CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Jan Sinčák

Determinants of Football Players' Market Value in the Czech Football League

Bachelor's Thesis

Prague 2020

Author: Jan Sinčák

Supervisor: PhDr. Radek Janhuba, M.A., Ph.D.

Academic Year: 2019/2020

Bibliographic note

SINČÁK, Jan. *Determinants of Football Players' Market Value in the Czech Football League.* Prague 2020. 54 pp. Bachelor's thesis (Bc.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor PhDr. Radek Janhuba, M.A., Ph.D.

Abstract

This thesis extends research on determinants of football players' market values by analysing data from the Czech football league. These determinants have been studied in the biggest European leagues but smaller competitions have so far been overlooked. This thesis further extends the research by including participation and performance of young players in national youth teams and European youth competitions. Data were collected over the course of three transfer periods and their analysis shows positive and significant effects of performance on players' market values. Moreover, the results indicate a large imbalance between market values of players of the best teams and the others. The estimation also shows positive effect of scoring in international and European matches of youth teams. To the best of our knowledge, this factor has not been studied yet. The results also hold when the number of matches played are used instead of number of minutes and when outliers are removed from the dataset.

Keywords

Football, Transfers, Market Values, Czech League, Game Statistics

Abstrakt

Tato práce pomocí analýzy dat z české ligy rozšiřuje dosavadní výzkum faktorů, které ovlivňují hodnotu fotbalistů. Dosud byl vliv těchto faktorů zkoumán pouze na datech z největších evropských lig a menší soutěže byly upozaděny. Tato práce rozšiřuje dosavadní výzkum s využitím dat o účasti a výkonu mladých hráčů v mládežnických reprezentacích a v evropských pohárech mládežnických týmů. Data pocházejí ze tří přestupních období a jejich analýza ukazuje kladný a statisticky významný efekt výkonu v zápasech na hráčovu tržní hodnotu. Výsledky ukazují i velký rozdíl v tržních hodnotách mezi hráči nejlepších českých týmů a zbytku soutěže. Kromě toho nalézáme i statisticky významný efekt skórování v zápasech mládežnických reprezentací a v evropských mládežnických pohárech. Dle našich poznatků je toto první práce, která si tohoto efektu všímá. Výsledky modelu jsou obdobné i při využití počtu odehraných zápasů namísto odehraných minut a při odstranění extrémních hodnot z datasetu.

Klíčová slova

fotbal, přestupy, tržní hodnoty, česká liga, herní statistiky

Declaration of Authorship

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

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

Prague, 4 May 2020

Jan Sinčák

Acknowledgments

I am very grateful to my supervisor PhDr. Radek Janhuba, M.A., Ph.D. for his help, time and guidance. I would also like to offer my special thanks to my family and friends for their undying support throughout my study.

Bachelor's Thesis Proposal

Research question and motivation: Transfer windows in football are eagerly awaited segments of a football season all around the globe, Czech Republic being no exception. Teams are spending large amounts of money, hoping that they will sell players they no longer need and sign their target players at the best transfer price. In order to predict the value a player may have, teams use a variety of models which give players a certain market value based on their performance on the pitch.

While a fair amount of research has been conducted in the best European leagues on what the factors that determine a player's market value are, there are no models based purely on the data from the Czech first football league. We believe that there are specific factors in the Czech league in which it differs from the best European leagues, making the foreign models less precise. One such factor could be the public conviction, that players from the Czech league have smaller market values than their equally skilled counterparts from the western leagues. Some factors may also be more important in the Czech Republic. Performance in the Champions League or Europa League may influence the transfer values differently because Czech teams do not participate in these competitions that often. We want to test how the factors, that determine a transfer value, differ in the Czech league and create a model which would suit these different conditions better.

Research questions:

1. Which factors can influence players' values differently in the Czech League?
2. Do these factors actually influence the players' market values differently?

Contribution: This thesis will extend the current research on football players' market values, taking into account specific aspects of the Czech first football league. If the results indicate presence of the specific aspects, the created model will be a better representation of the Czech transfer market.

Methodology: We are going to use detailed data on match performance (such as accuracy of passes, number of goals scored, number of successful tackles et cetera) of players who have been a part of a transfer in/out of/within the Czech first football league in the last several years. The exact specification of the methodology will be chosen based on the available data and diagnostics of estimated models. The match performance data will be obtained from the official website of the league and from a professional football platform wyscout.com. The data on market values and transfer prices will be provided by transfermarkt.de. We will analyze this data using a multiple linear regression and OLS estimation.

Outline:

1. Introduction
2. Theoretical Background
3. Description of the Data
4. Model
5. Results and Discussion
6. Conclusion

Bibliography:

He, Miao & Cachucho, Ricardo & Knobbe, Arno. (2015). Football player's performance and market value. Conference Paper, Machine Learning and Data Mining for Sports Analytics PKDD/ECML (pp. 87-95).

Majewski, Sebastian. Identification of Factors Determining Market Value of the Most Valuable Football Players. Journal of Management and Business Administration. Central Europe. 2016, 24(3), 91-104. ISSN 2450-8829.

Müller, Oliver, Alexander Simmons and Markus Weinmann. Beyond crowd judgments: Data-driven estimation of market value in association football. European Journal of Operational Research. 2017, 263(2), 611-624.

Newman, Dylan. Predicting Transfer Values in the English Premier League. Seminar paper, department of Economics, Duke University

Wicker, Pamela & Weimar, Daniel & Prinz, Joachim & Deutscher, Christian & Upmann, Thorsten. (2013). No Pain, No Gain: Effort and Productivity in Professional Soccer. International journal of sport finance. 8. 124-139.

Contents

List of Figures

List of Tables

Introduction

Football is one of the biggest sports in the world. Millions of fans watch not only the games, but they also keep an eye on transfers which are made during two transfer periods each season. The rumours about the sums of money spent on individual players receive lots of attention, as the sums are getting higher each year. But these sums tend to be much more important for football clubs since they need to balance out the earnings from selling their players and the amount of money spent on incoming players. The clubs therefore need to have means of evaluation of players based on their characteristics.

The topic of player evaluation has been studied by numerous papers which focused on the biggest European competitions. Smaller leagues have not been considered before since they do not have such a large audience. This thesis aims to study determinants of value of players from the Czech football league. It hopes to extend previous research from other leagues and to seek factors which may not have been considered before. Various performance indicators from multiple competitions are to be used and their contribution to players' market values analysed. This will allow us to perform a more precise analysis which will take into account possible differences between the Czech league and its previously studied counterparts.

The rest of the thesis is structured as follows. Chapter 1 provides an overview of existing literature on the topic of football players' evaluation and presence of superstars in football leagues. Chapter 2 deals with the question of credibility of market values provided by website *Transfermarkt*. In Chapter 3, we describe the dataset we use and sources of this data. Chapter 4 focuses on the methodology and description of our model. In Chapter 5, we present the results of our analysis and discuss them. Moreover, we perform robustness checks of our model. Finally, we conclude.

1 Literature Review

This section is composed of three subsections and summarizes existing literature on evaluation of football players. Section 1.1 provides an overview of early research which had to deal with little to no availability of detailed data on players' performance. Section 1.2 takes a look at latter papers with focus on theory of rise of superstars and its application in European football leagues. In Section 1.3, we delve into workings on effects of players' performance characteristics. Finally, Section 1.4 concentrates on characteristics not related to players' performance, such as height or position, as an explanatory factor of market values.

1.1 Early Research

Historically, academic literature has focused more on American sports such as baseball, basketball or American football while research on European football was left behind. Frick (2007) mentions several reasons why this was the case: unlike US sports teams, European football clubs did not disclose their salaries paid to players and transfer fees paid to other clubs. Only exceptions were the English league and German Bundesliga where highly detailed data were available thanks to a newspaper *Welt am Sonntag* and football magazine *Kicker*. Since then the situation has been getting better in other European countries too. Also, in the past, the market for football players has been very rigid compared to standard labour market. Transfer fees had to be paid even if the player's contract expired. Football leagues have also established strict rules on number of foreign players playing a match. This has changed in December 1995 when the European Court of Justice found these restrictions incompatible with the Treaty of Rome.

These changes led to a higher interest of researchers in European football markets, but early papers only used basic match and player statistics to find determinants of transfer fees. Dobson et al. (2000) studied determinants of transfer fees in English semi-professional (non-league) competitions. Not having detailed player statistics, they used more general statistics such as home stadium capacity or the average number of spectators. Some of the interesting factors affecting transfer fees in non-league divisions were the average attendance of buying and selling teams' games or ground capacity of buying team's stadium. Other papers have focused on professional leagues, but they have still worked only with basic player statistics.

Carmichael & Thomas (1993) first try to regress transfer fees on club characteristics from a dataset of 214 observations from season 1990/1991. They find goal difference, financial status and average attendance to be significant. After adding in age squared, match appearances and goals scored, these show as significant too. They also find that age squared has a negative effect because player's experience keeps increasing in time and this increases the transfer fee. But his physical abilities reach a peak at one point and then keep decreasing, making the transfer fee go down as well. This is captured by the square of age.

On a sample of 164 transfers from 1985 to 1990, Speight & Thomas (1997) seek effects of player and club characteristics on bargained transfer fee. They find a positive effect of age but negative effect of age squared. This corresponds to the study of Carmichael & Thomas (1993). Similarly, Speight & Thomas identify number of goals scored and number of appearances during career and goals difference of selling team as significant.

Carmichael et al. (1999) use a tobit model on data from season 1993/1994 because transfer fees cannot be negative. Then they compare results of tobit model to those of an OLS regression. They conclude that estimates of OLS regression are similar to marginal effects of the tobit estimation. They also introduce new independent variables: difference in goals scored between two last seasons and dummies representing lower league divisions which the player came from. Difference in goals scored shows the trend of player's ability to score and may be seen as a predictor of his future performance. Carmichael et al. show that they are significant in both tobit model and OLS regression.

1.2 Talent and Superstars

As more detailed performance and non-performance data has become available, researchers have started to use them extensively. This allowed them to use more complicated models which to lead to new questions. Why are top five per cent of players valued much more than the median player? What is the decisive factor since players' performances do not differ so enormously? What does it take to be a superstar (Lehmann & Schulze, 2008)? In their articles, Rosen (1981) and Adler (1985) take two different approaches to combat the issue and come up with unique solutions. Rosen (1981) argues that consumers prefer high quality performances (of athletes, artists, etc.) and lower quality alternatives are an imperfect substitute. In the same way as most people are not satisfied with a reasonably priced but less talented artist when they can enjoy a more expensive top class performance (Frey, 1998), spectators prefer matches of the best football players. Strong economies of scale allow better performing athletes bring large audiences to their games and small differences in performance therefore lead to massive disparity in earnings.

Adler (1985) offers another explanation which comes from network externalities. He argues that in order to value athlete's or artist's performance, people need to get to know them better and they do this by talking about the performer with others. The knowledge comes from consumption of athlete's performance and from discussing his performance with other knowledgeable people. The more the athlete is known, the easier it is to find his fans. As the athlete is becoming more popular, the number of his fans keeps increasing faster and faster thanks to the network externality. In the end it is media who have the power to create stars as they can spread knowledge about an athlete. This knowledge of the public then turns the athlete into a superstar.

The two approaches to formation of superstars are not mutually exclusive, but they rather complement each other. In an attempt to test these hypotheses, Franck & Nüesch (2012) not only use talent characteristics to predict player's market value but also a number of nonperformance-related press citations. Using data from five seasons (2001/2002 until 2005/2006) of German Bundesliga, they found clear evidence that both player's talent and his popularity increase his market value. Similar results were achieved using three different regression models. OLS, fixed effects regression and 95% quantile regression were used and all of them showed that both talent and popularity increase player's market value. The effects remain robust even when a different source of market values is used. Franck & Nüesch also treat players based on distribution of market values. Their paper shows that there is a difference in effect of independent variables on the market value for average players and superstars. They label a player as a superstar if his market value is at 95% quantile and find out that the marginal benefit of superstar's performance is much higher than in case of an average player. For instance, scoring a goal increases superstar's value four times more compared to an average player.

Garcia-del-Barrio & Pujol (2007) used data from season 2001/2002 of Spanish football league and looked at player's contribution to the club through sales of merchandise as one of the factors influencing his market value. They used the number of websites found when the player's name is searched on Google which is again a popularity measure. Despite the fact that sporting performance is an essential explanatory factor of player's market value, they found out that popularity of the player is also a major factor in determining his value. Garcia-del-Barrio & Pujol label this outcome as "the most relevant finding of their paper". Apart from this model, they also use another one to study a "winner-take-all" phenomenon. In this model, dummy variables are used for groups of the most popular players of the league, who are identified by number of results on Google. Dummies representing ten most popular players of the league and the most popular player of each team are deemed significant and the magnitude of these estimates exceeds all other estimates in the model. This leads to the conclusion that there is a winner-take-all setting in the Spanish league. It means that the most popular players of the league and each team enjoy much higher market values than the rest of the players, but this increase is out of proportion compared to their marginal economic and sport benefit to the team. The results also show an increase of value for players with experience in national team and international club competitions. Besides that, Garcia-del-Barrio & Pujol found that non-Spanish European players were systematically overrated, whereas players from countries outside of Europe suffered from being underrated. But the respective t-statistics were on the border of significance.

In contrary, Reilly & Witt (1995) found no evidence of racism and discrimination in English league transfer market. This issue was later re-examined by Medcalfe (2008) who arrived at the same conclusion. Such results may lead to the conclusion that there is no evidence of racial discrimination in these football leagues when it comes to evaluating players for their performance.

In his paper, Frick (2007) finds an interesting connection between players' wages and their market values. A comparison of multiple papers focusing on determinants of salaries and market values shows that the independent factors are very similar, and they affect the wages and values in the same way. Papers included in this comparative study use data from major European leagues (England, Germany, Italy and Spain) and they arrive at similar conclusions. This finding allows us to turn our attention to papers which study the connection between players' abilities and salaries.

Another similar research on player popularity and superstars in football but with a very different outcome was conducted by Lehmann & Schulze (2008). Their paper "What Does it Take to be a Star? — The Role of Performance and the Media for German Soccer Players" focuses on the relationship between players' salaries and their performance and popularity in late 1990s' German Bundesliga. While an OLS regression delivers expected results, quantile regression shows no evidence of the superstar phenomenon in German football league. Instead, their analysis points in the direction of diminishing returns to performance. This is in sharp contrast to Rosen's (1981) theory of superstars. Moreover, no evidence of Adler's (1985) explanation of emergence of superstars is found either. There is a highly significant and positive influence of media interest on players' salaries, but only in case of average players. Similarly to performance, effect of media coverage is deemed highly negative by a quantile regression model, leading to diminishing returns to media presence. Thus, media coverage is definitely an important aspect in determining players' salaries, but it cannot explain the superstar status in terms of wage. Another interesting finding is that neither performance nor popularity measures are regarded as significant in explaining superstars' earnings. This paper thus comes to very different conclusions compared to the research of Franck & Nüesch (2012), who identify both talent and popularity as significant factors in explaining superstars' value in Bundesliga.

In their paper, Herm et al. (2014) use most often mentioned talent attributes and several external determinants of market values. Among others, these are grades from football experts, total value of all players in the team and total number of hits reported by Google, similarly to Garcia-del-Barrio & Pujol (2007). On top of those, Herm et al. introduce a new independent variable, player's sports agent. This variable is set up as an average value of all players in the agent's portfolio. They argue that successful agents are willing to manage only successful and therefore valuable players. Like Franck & Nüesch (2012), they also compare multiple models. After accounting for effects of external attributes, the estimates of some talent variables become insignificant, while all but one estimate of external attributes are significant. An OLS estimation and 95% quantile regression are used and neither of them marks the player agents as significant in determining players' values. On the other hand, grades received, popularity, value of team and number of games played are considered to be significant.

1.3 Effects of Players' Performance

The first factor which comes to one's mind when thinking of determinants of market values, is most likely how well the player works on the pitch. This very general and not well-measurable foundation of player's performance can be decomposed into multiple elements.

The most important one is possibly the amount of play time a player gets. This metric is heavily influenced by the amount of confidence a manager has in the player. A player who underperforms will either get no play time at all or will be soon substituted by his teammate. Appearances in domestic games have thus been included in multiple papers and have been deemed significant with a positive effect on the market value (e.g., Franck & Nüesch, 2012; Wicker et al., 2013). Some papers have also differentiated between appearances in domestic and European competitions and games played for the national team. These matches are also proved to increase players' value and the magnitude of their effects is much larger (e.g., Garcia-del-Barrio & Pujol, 2007; Bryson et al., 2012). This is caused by the fact that only the best teams, that own the best players, get to play European competitions and only the best players have the privilege to represent their nation in international games. Number of minutes played has also been used instead of number of matches (Ruijg & van Ophem, 2014) in order to utilize a more precise measure of presence on the pitch.

The most attractive elements of a football match are probably goals and thus scoring is highly rewarded across competitions. Besides a raw number of goals in a season, multiple other measures such as goals scored throughout the whole career or a goal difference between the last two seasons were used and found significantly positive (e.g., Carmichael et al., 1999; Herm et al., 2014; Müller et al., 2017). Besides that, number of shots is another similar measure which has been included in several studies, but with mixed results. While He et al. (2015) have identified the number of shots on target as significant, Franck & Nüesch (2012) found no evidence of its effect on market values.

There is one more statistic, which is closely associated with the number of goals scored and that is the number of assists. An assist is a crucial final pass, which leads to a scored goal and it should therefore affect player's market value as well. Franck & Nüesch (2012) and Müller et al. (2017) have conducted analyses of influence of the number of assists on market values in multiple European football leagues. Their results indicate that assists have a significantly positive impact, similarly to the number of goals. Lehmann & Schulze (2008) come to the same conclusion in relation to players' salaries.

Other statistics have not been included in the models in general, different characteristics have been chosen by different researchers. Among those are passing in the form of the percentage of successful passes (Herm et al., 2014; Müller et al., 2017) and total number of completed passes (Müller et al., 2017). Both have been regarded as significant and with a positive effect. Medcalfe (2008) included the percentage of completed dribbles but found no evidence of its effect on transfer fees. This is in contrast to research of Müller et al. (2017) and He et al. (2015) who used a LASSO regression and deemed dribbling as significant.

We can identify one larger group of statistics which revolve around fights for the ball. This includes duels over the ball, aerial duels, tackles and also fouls which come as a consequence. There is little consensus among studies on either the sign of the effects or on their significance. While duels are believed to increase players' values (Franck & Nüesch, 2012; Müller et al., 2017), the models provide mixed results in case of the other statistics (Lehmann & Schulze, 2008; Medcalfe, 2008). In the same way as there is no unanimity on the effect of conceded fouls, there is no clear outcome on the matter of yellow and red cards. On one hand, by receiving a yellow card, the player faces the risk of being booked again and being sent off the pitch and thus weakening his team for the rest of the match. This should clearly lower the player's value. On the other hand, there is another possible view of this. The player may choose the lesser evil and receive a yellow card, while preventing the enemy team from getting into a dangerous position. Such behavior can be acknowledged by others and increase the player' market value. The effect of being booked was thus difficult to evaluate and as such, existing literature points in both directions. Müller et al. (2017) come to the conclusion, that yellow cards significantly decrease market values, whereas Franck & Nüesch (2012) see the effect positive, although insignificant. The same discussion may be lead in the case of red cards. But in this case there is no evidence, that red cards would influence players' market values (Franck & Nüesch, 2012; Müller et al., 2017).

Several researchers have included interaction terms based on players' position in their analyses. Because not all players have the same role on the pitch, it may be desirable to take this into account in the analysis. While attackers are expected to score goals, defenders are supposed to prevent the enemy from scoring and mifielders ought to deliver accurate passes in order to support their teammates. The number of tackles may therefore not be as important when evaluating a forward as it is in case of a central back and vice versa. Carmichael et al. (1999), Dobson et al. (2000), Lehmann & Schulze (2008) have included these interaction effects in their models and found them generally significant. To account for the different roles of each player, it is also possible to use performance indices made by football experts, as done by Garcia-del-Barrio & Pujol (2007). These indices take into account players' contribution based on their position and their tasks. The indices have been regarded as highly significant in explaining players' market values.

1.4 Nonperformance Characteristics

Several papers focus on players' physical attributes such as height or footedness. Bryson et al. (2012) studied the effect of footedness on footballer's salary. They argue that there are several reasons why two-footed players should enjoy higher remuneration compared to their one-footed teammates. First, being able to play with both feet should allow the player to score more goals since they would be able to shoot from more positions or make it harder for defenders to predict their next move. It should also allow them to complete a higher number of passes and to tackle opponents more precisely. Second, two-footed players are likely to make use of more positions on the pitch. Such ability is greatly valued by managers as the player is more flexible and can carry out more tasks on the turf. Finally, Bryson et al. (2012) believe that there may be a link between players' intelligence and two-footedness. This assumption comes from the study of Denny & Sullivan (2007), who argue that there may be a connection between a person's IQ and left-handedness. Bryson et al. (2012) claim that a more intelligent player could better foresee possible attack opportunities. And he would also be stronger in controlling the midfield, which is an important aspect of a football match. Using two datasets, one from Bundesliga and one from multiple European competitions, Bryson et al. (2012) found substantial evidence that two-footedness leads to a higher salary even after controlling for other performance measures. On top of this, they found no evidence, that using more two-footed players positively affects team performance. Based on these two outcomes they conclude that there is no mispricing of players in European competitions.

Herm et al. (2014) use similar arguments as Bryson et al. (2012) when they include two-footedness in their model as a possible factor affecting players' market value. Their model shows that there is positive influence of two-footedness on market values, but it is not strongly statistically significant. Moreover, this holds only when popularity measures are not accounted for. Influence of flexibility then disappears.

Another discussed physical characteristic is player's height, which was included in several papers. This factor was suggested by Bryson et al. (2012), because taller players should benefit from their height during heading. Being tall should be advantageous in various match situations; a tall striker will be able to beat most of his opponents when heading at the goal. On the other hand, a tall defender will be able to clear the ball from the penalty area during corner kicks more easily and a goalkeeper will benefit from his height when diving for shots on target. But Bryson et al. (2012) come to the conclusion that height is not significant. They argue that the insignificance is caused by little variation in height across the dataset. This issue was later revisited by Müller et al. (2017), who use the same arguments for including this variable. Unlike Bryson et al. (2012), they only use height and not its squared term. But they arrive at the same conclusion, stating that the effect of height is insignificant, as there is little variance in the data and little to no effect on market value in the model. Neither Wicker et al. (2013) find evidence of significance of height.

Finally, player's position on the pitch is also an important factor which could influence the market value. Players' positions were included in the paper of Garcia-del-Barrio & Pujol (2007). Their models based on data from the Spanish league included dummy variables representing defenders, midfielders and attackers with goalkeepers being left out as a reference group. Instead of individual performance characteristics, they used two different performance indices created by Spanish football journalists. As there are two indices, Garcia-del-Barrio & Pujol (2007) also use two models since the indices are undoubtedly correlated. The two models identically label attackers as significant and show that they are enjoying much larger market values compared to goalkeepers. This is not the case for defenders and midfielders, though. While one model deems defenders as insignificant and midfielders significant, the other model does the opposite. This is most likely caused by a different metric used to compute each performance index. The difference between goalkeepers and attackers is supported by other works too. Frick (2007) shows that goalkeepers are at the lower end of salary distribution and that attackers are enjoying up to 30% higher wages, as they are much more visible and can draw lots of attention (He et al., 2015).

2 Background on Crowd Estimations

This chapter begins with the discussion on transfer fees and market values in Section 2.1. These terms may seem interchangeable at first glance, but this section shows the difference between these two measures and explains what makes them different. Next, Section 2.2 provides insight into crowd-based estimation of football players' market values. First, it presents a brief overview of research on crowd estimations in general. Later it follows up with discussion on strengths and weaknesses of the process, which is used by *Transfermarkt* to estimate market values of football players.

2.1 Market Values vs. Transfer Fees

It is crucial to see the difference between market values and transfer fees. While a transfer fee represents the actual amount of money spent for a football player, market value may be defined as "an estimate of the amount of money a club would be willing to pay in order to make an athlete sign a contract, independent of an actual transaction" (Herm et al., 2014). Football clubs can use market values as an indicator of player's value even if he was not recently traded, which makes them very important during trade negotiations. The two numbers may differ due to bargaining power of the buying and selling club, sum of bonuses paid to player's agent or due to intentional overpaying (in order to prevent competitors from acquiring the player for his market value).

A major source of market values is a website *www.transfermarkt.com*, which has evolved into one of the most trusted pools of market values. The estimates it provides are proved to be very accurate approximations of actual transfer fees (Herm et al., 2014). This is why they have served as a foundation for numerous studies on football transfer market (e.g., Franck & Nüesch, 2012; He et al., 2015; Müller et al., 2017).

2.2 Wisdom of Crowds

The key to such accuracy does not lie in complicated econometric models but in crowd judgement. Transfermarkt relies on its community of contributors, who are fans of the sport and players. The idea is that a large group of common fans is able to judge player's value more precisely than a handful of football experts. The first study on "Vox Populi" was conducted by Sir Francis Galton (1907), who asked visitors of a cattle exhibition to guess weight of an ox. To his surprise, the median vote was less than one per cent off from the true weight. More researchers have since analysed this phenomenon. Among others Surowiecki (2005), who introduced the term "Wisdom of Crowds" for this type of community estimation. Wolfers & Zitzewitz (2004) conclude, that crowds are very efficient when it comes to aggregation tasks. And estimation of player's market value is such an aggregation task since there are many inputs in form of performance statistics, which have to be converted into a single variable (market value). This makes a crowd of fans so successful in their analyses.

Some of the challenges crowd judgement faces are lack of experience and knowledge, personal bias or intended manipulation and these may negatively impact outcome of the estimation (Lorenz et al., 2011). But one of the reasons why Transfermarkt is so successful is its ability to deal with these challenges. Herm et al. (2014) have described its system. The main feature of the estimation process is its lack of democracy. Herm et al. (2014) call it the judge principle. Transfermarkt does not give each estimate equal weight but uses a hierarchical system with the help of "judges". These judges are trustworthy members of the community and they have the power to change the final estimation by giving different weights to different estimates according to their belief. They are guarantors of accurate estimations as they should be more qualified and less prone to being biased than an average member.

Despite its advantages and undeniable accuracy, this system also comes with several limitations, as Müller et al. (2017) point out. One of them is the lack of unified estimation procedures. Thus, each user may take into account different performance indicators, which makes the result biased. Another concern is that there is no supervision of the judges and their procedures and there is no strict method of giving weights to members' estimations. As the judges' decisions are also based only on subjective measures, there is no possibility of reproducing their results. This may again in some cases lead to inaccurate estimations. The process also requires a large community of contributors in order to produce satisfying results. This may be a problem in case of smaller leagues and less known players as they do not have a large fanbase. Less famous divisions are also disadvantaged by the fact, that updates of the market values are rare. And updates on market values for major leagues are also limited. It is nearly impossible to frequently evaluate such a huge number of players and therefore Transfermarkt change the estimates every six or twelve months.

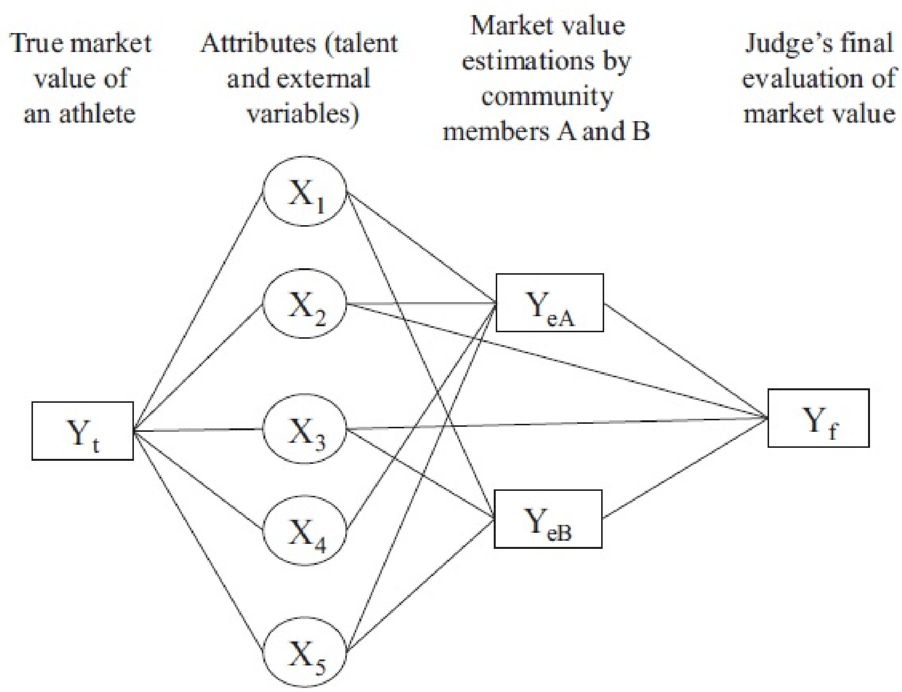

Figure 1: Transfermarkt Estimation Process (Source: Herm et al., 2014)

3 Data Description

3.1 Sources of Data

In order to create our dataset, we had to combine three sources of data. Our largest portion of data is on players' performance statistics. This was obtained from *www.wyscout.com*, which describes itself as a "professional platform for people working in the football world". This website offers numerous ways to inspect players' performance by accumulating match and career statistics, offering video clips of notable match moments and providing detailed analyses of football players. The performance data is available for more than four hundred football competitions from around the globe, including first and second Czech football leagues. We chose this platform as its data are easily available and used by many clubs from the best European competitions. This persuaded us that the collected data are relevant and accurate.

We collected data on players who made a transfer in, from or within the Czech first division during three different transfer periods (winter 2019, summer 2019 and winter 2020). Performance data were collected from matches which took place within one year before the player's transfer. We chose this limit because we are convinced, that clubs look at player's most recent performance when deciding whether to sell or to buy him and for how much money. The three transfer periods were chosen because we wanted to analyze as recent data as possible. And because Wyscout does not offer a simple way of downloading a compact dataset for matches which took place in 2017 or earlier. Our earliest match data therefore come from January 2018. Since we use data from one year before a transfer for our analysis, our boundary is set to the transfer window of winter 2019. In case of the summer transfer window, the cut-off date for data collection was 30$^{\text{th}}$ June 2019. This date was chosen because according to the league rules, the season must finish not later than June 30$^{\text{th}}$. For winter transfer periods, we chose 31$^{\text{st}}$ December as the last day.

The other component of our dataset are market values. These were collected from *www.transfermarkt.com*. Transfermarkt specializes in collecting data on players' transfer fees and estimating their market values. We have collected market values on all of our observations. We have chosen market values instead of transfer fees because Czech teams do not tend to make fees public very often. Most of the times these figures are leaked and do not come from official club statements. Transfermarkt also provided us with data on players' age at the time of a transfer and information on buying and selling clubs.

Final source of our data was a website *www.livesport.cz*. Livesport is a Czech company which specializes in providing live match scores in more than thirty sports and it has become one of the leaders in the market of sports scores. Livesport has served as a complementary data source as we have used it to fill in some missing data on players' age and to check that Wyscout provided us with correct number of matches for each player. We have also used it together with Wyscout to determine players' positions on the field and to split those into three groups (defender, midfielder, forward).

3.2 Data Characteristics

3.2.1 Dependent Variable

The dependent variable of our interest is the market value in euros. Each observation is accompanied by a market value which represents an estimate of the transfer fee at the time of transfer. Market values were chosen because transfer fees are mostly kept private by the clubs. We have acquired a total of 178 observations from the three transfer periods. Market values were corrected for inflation with base year set to 2020. Table 1 provides descriptive statistics of our dataset. Table 2 shows frequency distribution of qualitative data.

3.2.2 Independent Variables

We have used a total of 19 independent variables, majority of which are performance characteristics according to prior designation. Apart from the data provided by our sources, we have also created two categorical variables. The first one is *Period* which serves as a distinction of the three included transfer periods and takes on values “W19”, “S19” and “W20” which represent Winter 2019, Summer 2019 and Winter 2020. The other is a dummy variable *Topteam*. This variable marks transfers made by three teams, namely SK Slavia Praha, FC Viktoria Plzeň and AC Sparta Praha. These are the most ambitious clubs with highest budgets. It is therefore possible they would overpay their incoming players in order to prevent others from acquiring them. These teams are also most visible to the foreign clubs who look for new players in the Czech league and who offer much higher transfer fees compared to local standards. Slavia and Plzeň enjoy this status thanks to their recent successful performances in European competitions and Sparta are still known for their past participations.

The rest of the variables come from above-mentioned data sources. Three of them are nonperformance characteristics with the remaining 14 being performance characteristics. Since we are not working with a large dataset, we are relying mainly on the most commonly used variables in prior analyses. But on top of those, we have also included several variables, which were not used very often.

The most important variable is most likely *Minutes*. Similarly to other studies, it marks the number of minutes a player has spent on the pitch during the season. This is possibly the most important explanatory variable of the market value and we expect it to have a strictly positive influence. Besides minutes, we will also use *Matches* in another estimation, and we will compare their effects. The effect of one additional match is expected to correspond to the effect of 90 minutes but *Minutes* were chosen as the primary variable because they better explain involvement of players who usually serve as substitutes and get to play less minutes, even though their number of matches may be high. These variables only apply to the Czech league, European and international competitions are described independently.

Goals and *Assists* reflect players' productivity which is again a very important indicator of their performance. Same as is the case of *Minutes* and *Matches*, these statistics only apply to the Czech league. Last variable which is unique to the Czech league is *Passespercent*. This variable denotes the total percentage of successful passes from all matches of the Czech league combined. All of these variables are expected to increase the market values.

Age is another variable which is likely to be significant, since it may be seen as player's experience. Therefore, we expect it to increase the market value. On the other hand, as players get older, their physical condition tends to decrease. They may not be able to run as fast as before, or they could be more prone to getting injured. This represents limitations of their benefit to the team and as such, decreases their value. This relationship is accounted for by a square term of age.

We have also included more categorical variables. *Position* assigns each player their position on the pitch (Defender, Midfielder, Forward). This distinction may be important because according to prior research, there can be differences in evaluation of players on different positions. Goalkeepers are not included in our dataset because their role on the pitch is different from other players and their performance is based on very different statistics. The other variable is *Foot* which distinguishes whether a player has a stronger right or left foot. Left footed players make up less than a third of players in our dataset but teams usually need a similar amount of left footed and right footed players because it is difficult for a right footed player to play on the left side of the pitch. Thus, left footed players should be more valuable.

The rest of our variables come from competitions other than the Czech league. The best Czech clubs get a chance each season to play against the best European teams in either the Champions League or Europa League. These competitions are watched by millions of people around the world and a player can therefore get lots of attention from abroad when he delivers a good performance. We believe this can dramatically increase his market value. For this reason, we have included variables *Eumatches*, *Euminutes* and *Eugoals* in our dataset. These describe players' involvement and number of goals in European competitions.

We have used a similar approach in terms of international matches. Only the best players are offered to play for their national team. These matches again get more attention from abroad compared to matches of the Czech league. We have therefore included variables *Natmatches*, *Natminutes* and *Natgoals*. Since there are not many international games played during a season, we have decided to include both matches from international competitions (e.g., UEFA Nations League) and from international friendlies. We believe that the influence of appearing in an international game is comparable no matter the circumstances.

The effect of appearing in European and international competitions has not been taken into account by many previous papers. This may be caused by the fact that they focused solely on the best European competitions, where many more clubs take part in European competitions and lots of players figure in the best national teams. The level of clubs' strength and the number of spectators is thus comparable between local and international competitions. But the Czech football league presents a much smaller market, where appearances in other competitions may be more important.

This reasoning has brought us to the last variables used in our dataset which are *Youthmatches*, *Youthminutes* and *Youthgoals*. These variables represent participation and performance in the youth national teams and in the European youth competitions, specifically U19 — U21. To the best of our knowledge, this is the first work that implements these variables in estimating players' market values. There are many young players with a similar level of skill and with little playing history but only the best ones are chosen to represent their country or to play in a European competition. This gives them some advantages compared to other young players. First, the players signal to potential buyers that they are the best of their generation.

Second, these matches give players additional experience, making them even better compared to the other youngsters. In our opinion, these facts should lead to an increase of value of players with experience from youth teams.

Table 1: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Value (€)	178	632,720	1,024,367	25,000	150,487	687,500	6,500,000
Age	178	24.43	3.58	17	22	27	36
Matches	178	17.97	10.74	0	9	26	36
Minutes	178	1,328	961	0	450	2,094	3,414
Goals	178	2.38	3.75	0	0	3	21
Assists	178	1.11	1.87	0	0	2	10
Passespercent	178	66.63	25.38	0	67	81	91
Eumatches	178	1.15	2.90	0	0	0	14
Euminutes	178	92.31	244.34	0	0	0	1,191
Eugoals	178	0.12	0.42	0	0	0	2
Natmatches	178	0.60	1.69	0	0	0	9
Natminutes	178	39.88	131.94	0	0	0	821
Natgoals	178	0.08	0.43	0	0	0	4
Youthmatches	178	0.52	1.57	0	0	0	9
Youthminutes	178	40.31	131.19	0	0	0	766
Youthgoals	178	0.04	0.22	0	0	0	2

Table 2: Frequency Distribution of Qualitative Data

	Winter 2019	Summer 2019	Winter 2020
Period	61	74	43
	Defender	Midfielder	Forward
Position	61	66	51
	Right	Left	
Foot	126	52	
	Yes	No	
Topteam	44	134	

4 Methodology

4.1 Choice of Model

The choice of methodology is derived from the structure of data used. Our dataset comprises cross-sectional observations measured during three subsequent time periods. One type of data which makes use of multiple time periods are panel data. But panel data require a fixed group of cross-sectional members which are studied across multiple time periods (Wooldridge, 2016). Since our dataset constitutes of independent random samples from three different time periods, we identified it as a pooled cross-sectional dataset instead. This seems to be the most appropriate description since none of the included players made a transfer during every single transfer period.

Pooled cross-sectional data offer some advantages in comparison to standard cross-sectional data. Their most important aspect is that including multiple time periods allows us to obtain more observations for our dataset. This helps us in obtaining more precise estimators with more power (Wooldridge, 2016).

The use of pooled cross sections also raises some complications, although very minor and relatively simple to deal with. As the population may have different distributions across time, we have to account for this and allow for a different intercept across time periods. This is easily done by including dummy variables representing each but one time period. This period is then chosen as the base time period. The dummies themselves may in certain cases also be of interest. They allow us to observe the changes in the explained variable across time after controlling for the explanatory variables.

A vast majority of the existing literature referred to in this thesis use the Ordinary Least Squares (OLS) method to estimate the models. In several cases this is accompanied by a quantile regression estimation in order to examine the possible differences between average players and superstars. Since we do not aim to analyze this phenomenon, we shall not use this method. Another possibility was presented by Carmichael et al. (1999) who used a tobit estimation and compared its results to estimates provided by OLS.

Their concern was that the results might differ because they included a non-negligible amount of free transfers in their dataset. This group of transfers was expected to drammatically influence the outcome, as it represented censoring but Carmichael et al. (1999) found out, that the results of OLS and tobit estimation did not substantially differ. This finding and the fact that we are not including free transfers in our analysis mean, that we will not use tobit estimation.

Last alternative to OLS is the implementation of the Least Absolute Shrinkage and Selection Operator (LASSO) method, as done by He et al. (2015). Since this is an advanced method and vast majority of the previous research relied on Ordinary Least Squares, we have decided to use solely OLS.

There are several assumptions which are required to hold in order for OLS to produce valid results. Since the relationship between our dependent and independent variables is expected to be linear and we have included a complete range of transfers based on our selection criteria, we believe we only have to consider the risks of heteroscedasticity and perfect collinearity. The homoscedasticity assumption requires the variance of the error term to remain constant in our model. After runing a Breusch-Pagan test, we cannot reject the hypothesis of homoscedasticity. In order to account for the possible presence of heteroscedasticity, we will compute heteroscedasticity-robust standard errors and present them together with OLS standard errors. Finally, the assumption of no perfect collinearity among explanatory variables is confirmed with the use of Variance Inflation Factor (VIF). Value of 10 was chosen as the treshold, as suggested by Wooldridge (2016). The only variables which surpassed this boundary were age and its squared term, which are highly correlated by their nature. No corrective measures were therefore taken.

4.2 Model description

Based on the models used in prior research and on our variables, we use the following regression model to estimate market value of player i:

$$\begin{aligned}\log Value_i =&\beta_0 + \beta_1 PeriodS19_i + \beta_2 PeriodW20_i + \beta_3 Topteam_i \\ &+\beta_4 Age_i + \beta_5 Age_i^2 + \beta_6 Forward_i + \beta_7 Midfielder_i + \beta_8 Foot_i \\ &+\beta_9 Minutes_i + \beta_{10} Goals_i + \beta_{11} Assists_i + \beta_{12} Passespercent_i \\ &+\beta_{13} Natminutes_i + \beta_{14} Natgoals_i + \beta_{15} Euminutes_i \\ &+\beta_{16} Eugoals_i + \beta_{17} Youthminutes_i + \beta_{18} Youthgoals_i + \epsilon_i\end{aligned}$$

In line with existing literature, we use a logarithmic transformation of the dependent variable. A one-unit change in the k-th numerical variable therefore induces a $\beta_k \times 100\%$ change in the explained variable. In the case of dummy variables, the dependent variable changes by $\beta_{dummy} \times 100\%$ should the player possess the respective trait. If not, the dummy variable has no effect on the dependent variable. As already stated in Section 3.2, we expect all coefficients but β_1, β_2 and β_5 to be positive. The effect of β_5 should be negative because it describes the effect of becoming less fit with increasing age. β_1 and β_2 may be of either sign as they only serve to distinguish the three time periods. Because of the purpose of β_1 and β_2, we will not include them in our regression analysis since we are not interested in their estimates.

5 Estimation Results

5.1 Regression Analysis

Table 3 presents results of the estimation with robust standard errors included. A total of 13 coefficients turn out to be significant in our estimation with robust errors. All of those have the expected sign. These results correspond to previous findings in other leagues.

What is interesting is the magnitude of the dummy *Topteam*. It states that the players who joined or left the top three teams of Czech league, were by more than 92% more valuable than players of other clubs. This estimate is highly significant and it therefore shows that there has recently been a huge gap between the best clubs and the rest of the table. We see this as clear evidence that Slavia, Sparta and Plzeň can afford to pay much higher fees for their incoming players. But they are also able to help their players improve up to the level of European competition and consequently sell them for dramatically higher sums of money to foreign clubs.

Similarly to prior research, we can see that *Age* and its squared term are very significant in the Czech league too. Their impact on market values follows an inverted U-shape with peak at 26.75 years. This is slightly higher than 24.16 years reported by Herm et al. (2014) and 25.4 years by Lehmann & Schulze (2008) but explainable. Player requirements of the Czech league are very likely smaller than those of the best European competitions. The players may not be able to sprint for as long and as fast as their foreign counterparts. They may be less skilled in dribbling and other abilities. This helps older players stay comparable in physical abilities to their younger colleagues for a longer time. Moreover, there is also a considerable number of players who used to play in the best foreign leagues but as they get older, they decide to come back to the Czech league before they end their career. These players are still very skilled despite their age, and thus valuable.

What may be a little surprising is that our estimates do not show any sign of differences in values across players' positions. Dummies for both midfielders and forwards are very insignificant. It seems that Czech clubs do not value each position differently. On the other hand, they are willing to pay more money for a left footed player, other things being equal. This premium is quite substantial, over 17% extra (p-value <0.04). This result was expected and comes from the relative scarcity of left footed players.

As expected, *Minutes* are highly significant, and their effect is positive. The benefit of one additional minute may be very small but it can be easily recalculated into effect of a whole 90-minute match. This gives us a total increase of market value by approximately 2.20 per cent after each match. *Goals* and *Assists* also increase players' market value, as expected. Both with high significance. But interestingly, assists are more valued than goals. While an assist increases player's value by 6.35 per cent, scoring a goal results in an increase of only 5.47 per cent. This may possibly be attributed to lower ability of Czech league's players to make a successful pass. While scorers can often find themselves right in front of the goal and score easily, an assist requires a precise pass which is not as simple. This may also mean that goals in foreign leagues are on average scored from a bigger distance and therefore more valued than in the Czech league. We however do not have any data to test this hypothesis. The estimate of *Passespercent* does not bring any surprising results. It shows that player's value increases by approximately 0.84% with each percentage point of successful passes (p-value < 0.001).

Our model also shows that participation in European competitions and in the national team increases market values. While there is no evidence of impact of scoring on the market value, taking part in a match greatly increases the value. Players who are chosen to represent their country in an international match enjoy an increase of their market value by slightly less than 0.12% with each minute played. Over the course of a whole match this raises the player's value by more than 10.6%. This is almost five times as much as the effect of one local match. This result is also very significant. The effect of playing in a European competition is less significant (p-value <0.05) and also slightly smaller. Each 90 minutes played translate into market value increased by 4.28%, which is about two times more than in a match of the

Czech league. It is however important for us to remind ourselves that there is only a small number of these matches played during a season. Even if the effect of one match is high, the overall increase of player's value is limited.

Interestingly, only participating in a youth competition does not bring any growth of market value. On the other hand, scoring a goal drastically increases it. The estimated effect of one scored goal is more than 51% (p-value <0.02). There is a possible explanation for this different behavior compared to the previously discussed competitions. Even though they keep getting more attention, the national youth teams are nowhere near as popular as the adults' national team. Thus, appearing in such game may not be of great benefit. But scoring a goal for the national team draws a lot of attention to the youngster and therefore greatly increases his value. The same reasoning can be used in case of the European youth competitions too.

Table 3: Estimation Results

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
Topteam	0.922595***	0.922595***
	(0.104707)	(0.101455)
Age	0.441882***	0.441882***
	(0.127744)	(0.120244)
Age^2	−0.008259**	−0.008259***
	(0.002479)	(0.002295)
Forward	−0.089042	−0.089042
	(0.109500)	(0.108139)
Midfielder	−0.125107	−0.125107
	(0.098240)	(0.089383)
FootLeft	0.172287*	0.172287*
	(0.082574)	(0.079408)
Minutes	0.000245***	0.000245***
	(0.000059)	(0.000056)

Continued on next page

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
Goals	0.054670***	0.054670***
	(0.014493)	(0.015841)
Assists	0.063538*	0.063538**
	(0.027333)	(0.023575)
Passespercent	0.008409***	0.008409***
	(0.001756)	(0.001897)
Natminutes	0.001187**	0.001187***
	(0.000393)	(0.000295)
Natgoals	−0.046073	−0.046073
	(0.112921)	(0.109445)
Euminutes	0.000475*	0.000475*
	(0.000218)	(0.000233)
Eugoals	0.171908	0.171908
	(0.122562)	(0.135784)
Youthminutes	−0.000302	−0.000302
	(0.000415)	(0.000349)
Youthgoals	0.512644*	0.512644*
	(0.233858)	(0.203235)
Constant	5.460370***	5.460370***
	(1.616684)	(1.556258)
Observations	178	
R^2	0.831	
Adjusted R^2	0.812	
Residual Std. Error	0.481 (df = 159)	
F Statistic	43.342*** (df = 18; 159)	
Note:	*p<0.05; **p<0.01; ***p<0.001	
	For full table, see Appendix.	

5.2 Robustness Checks

5.2.1 Regression on Matches instead of Minutes

We had the possibility to choose between two approaches of implementing the participation in matches into our model. We chose *Minutes* as the primary option because we believe it better describes real participation. Since it has been labeled as the most significant non-dummy variable of our initial model, we want to estimate a model which uses a different metric. The model which uses *Matches* is expected to perform slightly worse but it still should be very similar to our primary model. We present its estimation in Table 4.

In the alternative estimation, all but one previously significant variables keep their significance levels. On top of that there is some weak statistical significance of different positions on the pitch. Both midfielders (p-value <0.06) and forwards (p-value <0.08) are marked as less valuable compared to defenders by 16.8% and 18.7%, respectively. These results are somewhat surprising since they differ from results of prior research, but they are only borderline significant.

Other estimates behave as expected and their magnitudes are comparable to our primary model. We can compare the estimated effect of an additional match and the effect of 90 minutes. These turn out to be very similar. Increase of market value by 2.24% per one local match is slightly higher than 2.20% per 90 minutes. The same goes for European games, where the effect of a match is 4.8% while 90 minutes raise the value by approximately 4.3%. In the case of international competitions, the effect of a match is a bit smaller (9.8% compared to 10.6%). The effect of age peaks at 26.4 years which is slightly lower than 26.75 years from our primary model. All other estimates also change very moderately. Adjusted R-squared drops from 81.2% to 80.9% which confirms our belief, that the primary model should behave better.

Table 4: Regression on Matches

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
Topteam	0.923203***	0.923203***
	(0.106009)	(0.108115)
Age	0.425186**	0.425186***
	(0.129845)	(0.114805)
Age^2	−0.008045**	−0.008045***
	(0.002514)	(0.002162)
Forward	−0.186924	−0.186924
	(0.105591)	(0.104860)
Midfielder	−0.168582	−0.168582
	(0.097828)	(0.088323)
FootLeft	0.190197*	0.190197*
	(0.083070)	(0.079773)
Matches	0.022431***	0.022431***
	(0.005604)	(0.005220)
Goals	0.057224***	0.057224***
	(0.014067)	(0.015596)
Assists	0.063007*	0.063007**
	(0.027485)	(0.023907)
Passespercent	0.007503***	0.007503***
	(0.001884)	(0.002036)
Natmatches	0.098178**	0.098178***
	(0.033396)	(0.026228)
Natgoals	−0.037865	−0.037865
	(0.116910)	(0.113304)

Continued on next page

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
Eumatches	0.048094*	0.048094*
	(0.018563)	(0.020409)
Eugoals	0.096368	0.096368
	(0.126544)	(0.135951)
Youthmatches	−0.037201	−0.037201
	(0.033794)	(0.033344)
Youthgoals	0.586707**	0.586707**
	(0.224642)	(0.219028)
Constant	5.746614***	5.746614***
	(1.647339)	(1.504900)
Observations	178	
R^2	0.829	
Adjusted R^2	0.809	
Residual Std. Error	0.484 (df = 159)	
F Statistic	42.735*** (df = 18; 159)	
Note:	*p<0.05; **p<0.01; ***p<0.001	
	For full table, see Appendix.	

5.2.2 Removing Outliers from the Dependent Variable

We have based our initial estimation on a dataset in which outliers were not treated in any way since we wanted to use as many observations as possible. It is however important to check the results of our model when outliers are removed. Thus, we checked the data for possible occurrence of outliers among market values, as their presence might spoil our estimations. This was done using a boxplot. This method identified a total of 19 outliers which were consequently removed from our dataset, leaving it at 159 observations. Table 5 presents results of an estimation using this reduced dataset. After estimating our model on the new dataset, we can see the results are roughly similar, although there are a few important differences.

Even after removing outliers, estimate of *Topteam* suggests that value of a player transferred by Slavia, Sparta or Plzeň is higher by 87.2% (p-value <0.001) compared to transfers of other teams. This is only about five percentage points lower than the estimate on our initial dataset. This shows that these teams are not only able to make the biggest transfers of the league, but they can also make bigger transfers of average players. The peak effect of Age comes at approximately 27 years, which is slightly higher than 26.75 years from the first model.

What is interesting is the lower significance and magnitude of the estimate of *Goals*, while the opposite happened to the estimate of *Assists*. An additional assist should increase player's value four times as much as an additional goal. This change may be linked to our previous hypothesis, that it is easier for players from the Czech league to score a goal than to make a precise pass to the scorer. The highly valued and highly skilled players, which we removed from our dataset may have been more likely to make assists, which were therefore undervalued.

The estimates of effects of non-local matches are also somewhat different. The estimate of *Natminutes* becomes less significant and smaller, but interestingly, there is a significantly negative effect of *Natgoals* now. This behavior is very strange and after an analysis of our dataset, we found out that we were left with only three players who managed to score a goal for their national team. These players come from Finland and Bahrain, whose national teams are as of April 2020 ranked 58th and 99th in FIFA national teams ranking (for comparison, Czech Republic ranks 45th). We have therefore come to the conclusion that these three players may have underperformed in context of the Czech league, but they were still playing on the level of their national team. Finally, the estimates of *Euminutes* and *Youthgoals* are no longer significant. This may possibly be accounted to the low number of observations left.

There are no other notable differences between this estimation and the primary one. Adjusted R-squared has dropped from 81.2% to 74.7% which shows that our initial model performed slightly better.

Table 5: Regression without Outliers

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
Topteam	0.871776***	0.871776***
	(0.100483)	(0.097909)
Age	0.484179***	0.484179***
	(0.115808)	(0.115855)
Age^2	−0.008949***	−0.008949***
	(0.002246)	(0.002232)
Forward	−0.033428	−0.033428
	(0.101032)	(0.104252)
Midfielder	−0.038982	−0.038982
	(0.091091)	(0.080076)
FootLeft	0.200754*	0.200754*
	(0.078509)	(0.079760)
Minutes	0.000235***	0.000235***
	(0.000057)	(0.000057)
Goals	0.025028	0.025028*
	(0.016251)	(0.012511)
Assist	0.102885***	0.102885***
	(0.030293)	(0.022702)
Passespercent	0.007820***	0.007820***
	(0.001574)	(0.001832)
Natminutes	0.000860	0.000860*
	(0.000553)	(0.000367)

Continued on next page

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
Natgoals	-0.309348^{*}	-0.309348^{**}
	(0.143657)	(0.108096)
Euminutes	0.000350	0.000350
	(0.000230)	(0.000191)
Eugoals	-0.079384	-0.079384
	(0.168115)	(0.107456)
Youthminutes	-0.000056	-0.000056
	(0.000373)	(0.000336)
Youthgoals	0.331969	0.331969
	(0.213081)	(0.172263)
Constant	4.904066^{**}	4.904066^{**}
	(1.468067)	(1.486361)
Observations	159	
R^2	0.776	
Adjusted R^2	0.747	
Residual Std. Error	0.427 (df = 140)	
F Statistic	26.960^{***} (df = 18; 140)	
Note:	*p<0.05; **p<0.01; ***p<0.001	
	For full table, see Appendix.	

Conclusion

The aim of this thesis is to analyze the effect of football players' performance on their market value. While multiple papers on this topic have been conducted on data from the best European leagues, our analysis is based on data from the Czech football league. It builds on existing research in this field and enhances it by looking for factors which have not been regarded before. To the best of our knowledge, this is the first paper which takes into account participation and performance of young players in national youth teams and in European youth competitions and studies their effect on the players' market values. The thesis builds on data from the three most recent transfer periods in order to be able to capture the current situation in the transfer market as accurately as possible.

The analysis of individual performance indicators shows similar results to the previously carried out research. This shows that in evaluating players' performance, the Czech league and transfer market are comparable to their best European counterparts. We found positive and significant effects of the time spent on pitch and of performance in the form of goals, assists and percentage of accurate passes. Besides this, our results indicate that left footed players are valued slightly higher compared to their equally skilled teammates and that appearances in international and European competitions also increase the market value. We also show that young players' value increases by scoring goals in youth international and European matches, an effect which has not been observed before. Last interesting finding is the major difference between the value of players transferred by three front Czech clubs and by the rest of the league. The results have also been found robust to the use of the number of matches played instead of minutes and to removal of outliers.

The biggest contribution of this thesis lies in extension of the existing research on data from a league which is much smaller than the ones studied before. It also contributes just as much by the use of statistics from youth teams and by showing that these statistics play a role in evaluating players.

Potential extensions of this research could make use of quantile regression in order to seek presence of superstars in the Czech league, as discussed in Section 1.2. Besides this, it would also be possible to study the effects of popularity, as suggested by Garcia-del-Barrio & Pujol (2007) or Franck & Nüesch (2012). Moreover, it may be interesting to run an analysis similar to ours with the use of data from upcoming transfer periods. Numerous fields of the economy have recently been struck by the outbreak of COVID-19, and football teams and sports clubs in general are no exception. According to an analysis by Poli et al. (2020), there is expected a 28% loss of the total market value of players from the five biggest European competitions. An analysis of the real consequences of the pandemic would be interesting not only in the context of the Czech league but also other football competitions.

References

Adler, M. (1985). Stardom and talent. *American Economic Review*, *75*(1), 208–212.

Bryson, A., Frick, B., & Simmons, R. (2012). The returns to scarce talent. *Journal of Sports Economics*, *14*(6), 606–628.

Carmichael, F., & Thomas, D. (1993). Bargaining in the transfer market: Theory and evidence. *Applied Economics*, *25*(12), 1467–1476.

Carmichael, F., Forrest, D., & Simmons, R. (1999). The labour market in association football: Who gets transferred and for how much? *Bulletin of Economic Research*, *51*(2), 125–150.

Denny, K., & Sullivan, V. O. (2007). The economic consequences of being left-handed: Some sinister results. *The Journal of Human Resources*, *42*(3), 353–374.

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

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

Frey, B. S. (1998). Superstar museums: An economic analysis. *Journal of Cultural Economics*, *22*(2/3), 113–125.

Frick, B. (2007). The football players' labor market: Empirical evidence from the major European leagues. *Scottish Journal of Political Economy*, *54*(3), 422–446.

Galton, F. (1907). Vox populi. *Nature*, *75*(1949), 450–451.

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

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

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

Lehmann, E. E., & Schulze, G. G. (2008). What does it take to be a star? — the role of performance and the media for German soccer players. *Applied Economics Quarterly*, *54*(1), 59–70.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, *108*(22), 9020–9025.

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

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

Poli, R., Besson, R., & Ravenel, L. (2020). *Pandemic: 28% loss on players' transfer value.* Retrieved April 19, 2020, from https://football-observatory.com/IMG/sites/b5wp/2019/wp289/en/

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

Rosen, S. (1981). The economics of superstars. *American Economic Review*, *71*(5), 845–858.

Ruijg, J., & van Ophem, H. (2014). Determinants of football transfers. *Applied Economics Letters*, *22*(1), 12–19.

Speight, A., & Thomas, D. (1997). Arbitrator decision-making in the transfer market: An empirical analysis. *Scottish Journal of Political Economy*, *44*(2), 198–215.

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

Wicker, P., Weimar, D., Prinz, J., & Upmann, T. (2013). No pain, no gain: Effort and productivity in professional soccer. *International journal of sport finance*, *8*(2), 124–139.

Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, *18*(2), 107–126.

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

Appendix

Table 6: Full Results of Primary Estimation

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
PeriodS19	-0.028287	-0.028287
	(0.088923)	(0.077389)
PeriodW20	0.193233	0.193233^{*}
	(0.100450)	(0.096408)
Topteam	0.922595^{***}	0.922595^{***}
	(0.104707)	(0.101455)
Age	0.441882^{***}	0.441882^{***}
	(0.127744)	(0.120244)
Age^2	-0.008259^{**}	-0.008259^{***}
	(0.002479)	(0.002295)
Forward	-0.089042	-0.089042
	(0.109500)	(0.108139)
Midfielder	-0.125107	-0.125107
	(0.098240)	(0.089383)
FootLeft	0.172287^{*}	0.172287^{*}
	(0.082574)	(0.079408)
Minutes	0.000245^{***}	0.000245^{***}
	(0.000059)	(0.000056)
Goals	0.054670^{***}	0.054670^{***}
	(0.014493)	(0.015841)
Assist	0.063538^{*}	0.063538^{**}
	(0.027333)	(0.023575)

Continued on next page

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
Passespercent	0.008409***	0.008409***
	(0.001756)	(0.001897)
Natminutes	0.001187**	0.001187***
	(0.000393)	(0.000295)
Natgoals	−0.046073	−0.046073
	(0.112921)	(0.109445)
Euminutes	0.000475*	0.000475*
	(0.000218)	(0.000233)
Eugoals	0.171908	0.171908
	(0.122562)	(0.135784)
Youthminutes	−0.000302	−0.000302
	(0.000415)	(0.000349)
Youthgoals	0.512644*	0.512644*
	(0.233858)	(0.203235)
Constant	5.460370***	5.460370***
	(1.616684)	(1.556258)
Observations	178	
R^2	0.831	
Adjusted R^2	0.812	
Residual Std. Error	0.481 (df = 159)	
F Statistic	43.342*** (df = 18; 159)	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 7: Full Results of Regression on Matches

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
PeriodS19	−0.037291	−0.037291
	(0.089352)	(0.076894)
PeriodW20	0.199960*	0.199960*
	(0.101208)	(0.097739)
Topteam	0.923203***	0.923203***
	(0.106009)	(0.108115)
Age	0.425186**	0.425186***
	(0.129845)	(0.114805)
Age^2	−0.008045**	−0.008045***
	(0.002514)	(0.002162)
Forward	−0.186924	−0.186924
	(0.105591)	(0.104860)
Midfielder	−0.168582	−0.168582
	(0.097828)	(0.088323)
FootLeft	0.190197*	0.190197*
	(0.083070)	(0.079773)
Matches	0.022431***	0.022431***
	(0.005604)	(0.005220)
Goals	0.057224***	0.057224***
	(0.014067)	(0.015596)
Assist	0.063007*	0.063007**
	(0.027485)	(0.023907)
Passespercent	0.007503***	0.007503***
	(0.001884)	(0.002036)

Continued on next page

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
Natmatches	0.098178**	0.098178***
	(0.033396)	(0.026228)
Natgoals	−0.037865	−0.037865
	(0.116910)	(0.113304)
Eumatches	0.048094*	0.048094*
	(0.018563)	(0.020409)
Eugoals	0.096368	0.096368
	(0.126544)	(0.135951)
Youthmatches	−0.037201	−0.037201
	(0.033794)	(0.033344)
Youthgoals	0.586707**	0.586707**
	(0.224642)	(0.219028)
Constant	5.746614***	5.746614***
	(1.647339)	(1.504900)
Observations	178	
R^2	0.829	
Adjusted R^2	0.809	
Residual Std. Error	0.484 (df = 159)	
F Statistic	42.735*** (df = 18; 159)	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 8: Full Results of Regression without Outliers

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
PeriodS19	−0.042014	−0.042014
	(0.082738)	(0.074045)
PeriodW20	0.082615	0.082615
	(0.091997)	(0.087995)
Topteam	0.871776***	0.871776***
	(0.100483)	(0.097909)
Age	0.484179***	0.484179***
	(0.115808)	(0.115855)
Age^2	−0.008949***	−0.008949***
	(0.002246)	(0.002232)
Forward	−0.033428	−0.033428
	(0.101032)	(0.104252)
Midfielder	−0.038982	−0.038982
	(0.091091)	(0.080076)
FootLeft	0.200754*	0.200754*
	(0.078509)	(0.079760)
Minutes	0.000235***	0.000235***
	(0.000057)	(0.000057)
Goals	0.025028	0.025028*
	(0.016251)	(0.012511)
Assist	0.102885***	0.102885***
	(0.030293)	(0.022702)
Passespercent	0.007820***	0.007820***
	(0.001574)	(0.001832)

Continued on next page

	Dependent variable:	
	log(Value)	
	OLS	Robust SE
Natminutes	0.000860	0.000860*
	(0.000553)	(0.000367)
Natgoals	−0.309348*	−0.309348**
	(0.143657)	(0.108096)
Euminutes	0.000350	0.000350
	(0.000230)	(0.000191)
Eugoals	−0.079384	−0.079384
	(0.168115)	(0.107456)
Youthminutes	−0.000056	−0.000056
	(0.000373)	(0.000336)
Youthgoals	0.331969	0.331969
	(0.213081)	(0.172263)
Constant	4.904066**	4.904066**
	(1.468067)	(1.486361)
Observations	159	
R^2	0.776	
Adjusted R^2	0.747	
Residual Std. Error	0.427 (df = 140)	
F Statistic	26.960*** (df = 18; 140)	
Note:	*p<0.05; **p<0.01; ***p<0.001	