC (rank)	1	18	101	178	254	897
LL (rank)	1.00	2.00	5.00	11.72	21.00	76.00
MPA^2 (min. sq.)	0.007	3969.000	6265.972	5491.556	7266.271	8100.000
SUSP (dummy)	0.0000	0.0000	0.0000	0.1683	0.0000	3.0000
WSrating (rating 0-10)	0.00	6.51	6.76	6.28	6.99	8.19
CSG	0.000	1.000	3.000	4.792	7.000	34.000
YC	0.00	1.00	4.00	4.04	6.00	16.00

Age: Because football is a highly physical and demanding sport, age plays a crucial role in a player's valuation. As shown by previous research, age is one of the main determinants of a player's performance and value. In this model, age is expected to have a negative coefficient in the model, as young players have generally higher potential and are therefore more sought after in the transfer market by the clubs.

MV: with the success of a transfer being determined based on the change in a player's market value (MV), MV is expected to be strongly correlated with the dependent variable. Despite it seeming counter intuitive, MV is expected to have a negative coefficient, as for players valued at a high market value it might be difficult to live up to the expectations and standards set by their exceptional performance at the previous club. Furthermore, players with a lower market value arguably have more potential for growth, as they are potentially not at their peak.

NatT: a dummy variable, representing whether a player played at least once for the national team of his country. Choosing only the best players from the country, national team appearances are an indicator of a player's qualities and are therefore expected to be correlated with a player's success with a positive coefficient.

JC: variable quantifying the team in the EPL, which a player is joining. This variable is expected to have a negative coefficient, as the better a team placed in last season's final ranking, the better the team is expected to perform again and therefore that creates an optimal environment for a player to come into after the transfer. Although it could be argued that players transferring into the top performing teams will not get as many opportunities and playing time, so other interpretations of the coefficient are possible.

EngL: another dummy variable representing a player's experience in the EPL in the previous season, which should theoretically put the player in an advantageous position, as he would already be used to the playing style and would therefore need less time to adjust. Therefore, it is expected to have a positive coefficient.

Fee: This variable is also expected to have a positive coefficient, as players bought for higher amounts are expected to perform at high standards to generate a positive ROI.

LC: A variable quantifying the team, which a player played for before the transfer. The assumption on this variable is that the better a team is ranked, the higher performance of the player transferring from it should be (otherwise, the player wouldn't be able to play for that highly ranked team). Therefore, as the ranking is from 1 being the best team, it is expected to have a negative coefficient.

LL: Applying the same logic as with the LC variable, the left league variable is expected to have a negative coefficient, as a player's quality and performance and therefore his potential success should be correlated with the quality and performance of the league, in which he played.

PPG: This variable measures the direct impact a player has, when in game. It is expected to have a positive coefficient, as with higher PPG a player's impact is higher and therefore this player is expected to bring this influence to the new club.

MPA^2: Variable representing physical attributes such as stamina - the more minutes per appearance, the better a player's stamina can be assumed to be, alongside with quality attributes - mostly being decided by the coach, such as effectiveness and impact - the more minutes per appearance, the more effective and more impact a player has. Therefore, MPA is expected to have a positive coefficient. Furthermore, as the majority of datapoints (last 3 quantiles) is within 63 MPA and 90 MPA, as shown in Table 3.3, this variable will be included as square.

Table 3.3: Summary of the MPA variable

Statistic	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
MPA	0.8564	63.0000	79.1579	71.6063	85.2424	90.0000

YC: Variable indicating negative traits of a player, such as inability to conform or follow set rules of the game, un-sportsmanship, or poor decision-making. Thus, it is predicted to have a negative coefficient, as these qualities theoretically should lower a player's valuation.

SUSP: Even stronger representation of negative traits of a player than YC, as it represents similar traits, but with more severe consequences of getting disqualified. It is also expected to have a negative coefficient for similar reasons as YC.

CSG: Variable representing a positive performance of players, which is standardized for all positions by adding goals (for field players) and clean sheets (for goalkeepers). It is expected to have a positive coefficient as well, as scoring goals and keeping clean sheets are desirable outcomes of a player's performance, potentially increasing his MV.

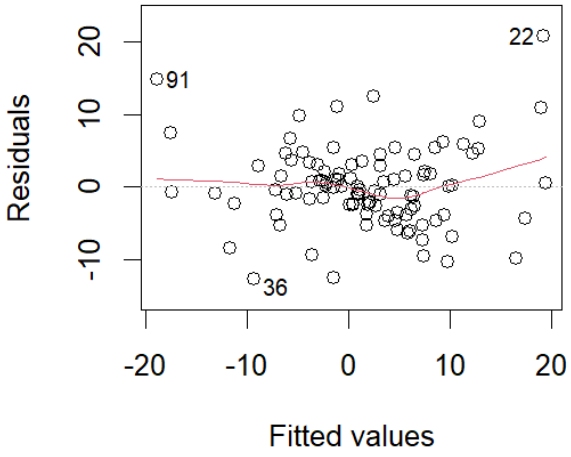
WSrating: In-depth rating of a player's performance in the previous season in one of the top leagues. It is expected to have a positive coefficient, as the better performance a player showcased the previous season, the more likely he is to bring that quality to the new team.

4. Results and discussion

Before diving into the interpretation of the results, it is necessary to test, whether the model fulfils all necessary assumptions of the OLS method, to produce unbiased and consistent estimates. These assumptions will be further on referenced to as MLR (Multiple Linear Regression) assumptions.

The first MLR assumption deals with the linearity of the relationship between the dependent variable and the independent variables. When looking at the Figure 4.1: Residuals vs Fitted values, it is visible that the residuals are roughly equally spread around the horizontal line without a clear pattern, which is a general indication of a linear relationship.

Figure 4.1: Residuals vs Fitted values



Second MLR assumption states that the sampling of the observations must be random. Considering the inclusion of all available transfers into the EPL from the given time-period from transfermarkt.co.uk, this assumption is also fulfilled.

No perfect collinearity is the condition, which the model needs to have to pass the third MLR assumption. This is tested using the Variance Inflation Factor (VIF). In general, VIF values between 5 and 10 or even larger than 10 indicate a high degree of correlation between the independent variables and potential multicollinearity issues. In the case of the initial model, all the VIF values are below 5, indicating that there is no major multicollinearity issue and the third MLR assumption is satisfied.

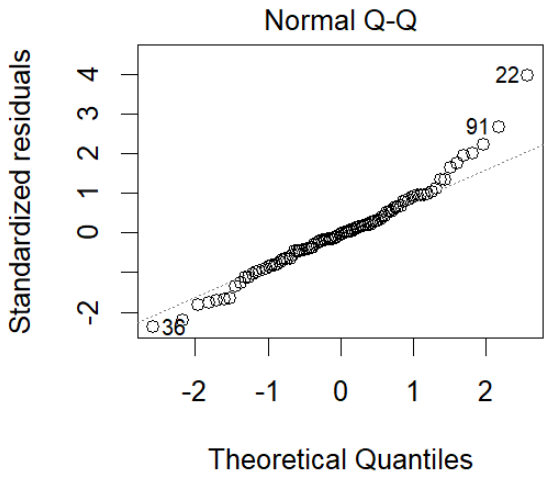
The fourth MLR assumption is concerned with zero conditional mean of *u* (disturbances) and as the intercept is included in the model, it is assumed that this assumption holds.

As MLR 1-MLR 4 hold, according to (Wooldridge, 2012), the OLS estimator is unbiassed and consistent.

Having an unbiassed and consistent estimator, it is important to satisfy MLR. 5 to obtain the best linear unbiased estimator (BLUE). For that, the model needs to satisfy the Homoskedasticity assumption. To test for constant variance, the studentized Breusch-Pagan Test was used. With the p-value of 0.051, the null hypothesis of present homoscedasticity is not rejected as the commonly used significance level of 0.05. To definitely conclude on the variance of the error, additional Non-Constant Variance Test was conducted confirming the results of the previous test at the significance level 0.05 with p-value of 0.09.

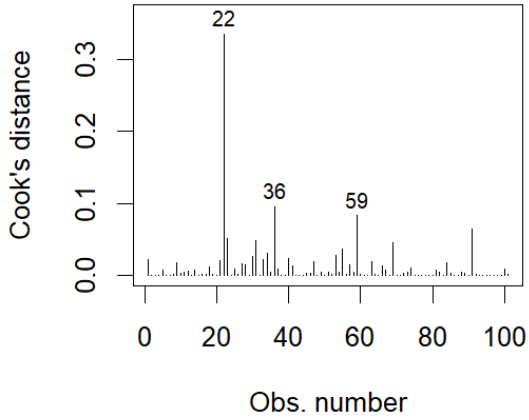
As it is important that the dataset doesn't include significantly influential outliers, which would skew the results, normal distribution of the error term is required under the sixth MLR assumption. When looking at Figure 4.2: Q-Q plot of standardized residuals, it is visible that the 22nd observation may be an influential outlier, distorting the picture.

Figure 4.2: Q-Q plot of standardized residuals



Furthermore, after conducting the Shapiro-Wilk Test, there is sufficient evidence to reject the null hypothesis that the data follows a normal distribution at the significance level 0.05, which can be partially attributed to the aforementioned outlier (confirmed, when looking at the Cook's distance in Figure 4.3: Cook's distance of observations from the), which needs to be examined.

Figure 4.3: Cook's distance of observations from the dataset



After analysing the data set, the 22nd row of data has been identified as the transfer of Rúben Dias, who transferred from Benfica to Manchester City on September 29th, 2020, for the amount of 71.6 million EUR, being valued at 35 million EUR at the time of the transfer, as reported by transfermakrt.co.uk. The reason for this transfer being identified as significantly influential is the magnitude of his "success". With the average success of a transfer being 1.71 million EUR and the standard deviation being 9.36 million EUR, as shows in Table 3.1, success of Rúben Dias cannot be considered representative.

This can be explained by several factors. One of them is the undervaluation of the player by the website at the time of the transfer, as only 14 days after the transfer, Dias has been valuated at 50 million EUR, 15 million EUR higher, than at the day of his transfer, as reported by transfermarkt.co.uk and shown in Figure 4.4.

Figure 4.4: Development of Rúben Dias's Marketvalue, transfermarkt.co.uk

Another factor causing this abnormality in the dataset is the exceptional performance of the player, which isn't captured by the included variables. Dias was only the fourth defending player and additionally only the second Portuguese player in the history of the English Premier League to win the Player of the Season award played. On top of his individual successes, he played a key role in the success of Manchester City in the 2020/2021 season, helping the team win the EPL title, EFL Cup and reach the UEFA Champions League Final (premierleague.com, 2021).

In general, the rule of thumb for Cook's distance is that its value greater than 4/n (n being the number of observations) is used to indicate highly influential points. Because that is the case with row 22 for reasons stated above and considering it is the most influential observation by far, as observed in Figure 4.3, it was excluded from the dataset, in order to ensure normal distribution of the dataset.

After conducting the Shapiro-Wilk Normality test with the adjusted dataset, there was a lack of evidence to reject the hypothesis that the data follows a normal distribution. All the necessary tests to check MLR. 1-5 with the adjusted dataset were again conducted to confirm that the model still fulfils all the needed assumptions.

Confirming that the OLS estimator fulfils all six MLR assumptions and is therefore the best unbiased estimator (BUE), the significance of each independent variable needs to be tested to assure that the model is the best possible fit for this particular dataset. Following a top-line assessment of the initial model, it became apparent that the *LL* variable wasn't significant (with the highest p-value of over 0.87), nor was it contributing to explaining the independent variable. This was confirmed with the increased adjusted R squared after the exclusion of the *LL* variable.

The new model, which became the full linear regression model, from which the reduced models were created to find the best fit for the data, was as follows:

Equation 4.1: Full linear regression model

$$successMV_{\mathrm{i}} = \beta_0 + \beta_1 age_i + \beta_2 MV_i + \beta_3 NatT_i + \beta_4 JC_i + \beta_5 EngL_i + \beta_6 fee_i + \beta_7 LC_i + \beta_8 PPG_i + \beta_9 MPA_i^2 + \beta_{10} SUSP_i + \beta_{11} WSrating_i + \beta_{12} CSG_i + \beta_{13} YC_i + \epsilon_i$$

As shown in Table 4.1, variables *age, MV, fee* and *LC* in the full model are statistically significant at the significance level 0.01, *WSrating* and *CSG* being statistically significant at a significance level 0.05 and *MPA* having statistical significance at the level o.1. Therefore, these variables contribute to explaining the transfer success of a player. On the other hand, *NatT* (p-value 0.30)*, EngL* (p-value 0.17)*, PPG* (p-value 0.76)*, SUSP* (p-value 0.85) and *YC* (p-value 0.63) are less likely to be contributing to explaining the dependent variable and the model excluding these variables is considered to be as the better fit for the data.

For finding the model with the best fit, Akaike Information criterion (AIC) was used, as it represents a commonly used test for model fit. Looking not only on the decrease or increase of the AIC values, but also on the change in adjusted R squared, *JC*, *PPG*, *SUSP* and *YC* have not shown to be contributing to the explanation of the independent variable in any significant way, as the exclusion of these variable decreased the AIC value and increased the adjusted R squared, which are both indicators for the goodness of fit of the model (Wooldridge, 2012). On the other hand, *NatT* and *EngL* variables have shown to be significant in explaining the independent variable, as their exclusion decreased the adjusted R squared and/or increased the AIC value. Therefore, *JC*, *PPG*, *SUSP* and *YC* have been excluded, *NatT* and *EngL* have been kept and the final reduced linear regression model is the following:

Equation 4.2: Final reduced linear regression model

$$successMV_{\mathrm{i}} = \beta_0 + \beta_1 age_i + \beta_2 MV_i + \beta_3 NatT_i + \beta_4 EngL_i + \beta_5 fee_i + \beta_6 LC_i + \beta_7 MPA_i^2 + +\beta_8 WSrating_i + \beta_9 CSG_i + \epsilon_i$$

Table 4.1: Summary and comparison of full linear regression model and reduced linear regression model

	Dependent variable: successMV	
	Full model	Reduced model
age	-0.8054*** (0.1717)	-0.7988*** (0.1647)
MV (market value)	-0.6494*** (0.0676)	-0.6574*** (0.0644)
NatT (appearance for national team)	1.3924 (1.3469)	1.4580 (1.3041)
JC (ranking of joined club)	0.1085 (0.1197)	
EngL (appearances in previous season Of the EPL)	2.0943 (1.4962)	2.0131 (1.4323)
fee (transfer fee paid)	0.3484*** (0.0600)	0.3416*** (0.0582)
LC (ranking of the left club)	-0.0120*** (0.0032)	-0.0111*** (0.0030)
PPG (average points per game earned)	-0.2069 (0.6621)	
MPA^2 (minutes played per appearance)	0.0005* (0.0003)	0.0005** (0.0003)
SUSP (number of suspensions)	-0.2285 (1.2357)	
WSrating (Whoscored rating 0-10)	0.7384** (0.3236)	0.8248*** (0.3045)
CSG (number of goals and clean sheets)	0.2625** (0.1004)	0.2587*** (0.0965)
YC (number of yellow cards)	0.0859 (0.1760)	
Constant	17.6056*** (4.9536)	17.9768*** (4.4656)
Observations	100	100
R2	0.6565	0.6516
Adjusted R2	0.6046	0.6168
Residual Std. Error	5.3898 (df = 86)	5.3060 (df=90)
F Statistic	12.6435*** (df = 13; 86)	18.7048*** (df = 9; 90)

Note: *p<0.1; **p<0.05; ***p<0.01

Following the exclusion of five variables, the reduced model still fulfils all of the MLR assumptions, making it the best unbiased model and the best fit for the data. Furthermore, when looking at the decreased p-value of the MPA^2, *WSrating* and *age* variables in the reduced model, they are now even more significant explanatory variables, with MPA^2 being now significant at the 0.05 level and the *WSrating* and *age* being significant at the 0.01 level.

As shown in table Table 4.1, a total of 100 observation was used and the results can therefore be considered representative, especially when the distribution of the data is relatively close to normal distribution, as presented in the Methodology section.

When looking at the R squared and the adjusted R squared, over 61% of the variance in the dependent variable is explained by the model. This is considered to be a moderately high value in the field of sports science, representing a good fit, as values of this magnitude or lower are commonly used to interpret trends in football (Majewski, 2016; Preston, 2012).

Furthermore, shown F-statistic of 18.70 on 9 and 90 degrees of freedom and the associated p-value (< 0.01) indicate that together the independent variables are explaining the dependent variable with statistical significance.

Overall, results presented in Table 4.1 suggest that the reduced model provides a good fit to the data with some variance of a player's success remaining unexplained by it. When looking at the signs of the estimated coefficients, there are variables confirming the initial assumptions and variables, which show different signs, than expected.

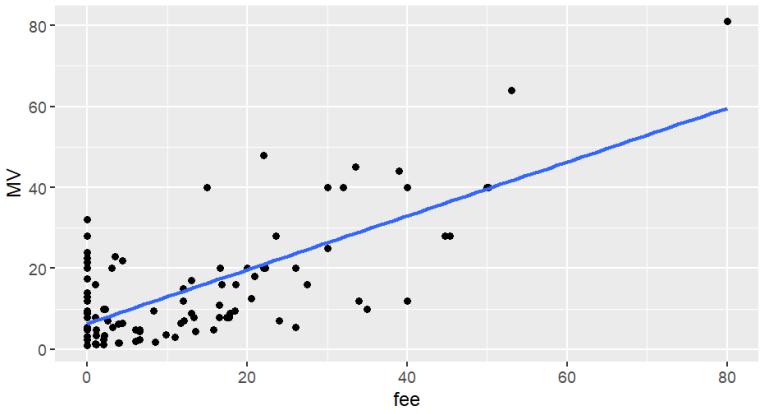
The β_1 coefficient, estimating the effect of a player's age on the success of his transfer is showing a negative sign, with every year, a player is older, his success is expected to decrease by 0.8 million EUR. This is in line with the initial assumption that young players have generally higher potential, are more sought after in the transfer market by the clubs and are therefore more likely to succeed.

With the success being defined as absolute change in market value of a player, it is only logical that the MV is a significant explanatory variable of said success, as expected in the assumptions. Furthermore, in line with the expectations, the β_2 coefficient has a negative sign, which can be attributed to different reasons. One of which being the expectations highly valued players face, which are categorically harder to live up to.

According to Table 4.1, the higher a player's value is, the less it is going to increase, as for every additional 1 million EUR in his estimated market valuation, he is estimated to decrease his success by 657 thousand EUR.

As shown in the Figure 4.5, there is a linear relationship between a player's transfer fee and his market value. It is therefore logical that the β_5 coefficient is statistically significant and, as illustrated in Figure 4.5, has a positive sign. For every 1 million EUR spent on a player's transfer, said players success is expected to increase by 342 thousand EUR. Although a transfer fee is strongly correlated with a player's market value, there are other factors influencing the amount, for which a player is sold, such as a club's negotiating position (e.g. with big clubs interested in the player, selling clubs can ask for higher amounts) or the pressure from the player and/or his agent (e.g. expiring contract, unwillingness to stay at the club, which decrease the value of the transfer, but not the value of the player). For these reasons, transfer fee and market value can be considered as two independent variables with no perfect collinearity, despite the correlation.

Figure 4.5: Scatter plot with trendline with MV on the Y-axis and fee on the X-axis



As show by the Table 4.1, the club, from which the player transferred plays an important role in his success. In line with the initial assumption, the higher a club is ranked in the Kickalgor.com ranking, the more successful a player's success is estimated to be. With 11 thousand EUR increase in a player's transfer success for every climbed rank of the previous club, the quality of the clubs, from which the players are transferring into the EPL plays a significant role in their market value growth and therewith, in their success.

Keeping the transformation of the MPA variable in mind, for every minute squared, which a player spends on average on the field during his appearances, his transfer success increases by 0.5 thousand EUR. As expected, the β_7 coefficient of this performance metric has a positive sign.

The positive relationship between a player's performance and his estimated success is further confirmed with the β_8 and β_9 coefficients. With every goal or a clean sheet, a player's success increases by 259 thousand EUR, based on the Table 4.1. Similarly, for every one-point increase of a player's average seasonal rating in one of the world's top leagues awarded by the Whoscored.com analytics, the estimated transfer success increases by 850 thousand EUR.

Seeing as a player's performance is the key determinant for his transfer, it is rational to assume, and the results of the regression confirm, that it should also be one of the main factors influencing his success.

Despite not being strongly statistically significant (level 0.01 or lower), the *EngL* and *NatT* variables are marginally important in explaining the success of a player's transfer, as excluding them decreases the adjusted R squared. The signs of the coefficients of β_3 and β_4 are in line with the initial assumptions, as they represent a certain advantage for the transferring player (being it an appearance for the national team, showcasing a player's elite performance, which is recognized by the selection to represent his home country, or his previous experience in the EPL in the past season, which would arguably allow him to adapt quicker, as he has experience against the players and teams from the league.), speaking to his qualifications and quality as a football player. Because of the relatively high p-values (0.17 and 0.27 respectively), estimation of the success of a transfer based on these variables should be perceived as a proximation.

Lastly, despite β_0 being statistically significant, no conclusions can be made on the true value of the intercept, as it is not possible to observe a player of age 0 or a player valuated at 0 EUR, nor is it possible to have a player transferring from a club ranked 0 on the Kickalgor.com ranking.

In conclusion, there is strong evidence to suggest that a player's performance (in this thesis represented through number of goals scored and clean sheets, minutes per appearance played and average league rating provided by Whoscored.com), his age, estimated market value by transfermarkt.co.uk and characteristics of the transfer (fee, for which he transferred to the new club and the quality of the club, from which the player transferred – using the league ranking from Kickalgor.com), are significant in explaining and predicting the success of the transfer.

On the other hand, based on the results of the regression, there is a lack of evidence to suggest that the performance of the team, the player is joining (quantified by the finishing position in the previous season), is related to the success of the transfer.

Similarly, the nationality of a player, quantified by the rank of a player's nation in the official FIFA rating, isn't statistically significant in explaining the transfer success. This can partially be explained by the EPL being a very diverse league with players from over 100 countries participating (Shergold, 2021), where the language barrier usually doesn't play a significant role anymore and there is more emphasis on individual performance. Furthermore, negative traits of a player, such as yellow cards received or number of suspensions, when included as individual variables, don't seem to be directly related to a player's transfer success either. They do however play a role, as they are included in a player's Whoscored.com rating and cannot therefore be completely disregarded.

5. Data limitation

Despite the model being a good fit for the data, explaining the success of a transfer relatively well, there are certain limitations the model has, which need to be kept in mind.

One of the main limitations is the lack of position-specific performance metrics. Because of data availability reasons, the dataset only includes basic metrics, which may interpret the capabilities of some positions better, than those of others. To be specific, a striker's success can be accurately evaluated based on scored goals, provided assists or MPA – although it could be argued that metrics, such as shots on goal or chances created would be more optimal at evaluating performance of attacking players. On the other hand, a defender's capability and performance is only partially explained by metrics mentioned above. Further data points (such as successful tackles, clean sheets and goals conceded) would be necessary to obtain a full indicator of performance. Although more detailed statistics are included in the *WSrating* variable, providing a more complex insight into each player's performance, inclusion of specific metrics could bring further insight into the variance of a transfer success of a player.

Another important limitation of this model lies within the definition of success, used in this thesis. As the focus is on short term success (time period of one season), by default it omits players, which only became successful after a longer period of time. For instance, Rayan Aït-Nouri, player from the data set used, transferring to Wolverhampton Wanderers on loan from Angers SCO, was valuated at 20 million EUR at the time of the transfer, as reported by transfermarkt.co.uk. After the end of the 2020/2021 season, his estimated market value dropped to 18 million EUR, classifying his transfer as unsuccessful. However, at the end of the season 2021/2022 his market value increased to 22 million EUR, increasing by 2 million EUR against his market value at the time of the transfer. Reason for that could be for example time needed for assimilation, as the player was not only joining a new club, but also entering a new league. Although seemingly problematic, this creates opportunity for further research, which would focus on studying the transfer success from a different angle, since this thesis, as mentioned before, focuses on the short-term success of a transfer.

Shortly after the beginning of the 19/20 season, a worldwide outbreak of the corona virus shook the entire world. Besides affecting every aspect of everyday lives of hundreds of millions of people worldwide, this pandemic also had a significant impact on football. The 19/20 and 20/21 seasons in particular have been affected by it, with matches

getting postponed or cancelled and clubs' revenues (match revenue, TV revenue, commercial revenue, etc.) and attendance dropping. Consequently, both salary and transfer expenses decreased, as forecasted by Quansah et al., 2020, exposing the model to possible variance throughout different seasons. It would therefore be beneficial, to also examine different seasons without the presence of COVID-19 effect to see, whether the model would yield the same results.

6. Conclusion

This thesis aims to investigate the relationship between a player's transfer to the English Premier League and between his performance before the transfer, his characteristics and the characteristics of the transfer. For this, data for more than 100 transfers from the summer transfer window of 2020 from various sources was collected. OLS was then used to estimate the effect of each variable on the transfer success, which was defined as change in estimated market value in millions EUR.

The analysis of collected transfers showed that age is one of the most significant estimators of a player's transfer success. Younger players have shown to be more successful, as indicated by the results of the analysis. This is indirectly supported by previous research (Majewski, 2016) and is also in line with current transfer market strategies of football clubs world-wide, where football clubs target young talents, as they recognize their potential.

Furthermore, the transfer fee, for which the players are bought, is also statistically significant in explaining and predicting a player's transfer success. As clubs are paying millions of EUR, it is only logical that the more they spend on a player, the more likely he is to succeed, as maximized return on investment is expected with the main goal of the investor being profit.

When looking at the market value, transfers of players with lower market value (at the time of the transfer), estimated by transfermarkt.co.uk, are more likely to be successful. This can be attributed to several reasons, one of which being the peak, which highly valued players experience and which makes it hard for them to stay in their top form for a longer period of time. Also, this finding is in line with the previously mentioned strong relationship of age and a transfer success, as younger players are generally valued lower than older, more experienced footballers.

Apart from a player's age, market valuation and transfer fee, the success of his transfer also depends on the player's performance and on the performance of his previous club. All should be considered, when making decisions on transfer strategies.

On the other hand, there is lack of evidence to suggest that the performance of the club, where the player is transferring to and his nationality would play a significant role in the success of his transfer.

The main contribution of this thesis lies in providing deeper understanding of EPL transfer market mechanics and potentially in incorporating its findings in local club's transfer strategies. Based on the results of the transfer analysis, it is recommended that the main focus of clubs should be targeting younger players from high quality clubs, rather than older, highly valued players. It is also important to focus on individual performance of the transfer targets, rather than on their nationality, as that no longer seems to be a determinant of success in the EPL.

Possible future research could aim to replicate this analysis in other football leagues to see, whether the findings of this work are generally applicable or specific to the English Premier League. Also, as this thesis focuses on short term transfer success (one season), alternative approach can be taken to investigate long term success of a player's transfer.

7. Bibliography

Al-Asadi, M. A., & Tasdemir, S. (2022). Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques. *IEEE Access*, *10*, 22631–22645. https://doi.org/10.1109/ACCESS.2022.3154767

Barros, C. P., & Leach, S. (2006). Performance evaluation of the English Premier Football League with data envelopment analysis. *Applied Economics*, *38*(12), 1449–1458. https://doi.org/10.1080/00036840500396574

Bhilawa, L., & Fahriansyah, R. (2022). THE INFLUENCE OF PERFORMANCE, AGE, AND NATIONALITY ON THE MARKET VALUE OF FOOTBALL PLAYERS. *Assets: Jurnal Akuntansi Dan Pendidikan*, *11*(1), 1. https://doi.org/10.25273/jap.v11i1.8422

Burdekin, R. C. K., & Franklin, M. (2015). Transfer spending in the English premier league: the haves and the have nots. *National Institute Economic Review*, *232*, R4–R17. https://doi.org/10.1177/002795011523200102

Carlos, L.-P., Ezequiel, R., & Anton, K. (2019). How does Video Assistant Referee (VAR) modify the game in elite soccer? *International Journal of Performance Analysis in Sport*, *19*(4), 646–653. https://doi.org/10.1080/24748668.2019.1646521

Ćwiklinski, B., Giełczyk, A., & Choraś, M. (2021). Who Will Score? A Machine Learning Approach to Supporting Football Team Building and Transfers. *Entropy*, *23*(1), 90. https://doi.org/10.3390/e23010090

Deloitte Football Money League 2023. (2023). Www.Deloitte.Com. https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html

Dias named 2020/21 EA SPORTS Player of the Season. (2021, June 6). *Premierleague.Com*. https://www.premierleague.com/news/2164971

https://www.fifa.com

FIFA. (2023b). *One Month On: 5 billion engaged with the FIFA World Cup Qatar 2022™*. Fifa.Com.

Gebbett, M. (2018). *Predictive Analysis of Success in the English Premier League*.

Gerhards, J., & Mutz, M. (2017). Who wins the championship? Market value and team composition as predictors of success in the top European football leagues. *European Societies*, *19*(3), 223–242. https://doi.org/10.1080/14616696.2016.1268704

He, M., Cachucho, R., & Knobbe, A. (2015, September 11). *Football player's performance and market value*.

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. https://doi.org/10.1016/j.smr.2013.12.006

Huth, C. (2020). Who invests in financial instruments of sport clubs? An empirical analysis of actual and potential individual investors of professional European football clubs. *European Sport Management Quarterly*, *20*(4), 500–519. https://doi.org/10.1080/16184742.2019.1684539

Keppel, P. (2020, December 18). How the volunteers of data website Transfermarkt became influential players at European top football clubs. *Www.Ftm.Eu*. https://www.ftm.eu/articles/transfermarkt-volunteers-european-football

Locks, R., Utsunomiya, H., Briggs, K. K., McNamara, S., Chahla, J., & Philippon, M. J. (2018). Return to Play After Hip Arthroscopic Surgery for Femoroacetabular Impingement in Professional Soccer Players. *The American Journal of Sports Medicine*, *46*(2), 273–279. https://doi.org/10.1177/0363546517738741

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. https://doi.org/10.7206/jmba.ce.2450-7814.177

Margareta, L. M., & Malinda, O. (2022). The Effect of Performance, Age, Transfer Fee and Salary to the Market Value of Professional Players (Empirical Studies in European Leagues Football Clubs). *International Journal of Global Operations Research*, *3*(3), 101–107. https://doi.org/10.47194/ijgor.v3i3.148

Milenovský, M. (2015). *Does Money Guarantee Success in Football?* [Undergraduate Thesis]. Charles University, Faculty of Social Sciences.

Morrow, S. (1999). *The New Business of Football: Accountability and Finance in Football*. Palgrave Macmillan.

Newman, D. (2016). *Predicting Transfer Values in the English Premier League* .

Niederer, D., Engeroff, T., Wilke, J., Vogt, L., & Banzer, W. (2018). Return to play, performance, and career duration after anterior cruciate ligament rupture: A case-control study in the five biggest football nations in Europe. *Scandinavian Journal of Medicine & Science in Sports*, *28*(10), 2226–2233. https://doi.org/10.1111/sms.13245

Phatak, A., & Gruber, M. (2019). Keep Your Head Up—Correlation between Visual Exploration Frequency, Passing Percentage and Turnover Rate in Elite Football Midfielders. *Sports*, *7*(6), 139. https://doi.org/10.3390/sports7060139

Premier League. (2022). *Premier League explained*. Www.Premierleague.Com.

Preston, T. (2012). *THE DETERMINANTS OF FOOTBALL TRANSFER MARKET VALUES: AN AGE OF FINANCIAL RESTRAINT*.

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. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3689563

Quansah, T., Frick, B., Lang, M., & Maguire, K. (2021). The Importance of Club Revenues for Player Salaries and Transfer Expenses—How Does the Coronavirus Outbreak (COVID-19) Impact the English Premier League? *Sustainability*, *13*(9), 5154. https://doi.org/10.3390/su13095154

Saravia, A. (2022). You play the way you live: aggressive soccer play and social capital. *Soccer & Society*, *23*(2), 115–129. https://doi.org/10.1080/14660970.2021.1919636

Scholz, T. M. (2011). Strategic Renewal of Complex Systems Based on the Case of the 1. FC Kaiserslautern. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1907626

Shergold, A. (2021). It's the Premier League of Nations! Players from 113 different countries have featured in England's top-flight with France, Ireland and Spain producing the most... but do you recall these obscure one-offs from Malta, Faroe Islands, Gambia and Suriname? *Www.Dailymail.Co.Uk*. https://www.dailymail.co.uk/sport/football/article-10297073/Premier-League-nations-113-different-nationalities-played-English-flight.html

https://www.kickalgor.com

https://www.transfermarkt.co.uk

Wand, T. (2022). Analysis of the Football Transfer Market Network. *Journal of Statistical Physics*, *187*(3), 27. https://doi.org/10.1007/s10955-022-02919-1

https://www.whoscored.com

Wooldridge, J. M. (2012). *Introductory Econometrics : a Modern Approach*. South-Western Cengage Learning.