CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

How team strategy in football influences players' market value

Bachelor's thesis

Author: Milan Knapp

Study program: Economics and Finance

Supervisor: Petr Pleticha, M.Sc.

Year of defense: 2020

Declaration of Authorship

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

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

Prague, April 20, 2020

Milan Knapp

Abstract

This work investigates the effect of team strategy in professional football on the value of players on the transfer market. The research is conducted on player-level data from the English Premier League, German Bundesliga, Spanish La Liga, Italian Serie A and French Ligue 1 in season 2018/2019. Price is explained by player-related attributes like age and height, performance data and by team-related statistics. We are specifically interested in the significance of team data and their relationship with playing strategy. Results of the work show strong evidence that different playing strategies influence players' value which makes optimization for maximal value of the team squad possible.

JEL Classification	C12, C01, Z20, C51
Keywords	Football, Players' Value, Team Strategy, Football Transfer Market

Title	How team strategy in football influences players' market value

Author's e-mail	milan.knapp10@gmail.com
Supervisor's e-mail	petr.pleticha@cerge-ei.cz

Abstrakt

Tato práce zkoumá efekt týmové strategie v profesionálním fotbale na cenu hráčů na přestupovém trhu. K výzkumu jsou použita data o hráčích z anglické Premier League, německé Bundelisgy, španělské La Ligy, italské Serie A a francouzské Ligue 1 za sezónu 2018/2019. Cena na přestupovém trhu je vysvětlena pomocí jednotlivvých hráčských atributů jako věk, či výška, výkonnů na hřišti v danném období a také pomocí týmových statistik. Týmové statistiky jsou hlavními zkoumanými proměnnými této práce a je testován jejich význam. Výsledky práce ukázaly, že herní strategie ovlivňuje cenu hráčů, toto zjištění umožňuje týmům optimalizovat svou strategii na maximální hodnotu týmu.

Klasifikace JEL C12, C01, Z20, C51
Klíčová slova Fotbal, Hodnota hráčů, Týmová Strategie, Fotbalový přestupový trh

Název práce Jak týmová strategie ovlivňuje cenu hráčů na přestupovém trhu

E-mail autora milan.knapp10@gmail.com
E-mail vedoucího práce petr.pleticha@cerge-ei.cz

Acknowledgments

I would like to express my gratitude especially to Petr Pleticha, M.Sc. for guiding me along the way and providing me with valuable insights on how to conduct the research properly, how to build the model and what to focus on. Furthermore, I would like to show my greatest appreciation for Petr Kasalický for his help with collecting the data. I owe my deepest gratitude to Eliska Sloukova and Thao Trang Do for answering all the questions (no matter how stupid the questions were) with an infinite patience. Lastly, many thanks go to Bea Davidova and Petr Havlicek for helping me to come up with the thesis topic.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record

Knapp, Milan: *How team strategy in football influences players' market value.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2020, pages 57. Advisor: Petr Pleticha, M.Sc.

Contents

List of Tables

Acronyms

FIFA Federation Internationale de Football Association

UEFA Union of European Football Associations

CL Champions League

EL European League

Bachelor's Thesis Proposal

Author	Milan Knapp
Supervisor	Petr Pleticha, M.Sc.
Proposed topic	How team strategy in football influences players' market value

Motivation Football clubs in Europe are not playing for fun but mainly for a profit. Clubs are spending hundreds of million euros on the football transfer market every year to get the best players available. Some teams are also investing a lot of money into finding new talents all around the world in hope to get them to their team, make them top star players and sell them on the market. It is indisputable that value of the football players depends on their personal attributes and performance, but the scope of this text is to find if and how the strategy of the whole team influence players' value. The main hypothesis is that offensive players are valued higher on the market, so the teams that play offensive football are more likely to raise higher valued football players than the teams playing defensively.

Hypotheses

Hypothesis #1: Does the team strategy influence the value of players on the market?

Hypothesis #2: If the strategy influences the value, is it right to conclude that offensive strategy leads to higher valued players?

Methodology The first part of the thesis will be focused on measuring offensiveness of teams' strategies depending on teams' statistics like shots, corners, yellow cards, etc. When the definition of the strategy will be specified, the "strategy variable" will be used in a model estimated using OLS regression with Player's value as a dependent variable and a significance of this variable will be tested. The research will be done on the case of the Premier League since it is the most prestigious league in Europe. The data used in this work will be obtained from publicly available sources like the official website of the league, football statistics databases and others. Players' market value will be obtained from transfermarkt.de since according to research wisdom of crowds estimates the transfer fees well. (Herm et al., 2014)

Expected Contribution The main contribution of this work is to find evidence that playing offensive football leads to higher revenue from selling players on the European football transfer market. The effect of players' personal attributes, performance as well as effect of market information on players' value has been examined (He et al., 2015) also the influence of the league in which the player operates has been tested (Ante, 2019) but the effect of team strategy on players' value has been so far neglected. Finding some evidence that team strategy influences players' value can lead to further research and potentially to adjustments in the teams investing and playing strategies.

Core bibliography

ANTE, L. (2019). Determinants of Transfers Fees: Evidence from the Five Major European Football Leagues.

He, M., CACHUCHO, R., & KNOBBE, A. J. (2015). Football Player's Performance and Market Value. In MLSA@ PKDD/ECML (pp. 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.

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

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

Author Supervisor

Chapter 1

Introduction

Professional football (soccer in the US English) is the most popular sport in the world, with about 4 billion fans (Sawe 2018) Professional football became a huge multi-billion industry (Ante 2019) making money on selling tv licenses, club merchandise for fans, match tickets, prize money for winning a league or a cup, and others. Clubs and sponsors pay the best players in the world millions of euros per year (BoE 2019). These circumstances, especially in the last ten years, led the transfer market with the players to grow to unprecedented size. While the most expensive transfer in 2000 was Real Madrid buying Luis Figo for £37m from Barcelona, currently in February 2020, the most expensive transfer is Neymar joining PSG for £198m in 2017 according to Transfermarkt. There are clubs famous for their work with young talents and making a profit on selling them to the most prestigious clubs in Europe. Especially famous for that is Ajax Amsterdam, which only in summer 2019 earned over €200m on selling their stars. Considering the prices on the transfer market, wisely buying and selling the players became crucial for the financial stability of any club. For some, money from having the best scouts, acquiring emerging stars, and selling them later is not the primary source of financing, but they still want their players to at least keep their value and increase it if possible. While the transfer fees are not always shared publically, we usually have reliable approximate information about the transfer. The issue is that most of the players are not changing clubs every year, and the transfer fees which were paid for them at a time of the transfer are not reflecting the changes in performance that happened after that. Market values are used in order to estimate the hypothetical transfer fee. The conceptual difference between the two is that the transfer fee is what a club pays for the player, and market value is an estimate on how much the clubs should be willing to pay for the player. The transfer fee may differ from the estimated market value for reasons including the remaining time till the end of the contract, strategic importance of the player, or the differences in bargaining power on each side (Herm *et al.* 2014). Despite the conceptual differences, market value estimates still serve as a good proxy, explaining around 90% of the variance in transfer fees (Herm *et al.* 2014).

The objective of this work is to examine if the team playing strategy can influence the value of its players on the transfer market. The hypothesis comes from the fact that (1) the offensive players are the most expensive on the transfer market (there is no defender in the top 10 most expensive transfers), and (2) playing offensive football is found more attractive by most of the fans and experts. The secondary goal is a replication of previous results on a new data and describtion of a possible shift in the preferences on transfer market. The scope of the work is to use existing research on the drivers of players' value (e.g., Ante (2019); Müller *et al.* (2017)) to estimate players' value using OLS regression and to combine it with the research on team strategy (e.g., Santos (2014); Guedes & Machado (2002)) by including set of team variables into the model and testing what influence on the market value these variables have, keeping other attributes fixed. The research is conducted on a dataset including all field players that played in one of the top five European leagues during the season 2018/2019.

The thesis is structured into chapters as follows: Chapter 2 introduces the key studies in one of the three fields related to this research. In the first part, the literature on the primary value drivers in today's transfer market is reviewed. The second part focuses on how the market value can be estimated because there is no single estimation procedure established, and two different approaches are valid (Müller *et al.* 2017). One is a crowd-based estimation on websites like Transfermarkt (transfermarkt.com), and the other is a data-driven approach. In the last part of the chapter, the literature on the team strategy is reviewed with a focus on how to measure the offensiveness of the strategy. Chapter 3 describes the collected variables in the dataset, how they were collected, and provides descriptive statistics of the data. Chapter 4 si dedicated to methodology description and model building. The last part of the work presents the results of the regression and a conclusion about the influence of a team strategy. Tables with all the results can be found in the Appendix.

Chapter 2

Literature Overview

Existing research focuses mainly on defining what the main variables influencing the player's value are and how the market value can be estimated to predict the actual transfer fees (Carmichael & Thomas 1993; Müller *et al.* 2017; Ante 2019). The focus on finding the main value drivers is understandable because it helps the managers to choose the players with the best return on the investment in terms of game results and financial results. The team strategy of the club in which the player currently works was, to our knowledge, so far neglected in the research as a possible value driver. We think that the reason may be that because the necessary data became available recently, and the clubs so far focused more on using this data to get the edge over others in identifying the players with the highest potential. Focus on finding what increases the value of their players when they decide to sell them was not of the highest importance.

2.1 Main drivers of players' market value

There are many player attributes identified by academic research as influencing players' value. Ante (2019) proposes a definition of dividing these attributes into three categories. The first category is defined as (1) Personal characteristics not influenced by actual performance. That means attributes that a player has or miss, and whose change is not dependent on the player's on-pitch results, for example, age or height. This category reveals what the highly valued characteristics on the market in the time of the research are, and it can serve as an indicator of a change in priorities over time as football is slowly evolving. By comparing research results conducted in different time periods, these changes of behavior on the transfer market can be documented. One of the examples of how football is changing is that the top football players of today are, on average, more than a year and a half older than top players thirty years ago (Kalen *et al.* 2019).

The next category is (2) Player performance on the pitch. This category covers the performance-related metrics, and they can differ for each player position as the expectations are different for defensive and offensive roles (Müller *et al.* 2017). The last type of metrics we may want to account for is (3) Player popularity among people. It accounts for the possibility that this metric unrelated to the actual performance of a player may influence his transfer fee. Franck & Nueesch (2008) tests this hypothesis on top players in Bundesliga, and the findings confirmed that popularity among people influences the transfer fee.

2.1.1 Personal characteristics

Multiple studies agree that age is one of the main variables influencing market value. (Carmichael & Thomas 1993; Ante 2019; Müller *et al.* 2017; Ruijg & van Ophem 2015). The age is usually included in the model in its quadratic form to take into account the non-linear nature of the variable. It is a good proxy for potential and experience (Carmichael & Thomas 1993). At the beginning of the career, the player is inexperienced, but the potential to grow is at its peak. With each new season, the player gets more experience, and his value increases. The negative effect of age is that the potential to grow lowers with each season, and the player is one year closer to the end of the career, which usually does not exceed twenty years.

When the player is young, the gain of experience causes substantial growth in value, and the negative effect of aging is not that important. Typically, the players reach their maximum in their mid-twenties, and from this point, the effect of aging becomes stronger than the gain of new experience. Academical research agrees with the described process (Müller *et al.* 2017; Ruijg & van Ophem 2015). To our best knowledge, other metrics were not previously used to take experience into account, and every time age squared was used, a negative correlation between age squared and market values was confirmed. (Carmichael & Thomas 1993; Carmichael & Forrest 1999; Müller *et al.* 2017; Ruijg & van Ophem 2015).

The height of the player is another personal characteristic influencing the value. Tall players are in general preferred over short and average ones, with the other variables being the same (Carmichael & Forrest 1999). The reason for this is that tall players provide the coach with more options in strategy, and they are beneficial in more standard on-pitch situations (e.g., both offensive and defensive corner kicks). The research also found evidence that these players are more likely to score goals from the air, win aerial tackles and prevent the other team from scoring (Dobson & Gerrard 1999; Ante 2019). It is important to mention that the benefit of height is not the same for all playing positions as not all roles often get into situations when the height provides significant benefit. (Ante 2019).

Lastly, the country or a continent of origin plays a role in players' valuation. Some of the previous research included this information as a variable in the model, and both Feess & Muehlheusser (2003) and Ante (2019) found evidence that keeping other variables fixed, a continent of origin may influence the value. Results in Ante (2019) show that being from South America has a positive effect on market value, but being from Asia may have a negative effect. This effect can be explained based on assumptions about the general population of these countries.

The effect may come from the fact that in some parts of the world, football is more incorporated in the national culture (like in Brazil, for example) than in the others, and more top-class football players come from these parts of the world. The countries consequently achieve better results in international competitions and are stereotypically seen as "Home of football". When the players from these states are compared with others at the same performance level, the perceived value may be higher only due to the stereotypical assumptions that someone from South America will be a better player than someone from Asia. Another explanation may be that the difference does not come from any stereotype but from the fact that the state of origin may be a proxy for a playing style the player has, and this style may be valued higher on the current transfer market.

2.1.2 Performance

Personal characteristics are similar to millions of football players around the world. What makes clubs willing to pay millions of euros for a player is his performance on the pitch as it has a direct effect on the match results and, therefore, on the financial results of the club itself.

Each club has several options to choose from when it comes to a decision who should play in the match. It is logical to assume that the better performing players will get more field time in comparison with their worse performing teammates. When we assume that field time is a proxy for performance, the result should be a higher valuation of these players. This assumption was confirmed multiple times by previous research (Carmichael & Thomas 1993; Bryson *et al.* 2009; Ruijg & van Ophem 2015; Müller *et al.* 2017; Ante 2019).

There are a few different methods on how to account for playing time, and each of them expresses the same information in a slightly different way. Carmichael & Thomas (1993) uses the number of match appearances during the season and suggests that it serves well as a proxy for the current form and general fitness of the player. Bryson *et al.* (2009) divides the number of appearances into two variables (1) being in starting eleven and (2) going to the match from the bend. The approach of Ruijg & van Ophem (2015) is similar to the one in Bryson *et al.* (2009) but uses the form of a ratio between substitutes starts and the total number of appearances. Ruijg & van Ophem (2015) research also adds an average of minutes played in a match as a variable of the field time. The number of minutes is also used in Müller *et al.* (2017) and Ante (2019). Lastly, Bryson *et al.* (2009) takes into account that some of the games are more important than others and uses a dummy variable for appearances in the Champions League and UEFA Cup (predecessor of Europe League). These two cup competitions are very prestigious, and only the best teams across Europe participate. As a result, starting in one of these can be a sign of exceptional performance.

In this study, We use the number of minutes of playing time because it captures most accurately how much time the player spent on the pitch. We see a disadvantage of using the number of starts in starting eleven or from a bench in the fact that this approach cannot account for differences between players regularly playing the whole 90 minutes and players who are often substituted. The information about starts in CL and EL is included in the form of two dummy variables to measure the effect of participating in these international matches.

Including the right performance metrics is not as easy as it may appear to the casual spectator. The number of goals scored is the first variable that comes to mind, and to our best knowledge; it is present in all the related research. It has a proven positive effect on the market value (Carmichael & Thomas 1993; Dobson & Gerrard 1999; Müller *et al.* 2017).The problem with the number of goals is that this metric is favoring strikers and offensive midfielders because they have attacking roles on the pitch. (Ante 2019). Defensive players may score goals as well, but it is not their primary role on the pitch, and as a result, measuring their performance by the number of goals scored would be unfair.

Assists and shots are other variables used in the models as well (Müller *et al.* 2017; Ante 2019), and they help us to better account for the effectivity of a player in a scoring opportunity and also for his propensity to help other team players to score a goal. Unfortunately, the problem remains the same. These metrics are also related to the offensive part of the game, and defenders may be undervalued when only these metrics are used.

Other metrics that are used in the academical research for measuring the performance more accurately are passes and a percentage of successful passes (Müller *et al.* 2017). While the number of passes may have a positive correlation with being an offensive player, the percentage of successful passes is crucial for defenders as well, since one imprecise pass may lead to conceding a goal and, as a consequence, losing the game. It is an essential variable for offensive roles as well because a high percentage of lost balls from inaccurate passes will significantly lower the chance of the team to score. Due to the differences in defensive and offensive roles, midfielders and strikers are more prone to risky passes because one good pass may result in scoring a goal. Defenders are more likely to choose less risky options because of the risk of conceding a goal as described earlier.

Lastly, discipline and skill to provoke the opponent's indisciplined behavior is another valued aspect of players' performance in the field. To account for this, the number of yellow and red cards is used in the literature (Kiefer 2012; Ruijg & van Ophem 2015; Müller *et al.* 2017). Müller *et al.* (2017) finds yellow cards to be a significant variable negatively influencing the estimated value. It may be a signal of a player's unpredictable behavior in stressful situations. Fouls committed and suffered are often included to account for the same thing (Müller *et al.* 2017; Kiefer 2012; He *et al.* 2015). While fouls committed account for the frequency of how often the player breaks the rules himself, the fouls suffered measure the ability to force the opponents to break a rule. Results from Ante (2019) and Kiefer (2012) agree that fouls suffered may be positively correlated with the players' value, and the same results apply for fouls committed as well. Nevertheless, none of the research shown strong evidence for that.

2.1.3 Popularity

The last of the three categories influencing the transfer fee is independent on the actual performance and suggests that more popular players will be valuated higher while keeping other variables fixed. It may be explained by the fact that buying a so-called superstar may increase the popularity of the club around the world, increase the sales of tickets and merchandise (Ante 2019). Adler (1985) concludes that to become a star, superior talent over others is not necessary, but it can be explained by better public knowledge of the performer. This theory was tested if it applies to the football environment in Franck & Nueesch (2008) on all players playing in Bundesliga in the season 2004/2005. It was found that the player's popularity is a significant predictor of the stars' market values. Evidence for Rosen (1981) stating that the emergence of superstars relies primarily on superiority in observable talent was not found as Rosen (1981) and Adler (1985) are two competing theories on superstar formation.

Due to the rise of the Internet and social media, current literature uses different methods to account for popularity than the ones used in the past. Franck & Nueesch (2008) uses the existence of a dedicated player's homepage, the number of Google hits, and name appearances in the press to measure the popularity of an individual player. Kiefer (2012) takes the emergence of new social media channels into account by using the number of likes on social site Facebook as a variable of player's popularity. Müller *et al.* (2017) uses a wide variety of measures including Wikipedia page views, Google hits, press citations, and YouTube videos to account for popularity among different age and social groups, that comes from the fact that not all communication channels are used across population equally. Ante (2019) included the number of followers on all leading social media platforms, including Facebook, Instagram, and Twitter, to account for popularity.

2.2 How to estimate the market value

Considering the size of the football industry, it is surprising how long the industry ignored possibilities that come with data. Professional football was called "the least statistical" of all major sports by the New York Times in 2010 (Müller *et al.* 2017). For a long time, market values have been estimated by managers and football experts. Value estimation has changed dramatically in the last few years, with much more detailed performance statistics available and with the growth of the Internet.

The growth of the Internet makes the crowd-based estimation possible, which is nowadays the number one source of value estimates (Müller *et al.* 2017; Herm *et al.* 2014). The biggest crowdsourcing website Transfermarkt (www.transfermarkt.com) is a platform where millions of football fans are estimating the market values of professional football players. To ensure that the estimates do not suffer from people biasedness, not every vote is equal on Transfermarkt. There is a system of selected judges taking place, and they judge the estimates of others and have the final say in deciding about the given estimate. The process is nicely described in Herm *et al.* (2014).

These crowd-based estimates have been found very accurate and more precise than estimates created by football experts (Herm *et al.* 2014). According to Peeters (2018), Transfermarkt estimates also predict the performance of the national team better than the more traditional measures like FIFA ranking or ELO rating. The Transfermarkt estimates are also commonly used in media and during the transfer fees negotiations. Scientific studies Franck & Nüesch (2012); He *et al.* (2015) use Transfermarkt market values as a foundation for their research.

While crowd-based estimation is nowadays the leading method of value estimation, Academic research questions its efficiency (Müller *et al.* 2017). One of the possible downsides of crowd-based estimation is the described system of judges because the question "Who judges the judges' unbiasedness?" arrises (Müller *et al.* 2017). The availability of data allows for precise data-driven estimation. There are research companies like CIES Football observatory (https://football-observatory.com) providing their econometric estimates online and growing their popularity.

Müller *et al.* (2017) has found that data-driven estimation has good results estimating the actual transfer fees. In comparison with the Transfermarkt estimation, the data-driven results were slightly less efficient, but the difference was not statistically significant. Müller *et al.* (2017) also states that data-driven estimation worked better for low- to medium-priced players, while crowd estimates were more accurate for the high-priced players. Lastly, Müller *et al.* (2017) concludes that both methods perform very well in estimating the transfer fees, but with more data available, the data-driven estimation will become more efficient in the future as it overcomes some of the problems of crowd-based estimates such as possible biasedness of the crowd.

2.3 Strategy in professional football

Strategy in football is a broad topic; it may include player formations, training approach, the process of selecting the right players, and others. In this work, we are interested in how offensively or defensively the team plays during the matches. The problem with the team strategy is the difficulty with the proper measurement as it is a complex issue. Team strategy depends not only on the decisions of the manager but also on the ability of the players to follow his instructions. Another issue is that the strategy can change at any time during the season or even during the match.

Palomino *et al.* (1998) found evidence that the strategy is changing during the match, and these changes depend on the score. When the match is tied, both teams will attack in order to score a goal and take the lead. When one of the team is ahead, it usually starts to play more defensively, and on the contrary, the losing team will attack more to increase its chances of scoring an equalizer. This change in the behavior will lead to a higher chance of scoring on both sides because the losing team puts more effort into the offensive phase, and mistakes in defending are more likely to occur, which creates new scoring opportunities for the leading team (Santos 2014). The strategy is not changing only during the game, but it will differ from game to game depending on the strategy and strength of the opponent (Santos 2014).

The most challenging part of measuring the effect of the strategy is the issue of differentiating between the influence of the strategy from the differences caused by the quality of the two teams (Santos 2014). Santos (2014) controls for the asymmetry among the teams because the effect of different strengths may be much more significant than the effect of a manager's tactical decisions. Santos (2014) divides the teams into three groups by quality, which was represented by the average number of points won per game, and he also controls for the effect of playing on the home stadium and the opponents' stadium. Santos (2014) found evidence that teams following an offensive strategy are more likely to both score and receive goals, and they are less likely to commit fouls and receive yellow cards. We use findings of Santos (2014) and accounts for the team quality by including the average points gained per league match into the model.

While Santos (2014) focuses on the strategy changes during the match, Guedes & Machado (2002) and Moschini (2010) focus on how is the team strategy changing in the long term. Both papers investigate if the team acquired more offensive strategies after the introduction of "3-point rule" by FIFA in 1995 because it was designed in order to make football matches more attractive. While both mentioned papers conclude that it leads to more goals being scored and the probability of drawn matches decreased by 16 percent (Moschini 2010), data from Portuguese league have shown that mainly underdogs were incentivized to adopt a significantly more offensive strategy, while mainly the top tier teams where scoring those extra goals Guedes & Machado (2002).

Chapter 3

Data

3.1 Data description

This research uses two different datasets as a primary source. The *Dataset Players* contains information about each player who played in Premier League, Bundesliga, Serie A, La Liga or Ligue 1 during the season 2018/2019. In *Dataset Teams* there are statistics for each team playing in one of the leagues in that season. The selected leagues are the five most prestigious national leagues, and almost all the best players play in of these five. Most of the data were collected through API-Football which is a service hosted on Rapid API - a marketplace where thousands of API's are being offered. API-Football offers wide choice of football-related data in an easily accessible way. Python code and Jupyter notebook was used in order to collect the data. The market value estimates were scrapped from the Transfermarkt website for each player and completed manually for those players who were not found automatically.

3.1.1 Dataset Players

Data for all players in the top five leagues were collected. Together, there is information about 3112 players from which 665 are strikers, 1071 are midfielders, 1028 are defenders, and 349 are goalkeepers. For this research, the goalkeepers are excluded from the regression. The reason is that the performance of goalkeepers cannot be measured using the same metrics as for the field players, and they may cause bias in our data. Excluding goalkeepers from the analysis is a common practice in existing research (Bryson *et al.* 2009; Müller *et al.* 2017).

The players who did not have a value estimate given by Transfermarkt were also dropped out. The reason for a missing estimate was usually the age of the player. He was either too young and have not been evaluated by the Transfermarkt community yet, or he retired that season and did not have any value estimate assigned because he was no longer active. The number of observations in the regressed model is 2370. The dataset contains 20 variables for each player divided into three categories. The dependent variable Market Value was collected twice - at the beginning of the season, and after its end. Collecting the market value estimate makes measuring the influence of performance in only that one season possible in case it would be necessary. See table 3.1 for a complete list of variables.

Table 3.1: List of collected variables

Dependent variable: *Market Value*		
Independent variables		
Personal characteristics	Performance variables	
Age	*Minutes played*	*Passes*
Height	*Goals scored*	*Passes accuracy*
Position	*Assists*	*Fouls committed*
Team	*Yellow cards*	*Fouls suffered*
Continent of origin	*Red cards*	*Passes*
	Shots	*Appearance in CL*
	Shots accuracy	*Appearance in EL*

Information about *player's team* was collected in order to connect the data with the second dataset. As was previously mentioned, the effect of popularity on players' value was investigated in the past as well, but no variable measuring popularity is included. Collecting relevant data for popularity was left out of the scope of this work because the primary goal is to identify the effect of a team strategy, which is presumably uncorrelated with the popularity of individual players and, therefore, leaving the popularity out of the model does not cause any bias. If the scope of the work would be to estimate the players' value, omitting the information about popularity may lead to lower efficiency of the model. Summary statistics data can be found in table 3.2.

Football is a cooperative game where each of the eleven players has a different role, but the main differences are among strikers, midfielders, and defenders. We can expect that the statistics for each of the positions will differ and it may be useful to split the dataset by position for some of the analysis. Summary data for each position can be found in the appendix.

Table 3.2: Summary statistics for all players

Player statistics				
Variables	mean	sd	min	max
Value after season	10.98	17.64	.05	200
Age	26.58	4.20	18	42
Height	181.66	6.22	162	201
Hours played	21.90	16.93	0	57
Goals	1.96	3.51	0	36
Assists	1.35	2.07	0	15
Yellow cards	2.84	2.89	0	16
Red cards	.074	.281	0	3
Shots	18.71	23.14	0	177
Shots accuracy	25.53	20.99	0	100
Passes	494.62	477.87	0	2768
Passes accuracy	68.39	24.42	0	100
Fouls committed	17.99	15.84	0	98
Fouls suffered	16.94	17.55	0	112
Played in CL	.159	.365	0	1
Played in EL	.153	.360	0	1
Observations	2370			

3.1.2 Dataset Teams

In the second dataset, We have included all the variables that may be useful for testing the effect of team strategy on the dependent variable. Ten variables were collected for 98 teams from five leagues. All the collected variables can be found in table 3.3, together with the related descriptive statistics. All the team statistics are averages for one match.

Most of the football experts and fans believe that each league is different and has its specifics. Two league specifics are discussed often. Firstly, something that can be called "Style of football", which is pretty much a general term for everything including what type of players are preferred in the given league (e.g., if the employers prefer technical skills over strength), in what formations teams play, and others.

The second specific is team strategy. There is a belief that not only each team has its strategy, but also the whole league tends to be played more offensively or defensively. To examine if these differences among leagues exist or it is just a stereotype, statistics for each league were investigated. See Appendix for the tables with per league summary statistics. The evidence was found that

Table 3.3: Summary statistics for Team dataset

Team statistics				
Variables (Avg. per match)	mean	sd	min	max
Points	1.37	.457	.42	2.58
Goals scored	1.38	.444	.58	2.76
Goals received	1.38	.346	.58	2.13
Yellow cards	2.37	.570	1.29	3.97
Red cards	.061	.041	0	.21
Fouls committed	14.39	2.371	9.5	20.66
Fouls suffered	13.66	2.610	9	22.61
Shots	14.93	4.164	9.24	29.05
Shots accurracy	.34	.0338	.27	.43
Passes	344.47	90.706	158	622
Observations	98			

each league is different. For example, while Bundesliga, which has the highest average number of goals scored, has the average per match close to 3.2, the league with the least goals scored, which is League 1, has less than 2.6 goals per match. To take these differences into account, the dummy variable with information about the league will be included in the model.

Chapter 4

Methodology

4.1 OLS regression

For the empirical part of the thesis, the standard OLS regression is used to estimate the impact of team strategy on the values of individual players. Linear regression was used multiple times in the previous research for estimating the players' value (Carmichael & Thomas 1993; Ruijg & van Ophem 2015; Ante 2019). Other methods have been used as well, but there is no need for these advanced tools in this paper as they were used for particular reasons in each case. He *et al.* (2015) used Lasso regression to conduct the research due to the lack of observations in comparison with the number of independent variables. Lastly, Müller *et al.* (2017) uses machine learning techniques to estimate the players' value. He trained the model first on transfermarkt data, and then he was using it for tests if the data-driven estimation is more efficient than the crowd-based estimates. We have enough observations and the goal of the work is not to train a model for predicting values, which makes OLS model the best suitable option.

4.2 Model

The unit of observation in this model is an individual player who played in one of the top five football leagues during the season 2018/2019. The dependent variable in the main model is the natural logarithm of the player's value estimate from Transfermarkt, which was estimated in summer 2019 after the season ended. The natural logarithm of the variable is used because the data is highly skewed, which would cause bias in the regression results. Therefore, it is a log-level model, and the effect of independent variables will be interpreted as percentage change.

As the first step in building the model, all the player characteristics and performance variables were included in the model without any data about team strategy to test if the regression results will correspond with the existing literature. Age, height, the continent of origin, and a dummy variable for playing position were included to describe the player itself. Age was included in both linear and squared form to allow for a non-linear relationship of age and the dependent variable. Variable for a league in which the player participates was also included. As performance variables, the commonly used metrics are present - a number of hours played, goals, assists, yellow and red cards, fouls committed and suffered, and a dummy for playing in CL or EL.

We also use the following additional performance variables to test if they have any effect on the player's value - shots, shots accuracy, passes, and passes accuracy. To our best knowledge, shots, and shots accuracy have not been previously tested. The number of shots may increase the value of a particular player, especially of an offensive midfielder or a striker, because it expresses how often he can get into scoring opportunities. Shots accuracy then express how successful the player is when he gets into this situation, and while shooting on target more with higher frequency does not ensure that the player will score a goal, the probability of scoring a goal is zero if the attempt goes off target.

The number of passes may be a significant value driver for players with both defensive and offensive roles as it may be a good proxy for being a team player and having the mindset for playing technical football (teams with technical style are most likely to play more passes, and the players should be comfortable solving the situations in that way). Passes accuracy then shows how precise the player is in realizing his ideas, which may be very important as inaccuracies in the play are a widespread cause of getting a goal.

The included variables were mostly the same as in the related literature, and we successfully replicated the results. As a next step, independent variables from the Team dataset were added to examine if there will be any additional explanatory effect on the players' value. Santos (2014) uses a PCA method to convert a number of observable variables into one unobservable variable, which should express the team strategy. The scope of this work is not only to explore if there is any effect of the team strategy on the players' value but also to find which of the team metrics that are possibly related to the team strategy are influencing the value most significantly. Therefore, PCA is not used because it would make estimation of each of the effects difficult.

Team goals scored and team goals received are in the model to measure if there is any effect of being from a team that scores or gets a lot of/a few goals. Team cards and also committed fouls may correlate with the playing strategy as well because it describes how often the team uses unfair practices to stop the other team. The number of fouls suffered may be related to the style of play as well because some playing styles bring more contact among players and more fouls as a result. Team shots and team passes should be positively correlated with the offensiveness of a team strategy because the reason for playing offensively is to score more goals, and for scoring more goals, more attempts are needed as well as the passes leading to a scoring opportunity.

Lastly, following the research of Santos (2014), team points are included in the model to account for a different quality of a team. If we omit to include the team points, there would be no way how to differentiate teams that score many goals because they have adopted an offensive strategy from teams that scored more often solely because of their higher quality.

Generally, for cross-sectional data, the heteroskedasticity problem is likely to arise (Wooldridge 2016). To test for heteroskedasticity, the Breusch-Pagan test was employed. The test showed heteroskedasticity in the model. Robust standard errors are reported across this work to take the heteroskedasticity into account. The second issue with the model that can arise is the endogeneity problem. Endogeneity arises when some or all of the explanatory variables are correlated with the error term. It may lead to over or underestimation of these variables. Unfortunately, testing for endogeneity is extremely complicated, and because the cross-sectional data are used in this model, and endogeneity cannot be adequately tested. Our analysis is rather exploratory and it merely indicates certain relationship and eventual over- or underestimation is not a problem.

We can follow a simple logic to show that there is a causal relationship among the explanatory and explained variables. We may say that players are more expensive than others because they are better players, which is in line with the elementary economic theory that higher quality goods are more expensive.

The players are the goods in this example, and with increasing quality, price increases as well. If we would like to change the explanatory variable and prove that the causality is in the opposite direction, we would not be successful. We would need evidence that higher price cause better performance, which does not make any sense because if a mediocre player would be acquired for an outstanding price, it is doubtful that it would make him star of the team.

The OLS regression has the following form:

$$\begin{aligned}
\ln(\text{valueAfterSeason}) = \alpha &+ \beta_1\text{age} + \beta_2\text{age}^2 + \beta_3\text{height} + \beta_4\text{goals} + \beta_5\text{Assists}\\
&+ \beta_6\text{yellowCards} + \beta_7\text{redCards} + \beta_8\text{shots}\\
&+ \beta_9\text{shotsAccuracy} + \beta_{10}\text{passes} + \beta_{11}\text{passesAccuracy}\\
&+ \beta_{12}\text{foulsCommitted} + \beta_{13}\text{foulsSuffered} + \beta_{14}\text{CL}\\
&+ \beta_{15}\text{EL} + \beta_{16}\text{teamPoints} + \beta_{17}\text{teamGoalsScored}\\
&+ \beta_{18}\text{teamGoalsReceived} + \beta_{19}\text{teamCardsYellow}\\
&+ \beta_{20}\text{teamCardsRed} + \beta_{21}\text{teamFoulsCommitted}\\
&+ \beta_{22}\text{teamFoulsSuffered} + \beta_{23}\text{teamShots}\\
&+ \beta_{24}\text{teamPasses} + \gamma_1\text{position} + \gamma_2\text{league}\\
&+ \gamma_3\text{continent} + \epsilon
\end{aligned}$$

In the model, the dependent variable is a natural logarithm of value after season. β represents the scalar coefficient, while γ represents the vector coefficient because position, league, and continent contain sets of dummy variables representing the position of the player, league in which he plays, and the continent of origin.

In order to obtain the information not only about the transfer market as a whole but about what differs the best players in each position, the same model was run on three subsamples, each containing only the players from the same position. We expect that while some of the estimates will be significant for all playing positions, most of them will be significant only for some of the positions and not for the others. This effect may be crucial when the influence of the team strategy will be examined because while the value of defender may increase when the team plays defensively and, as a result, receive fewer goals, strikers would not look as good in that team.

4.2.1 Multicollinearity issue

When models contain a higher number of independent variables, a possible problem with multicollinearity may arise, and therefore, we should test for it. The variance inflation factor is used to find possible problematic variables. VIFs are calculated by regressing each predictor on all other independent variables and by using the R-squared values from this regression in the VIF formula. When there is a high degree of multicollinearity, the VIF will get higher, and we should consider removing the independent variables with high VIF. According to Wooldridge (2016), setting up a cutoff value for VIF when we consider multicollinearity a problem is arbitrary, but a value of 10 is often used. We have calculated VIF for our main model, and the results have shown that some of the variables should suffer from the multicollinearity issue, and we should consider dropping them. Wooldridge (2016) also says that even though we prefer the VIF to be smaller, it should not affect our decision about including or dropping the predictor if we think it is necessary to be included in the regression. The independent variables in our model which had VIF higher than 10 where:

- Age and Age squared, which is common when different powers of the same variable are included, and it is not an issue.
- Hours played - this variable had a VIF very slightly above 10, and a higher degree of multicollinearity can be expected because, as described earlier, better-performing players are likely to have more on-pitch time and as a result, time on the pitch is correlated with the performance metrics. We decided to leave in the model because it is an important variable and should not be left out because it will provide us with key information what time was needed to achieve the performance results each particular player had and we would not be able to differentiate among players who needed twice as much time to achieve the same success as others.
- Team points was another variable that could not be left out because it is the only metric that makes a differentiation between the effect of strategy and the effect of a higher quality of the team possible.
- Lastly, We did not remove team goals scored from the model as it is one of the main variables that change with the change in strategy (Santos 2014), and we need to evaluate the effect of different strategies.

We have included the VIF values in the appendix as well as a comparison of the main model with the model where the variables with VIF over 10 were excluded (except the age variables). The results of the regression stayed intact, except for change in significance for a few of the variables like passes and team passes which we do not consider an issue.

4.3 Testing the influence of team strategy

The main goal of this work is to uncover if a team strategy influences the value of players in that club. A set of team variables is included in the model, and their effect needs to be tested in order to identify which of them are influencing the value and which are not. Due to the complexity of an unobserved variable *Team Strategy*, we will test for an overall significance first using the f-test. Then as a next step, we will identify the variables which are significant on their own, independently on the others. As the last step in identifying the significant team variables, We will want to identify the ones which are not significant by itself, but they are jointly significant with others. As a result, we will be left with team variables which have no significant effect neither on their own nor jointly with the rest. The significant variables will then be interpreted how they are related to the strategy and if we can conclude something about the effect of *Team strategy*.

Chapter 5

Results

5.1 Value estimation results

5.1.1 Personal characteristics

The results of the main regression analysis can be found in Estimation results table, where the dependent variable is the natural logarithm of *Market value*. The results show that not all of the variables are statistically significant.*Age* was identified as very significant and we can tell from the coefficients that the players are reaching their peak at the age of 24 and from this point, the effect of aging overpowers the gain of the new experience. The regression results also show that taller players are demanded more, therefore more expensive, and each extra centimeter of *height* increases the players' value by 1.1%.

5.1.2 Dummy variables

Three sets of dummy variables were included in the model - *Continent of origin*, *League*, and *Position*. Regarding the *Continent of origin*, we can say that being from South America is statistically significant and increases the player's value by 20% in comparison with Africa which is used as a baseline. Other differences among continents are not significant. While strikers and midfielders are valued similarly, being a defender is associated with 25% lower average price on the market. It is caused by the fact that the effect of scoring a goal is easy to observe, has a direct impact on the score, and it is typical for midfielders and strikers to score. Defenders are not only less valuable on the market, but they also have lower wages in comparison with other players (Ante 2019).

Lastly, the results can give us information about how the *Leagues* influence the players' value, and we can sort the leagues by the coefficients to get their quality or "prestigiousness". Some of the players may be more expensive only because they play a competition that is perceived as better. It is interesting to point out that while UEFA (uefa.com) ranks the leagues in order: La Liga, Premier League, Serie A, Bundesliga, and Ligue 1, according to our results the ranking is Premier League, Bundesliga/La Liga, Serie A and Ligue 1. The different orders may come from the criteria that were used to evaluate the leagues. UEFA based its ranking on national coefficients, which are calculated based on the international matches, and therefore it takes only the best clubs of each country into account. In our paper, we do not measure the success in international matches, but the price premium of the leagues.

5.1.3 Performance

Yellow cards, *Red cards*, *Passes*, and *Fouls committed* were found statistically insignificant for the analysis. This was expected, as the results were similar in the previous research Ante (2019); Müller *et al.* (2017). The rest of the performance variables was found significant. Strong evidence for the expected outcome that *Goals scored* is increasing the value can be found in the regression results. Each goal scored increases the value on average by 2%. The variables which were not, to our best knowledge, tested in previous research are *Shots* and *Shots Accuracy* were found significant on 99.9% significance level, both increasing the player's value. None of the performance variables was identified as lowering the players' value, even though negative aspects like *Fouls committed* and both cards were included.

5.1.4 Team strategy

In the previous parts of the chapter, we have successfully replicated the results of previous research and discovered some new findings. The main scope of this work is to analyze the set of team-related variables and their effect on the players' value. By conducting an F-test of all team strategy related variables together, we got the $F(8, 2332) = 8.92$, and it is strong evidence that we can reject the hypothesis that team strategy does not affect the value of the players in that team. As a next step, we would like to know which of the team statistics are the most important on their own.

Goals received have a significant adverse effect on the players' value. Interestingly, the *Team goals Scored* are not significant in determining the player's value. This may look surprising at first, but it is likely to be caused by including variables for the average number of shots and shots accuracy as well, which are both significant, and they are correlated with the number of goals scored. For the negative effects, we do not have similar data, for example, how many times the opposing team shot at our goal. It is likely that due to this fact the estimated effect of receiving a goal cannot be directly compared with the estimated effect of scoring a goal because receiving a goal may be overestimated as it also includes the effect on how easy is for the opponent to get into scoring opportunity and others.

Team cards are surprising part for us because while the estimated effect of *Yellow cards* is not statistically significant, evidence for the positive effect of *Red cards* is strong. *Fouls Committed* and *Fouls suffered* behaves as expected. Both of them are significant, and the estimated effect of *Fouls Committed* is negative, and *Fouls suffered* has a positive effect.

There are only three statistically insignificant variables - *Goals Scored*, *Team yellow cards* and *Team passes.* These three variables were tested for joint significance in all possible combinations but with negative results every time. To show the robustness of the model and the underlying assumptions, three different models were created, and their comparison with the original model can be found in the appendix. In the first model, information about the player position was left out. In the second, we do not control for the continent of origin, and in the last one, we use only player-level data to see how the model behaves when no information about the team performance is included. All three alternative models provides very similar estimation results.

5.2 Subsample differences

5.2.1 Player-related variables

Estimation results for each position can be found in the the table Estimation results for an easy comparison with the main regression. *Age* is one of the variables significant for all the positions, which means that the speed of aging and gaining new experience is similar for all players. We can notice a difference in the case of height. *Height* has a positive effect on 99% confidence level for defenders, 95% for midfielders, and is statistically insignificant for strikers.

Table 5.1: Estimation results part 1

	Players' value			
	All players	Defenders	Midfielders	Strikers
Age	0.65***	0.74***	0.64***	0.50***
	(0.052)	(0.079)	(0.087)	(0.098)
Age2	-0.014***	-0.015***	-0.013***	-0.011***
	(0.00093)	(0.0014)	(0.0016)	(0.0018)
Height	0.011***	0.017**	0.011*	0.0074
	(0.0028)	(0.0052)	(0.0046)	(0.0052)
La Liga	-0.12	-0.083	-0.037	-0.21
	(0.080)	(0.13)	(0.12)	(0.17)
Ligue 1	-0.43***	-0.50***	-0.39***	-0.33*
	(0.069)	(0.11)	(0.11)	(0.14)
Premier League	0.77***	0.78***	0.71***	0.92***
	(0.062)	(0.10)	(0.10)	(0.12)
Serie A	-0.28***	-0.30	-0.28*	-0.19
	(0.072)	(0.12)	(0.11)	(0.15)
Hours played	0.028***	0.021***	0.026***	0.029***
	(0.0027)	(0.0043)	(0.0045)	(0.0068)
Goals	0.020**	0.044	0.018	0.016
	(0.0076)	(0.026)	(0.015)	(0.011)
Assists	0.019*	-0.00054	0.035**	0.029
	(0.0082)	(0.016)	(0.013)	(0.017)
Yellow cards	0.0054	0.0092	0.0018	-0.0024
	(0.0076)	(0.012)	(0.012)	(0.019)
Red cards	0.031	-0.0076	0.089	0.069
	(0.046)	(0.065)	(0.092)	(0.11)
Shots	0.0055***	0.012**	0.0066**	0.0064**
	(0.0014)	(0.0040)	(0.0024)	(0.0024)
N	2370	884	914	572
R^2	0.768	0.766	0.771	0.799

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.2: Estimation results part 2

	Players' value			
	All players	Defenders	Midfielders	Strikers
Shots accuracy	0.0045***	0.0026	0.0045*	0.010***
	(0.00096)	(0.0014)	(0.0018)	(0.0023)
Passes	0.000082	0.000087	0.000077	-0.00013
	(0.000070)	(0.00012)	(0.00011)	(0.00025)
Passes accuracy	0.013***	0.014***	0.014***	0.0081***
	(0.0011)	(0.0019)	(0.0019)	(0.0021)
Fouls committed	-0.0016	0.0029	-0.0036	0.000038
	(0.0017)	(0.0036)	(0.0027)	(0.0034)
Fouls suffered	0.0039**	0.0076**	0.0035	0.0013
	(0.0012)	(0.0024)	(0.0018)	(0.0024)
Played in CL	0.68***	0.75***	0.72***	0.52***
	(0.063)	(0.10)	(0.10)	(0.13)
Played in EL	0.40***	0.48***	0.38***	0.35**
	(0.064)	(0.10)	(0.11)	(0.12)
Team points	0.0084	0.25	0.0019	-0.15
	(0.17)	(0.28)	(0.28)	(0.34)
Team goals scored	0.019	-0.081	0.060	0.049
	(0.13)	(0.21)	(0.21)	(0.27)
Team goals received	-0.55***	-0.48*	-0.56**	-0.54*
	(0.13)	(0.21)	(0.21)	(0.26)
Team yellow cards	-0.034	-0.20	0.0048	0.095
	(0.070)	(0.12)	(0.11)	(0.15)
Team red cards	2.61***	3.19***	1.64*	2.93**
	(0.49)	(0.82)	(0.76)	(1.064)
Team fouls committed	-0.048***	-0.030	-0.041	-0.072*
	(0.014)	(0.023)	(0.022)	(0.032)
N	2370	884	914	572
R^2	0.768	0.766	0.771	0.799

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.3: Estimation results part 3

	Players' value			
	All players	Defenders	Midfielders	Strikers
Team fouls suffered	0.044***	0.061***	0.019	0.057*
	(0.011)	(0.016)	(0.018)	(0.022)
Team shots	0.032**	0.010	0.040*	0.046*
	(0.010)	(0.016)	(0.017)	(0.021)
Team shots accuracy	0.034***	0.024	0.027*	0.052***
	(0.0082)	(0.014)	(0.013)	(0.016)
Team passes	-0.00035	-0.00027	-0.00050	-0.00021
	(0.00018)	(0.00029)	(0.00029)	(0.00037)
Asia	0.084	0.045	-0.17	0.40
	(0.16)	(0.26)	(0.26)	(0.27)
Australia	-0.22	-1.12	0.29	0.26
	(0.42)	(0.85)	(0.26)	(0.36)
Central America	0.25	0.28	0.10	0.49
	(0.14)	(0.19)	(0.27)	(0.34)
Europe	0.052	0.13	-0.043	0.092
	(0.048)	(0.077)	(0.079)	(0.097)
North America	0.16	0.35	0.33	-0.38
	(0.18)	(0.20)	(0.28)	(0.36)
South America	0.20**	0.097	0.17	0.44***
	(0.066)	(0.10)	(0.12)	(0.12)
Defender	-0.26***			
	(0.059)			
Midfielder	-0.063			
	(0.050)			
_cons	-10.72***	-13.13***	-10.29***	-8.99***
	(0.97)	(1.61)	(1.59)	(1.83)
N	2370	884	914	572
R^2	0.768	0.766	0.771	0.799

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This can be observed in real situations because tall players are associated more with defensive roles. After all, it is where the height can bring more benefits. The effect of being from South America is significant only in case of attackers where the average value is 44% higher in comparison with others.

Interesting is the situation with *Goals*. It is a significant variable when we are estimating the value for the whole dataset, but it loses its importance when we are estimating the model for each position separately. While *Shots* are significant value drivers for all the positions, *Shots accuracy* is essential only in the case of midfielders and strikers. This may be surprising for a reader, but in modern football, defenders who can help in the offensive situations are becoming more important than ever before. In their case, the number of shots may signal that they are active in both parts of the game, but the accuracy is not that important because goals are not expected from them even if they take an active part in attacking.

5.2.2 Team Strategy variables

The estimated effect of team variables is very different from the original estimation, and the testing will be more complicated. For all positions, team strategy variables are jointly significant at 99% significance level. Received goals are also significant for everyone as in the main model; other metrics are important only for some of the subsamples.

Defenders' value is influenced by *Received goals*, *Team red cards* and *Fouls suffered*. When we conduct an f-test for the joint significance of the remaining team variables, they are still significant at 95% significance level, and therefore we test which of the variables can have some effect together with others and which are not important in determining the transfer fees. The pair *Fouls Committed* and *Team Yellow cards* are jointly significant, and there is a negative effect on the players' value. A combination where we left out *Team yellow cards* or *Team fouls committed* is insignificant, so we can conclude that the value of defenders is influenced by *Goals received, Red cards, Yellow cards, Fouls committed* and *Fouls suffered.*

The analysis for midfielders and strikers has been conducted using the same steps with the following results:

- Midfielders' value depends on *Goals received, Red cards, Team shots* and *Team shots accuracy* where each of the variables is significant. *Fouls* *committed* and *Team passes* jointly decrease the estimated value of midfielders.

- Strikers' value is significantly influenced by *Goals received, Red cards, Fouls committed, Fouls suffered, Team shots* and *Team shots accuracy* where each of them is significant, and no combination of the remaining team variables is jointly significant.

5.3 Team strategy

In the last part, we need to evaluate what playing strategy increases the value of players if any. The previous research found that teams acquiring offensive strategy are scoring more goals but also due to the higher openness of the game are more likely to suffer goals (Santos 2014). Teams with an offensive strategy are also less likely to commit fouls and get yellow cards Santos (2014).

To our best knowledge, there is no detailed research on the team strategy in football and its influence on match statistics. Therefore, We use assumptions about the effects of the strategy on metrics in this research that comes from the logic of football, the definition of what is an offensive strategy and a personal experience as a football player. Firstly, we assume that the number of shots is positively correlated with an offensive strategy because it is the goal of the strategy itself - to create more scoring opportunities. The assumption about shots accuracy is based on the fact that teams playing offensively have more practice in solving the situations that comes from attacking more often, and shots accuracy should be positively correlated with the offensiveness as well.

The assumption about red cards stands on the findings of Santos (2014) that yellow cards are less often given to the teams that attack more. Red cards should be the same case because the most dangerous fouls for which red cards are being given are much more likely to occur during the defensive part of the game. We assume that despite the strong evidence for red cards having a positive effect on players' value, it is not relevant because the red cards are unfavorable for all the teams in any situation. After all, the team has a disadvantage of fewer players for the rest of the match. As a result, all the teams are trying to avoid red cards, and the number of them is quite small for all the teams in the dataset. The number of red cards depends more on the discipline than on the offensiveness or defensiveness. The significant effect of red cards probably comes from the fact that teams with more funds got more red cards that season, not from the fact that team playing dangerously would be valued higher. Lastly, fouls suffered are likely to be positively correlated with offensive strategy as well because when the team is attacking more, the opponent must defend more, and according to Santos (2014), defensive play correlates with more fouls committed.

Now, when we have defined how strategy influences our observed team variables, we can evaluate if there is any evidence that offensive strategy leads to higher valuation of the players. Defenders value decreases when their team suffers many goals, commit fouls, and gets yellow cards. An increase in the value is associated with a higher average of suffered fouls. While receiving more goals is a typical effect of an offensive strategy, all the other aspects should be better with an offensive strategy. Midfielders' value will decrease with the defensive strategy due to the expected higher number of committed fouls, fewer shots, and worse shots accuracy. On the other hand, a decrease in value for playing in a team that receives many goals is significant. Lastly, strikers estimated value would be higher with an offensive strategy due to the importance of both fouls suffered and committed and the number of *Shots* and their accuracy.

As we may observe, more of the variables associated with an offensive style of play are significant as our focus shifts from the defensive players towards midfielders and strikers. The value of each group will decrease with the received goals, but each of them has different team metrics, which may compensate for that decrease. The question is if the positive effect of offensive play will overcome that decrease or if the teams should play defensively because these effects cannot compensate for the higher number of received goals. We may argue that while defenders' valuation will probably be higher with more defensive play because there are not that many team variables that may justify the offensive style, the effect of offensive strategy on midfielders and strikers may be positive.

Chapter 6

Conclusion

Football is a multi-billion business where making a profit is not always the primary goal, but with increasing costs, popularity and with an introduction UEFA Financial Fair Play Regulations, attention to selling the players the smart way was increased. The main goal of the work was to investigate if there is a relationship between team strategy and the value of each player in that team. The hypothesis was that because offensive players cost more (Müller *et al.* 2017; Ante 2019), teams with offensive strategies will be able to sell their players for a higher price on the transfer market even though their performance statistics will be the same as the players from other clubs.

Using linear regression on data from the top five European football leagues for season 2018/2019, we have defined the most influential variables. From player-related metrics, age, the number of goals scored, and shots Accuracy are just some of the most significant. The key finding is that team-related statistics have a significant effect on players' value even when we account for a different quality of clubs. It means that the potential transfer fee club may obtain when selling its players depends not only on the quality of the player and the current results of the team but also on the way the team achieved its results. This effect is significant for each playing position included in the research - defenders, midfielders, and strikers, but the relationship between independent variables and the dependent variable is different for each of the positions.

To link the team variables with different playing strategies, We use the findings of Santos (2014), the crowd-wisdom, and his personal experience. Therefore, changes in each of the team-related metrics are associated with the change in the playing strategy. Only one team metric was identified as significant for all positions, and it was the number of goals received. The number of significant metrics that are associated with offensive strategy increased with the offensiveness of the position for which the regression was evaluated.

The findings of this paper may be used in a multiple ways. Firstly, with the knowledge that the environment of each team influences the players' value, team scouts should be able to identify players with very similar qualities and get them with a discount when choosing the one from the team with cheaper players. We see another use of the work as a base for further research in the field of team strategy and application of this knowledge on players' valuation.

While the work proves, that team strategy has a significant effect on players' value and describes how each of the team metrics changes with the chosen playing strategy, it does not explore the relationships among these metrics into detail. Extension of the work may be research focusing more on the team strategy itself to estimate how much the team metrics change when the strategy is changed. It would allow for a more detailed interpretation of this paper. The problem with that research would be that the strategy metrics are changing simultaneously and the strategy is constantly changing as well, so detailed knowledge of football, well designed assumptions and the right estimating methods should be chosen carefully.

Bibliography

ADLER, M. (1985): "Stardom and talent." *The American Economic Review* **75(1)**: pp. 208–212.

ANTE, L. (2019): "Determinants of Transfers Fees Evidence from the Five Major European Football Leagues." *abc* .

BOE (2019): "Why are football players paid so much?" .

BRYSON, A., B. FRICK, & R. SIMMONS (2009): "The returns to scarce talent." *LSE Research Online Documents on Economics* .

CARMICHAEL, F. & D. FORREST (1999): "The labour market in association football: Who gets transferred and for how much?" *Bulletin of Economic Research* **51**: pp. 125–50.

CARMICHAEL, F. & D. THOMAS (1993): "Bargaining in the transfer market: Theory and evidence." *Applied Economics* **25**: pp. 1467–76.

DOBSON, S. & B. GERRARD (1999): "The determination of player transfer fees in english professional soccer." *Journal of Sport Management* **13(4)**: p. 259.

FEESS, E. & G. MUEHLHEUSSER (2003): "The impact of transfer fees on professional sports." *The Scandinavian Journal of Economics* **105(1)**: pp. 139–154.

FRANCK, E. & S. NUEESCH (2008): "Mechanisms of superstar formation in german soccer." *EUROPEAN SPORT MANAGEMENT QUARTERLY* **8(2)**: pp. 145–164.

FRANCK, E. & S. NÜESCH (2012): "Talent and/or popularity." *Economic Inquiry* **50(1)**: pp. 202–216.

GUEDES, J. C. & F. S. MACHADO (2002): "Changing rewards in contests." *Empirical Economics* **27(4)**: pp. 607–630.

He, M., R. Cachucho, & A. Knobbe (2015): "Football player's performance and market value." In "Mass," .

Herm, S., H. markus Callsen-bracker, & H. Kreis (2014): "When the crowd evaluates soccer players' market values." *Sport Management Review* **17(4)**: pp. 484–492.

Kalen, A., E. Rey, A. S. de Rellan-Guerra, & C. Lago-penas (2019): "Are soccer players older now than before? aging trends and market value in the last three decades of the uefa champions league." *FRONTIERS IN PSYCHOLOGY* **10**.

Kiefer, S. (2012): "The impact of the euro 2012 on popularity and market value of football players." *Discussion Papers of the Institute for Organisational Economics* .

Moschini, G. (2010): "Incentives and outcomes in a strategic setting." *Economic Inquiry* **48(1)**: p. 65.

Müller, O., A. Simons, & M. Weinmann (2017): "Beyond crowd judgments." *European Journal of Operational Research* **263(2)**: pp. 611–624.

Palomino, F., L. Rigotti, & A. Rustichini (1998): "Skill, strategy and passion." *Discussion Paper* .

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

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

Ruijg, J. & H. van Ophem (2015): "Determinants of football transfers." *APPLIED ECONOMICS LETTERS* **22(1)**: pp. 12–19.

Santos, R. M. (2014): "Optimal soccer strategies." *Economic Inquiry* **52(1)**: pp. 183–200.

Sawe, B. E. (2018): "The most popular sports in the world." Worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html.

Wooldridge, J. M. (2016): *Introductory econometrics: a modern approach.* Cengage.

Appendix A

Title of Appendix A

Table A.1: Summary statistics for defenders

	Defenders			
Variables	mean	sd	min	max
Value after season	8.87	12.87	.05	90
Age	26.94	4.28	18	42
Height	183.63	5.88	168	199
Hours played	24.21	17.34	0	57
Goals	.69	1.12	0	8
Assists	.84	1.47	0	13
Yellow cards	3.15	2.96	0	16
Red cards	.10	.34	0	3
Shots	8.83	8.92	0	46
Shots accuracy	22.57	22.57	0	100
Passes	608.66	511.32	0	2768
Passes accuracy	69.97	24.32	0	100
Fouls committed	16.32	13.01	0	60
Fouls suffered	12.90	12.57	0	78
Played in CL	.17	.37	0	1
Played in EL	.16	.37	0	1
Observations	884			

Table A.2: Summary statistics for midfielders

	Midfielders			
Variables	mean	sd	min	max
Value after season	11.21	16.36	.05	130
Age	26.41	4.10	18	39
Height	179.73	5.90	162	196
Hours played	21.44	16.57	0	57
Goals	1.61	2.36	0	17
Assists	1.54	2.19	0	14
Yellow cards	3.09	3.09	0	16
Red cards	.063	.24	0	1
Shots	19.40	20.13	0	105
Shots accuracy	24.30	19.03	0	100
Passes	534.44	494.47	0	2742
Passes accuracy	70.50	24.36	0	100
Fouls committed	19.57	17.20	0	98
Fouls suffered	19.35	18.94	0	103
Played in CL	.15	.36	0	1
Played in EL	.16	.36	0	1
Observations	914			

Table A.3: Summary statistics for strikers

	Strikers			
Variables	mean	sd	min	max
Value after season	13.87	24.31	.05	200
age	26.29	4.19	18	41
height	181.69	6.27	163	201
hoursplayed	19.09	16.37	0	57
goals	4.49	5.59	0	36
assists	1.85	2.47	0	15
yellow cards	1.96	2.18	0	12
red cards	.049	.224	0	2
shots	32.88	33.13	0	177
shotsAcc	32.08	20.05	0	100
Passes	254.77	268.10	0	1485
passAcc	62.56	23.81	0	100
fouls committed	18.04	17.26	0	90
fouls suffered	19.33	20.47	0	112
Played in CL	.15	.36	0	1
Played in EL	.14	.35	0	1
Observations	572			

Table A.4: Summary statistics for Bundesliga

	Bundesliga			
Variables (Avg. per match)	mean	sd	min	max
Points	1.38	.50	.56	2.29
Goals scored	1.59	.49	.76	2.59
Goals received	1.59	.37	.85	2.12
Yellow cards	2.07	.44	1.29	2.91
Red cards	.043	.037	0	.12
Fouls committed	14.28	2.57	10.21	19.12
Fouls suffered	13.38	1.72	11.12	16.44
Shots per match	14.71	3.45	10.39	22.53
Shots accurracy	.35	.040	.27	.4
Passes	355.94	91.53	239	551
Observations	18			

Table A.5: Summary statistics for Premier League

	Premier league			
Variables (Avg. per match)	mean	sd	min	max
Points	1.41	.55	.42	2.58
Goals scored	1.41	.48	.58	2.5
Goals received	1.41	.41	.58	2.13
Yellow cards	1.99	.38	1.5	2.84
Red cards	.045	.031	0	.11
Fouls committed	12.82	2.43	9.5	17.34
Fouls suffered	12.11	2.71	9	19.79
Shots	16.18	5.14	9.58	29.05
Shots accurracy	.34	.028	.30	.40
Passes	362.10	114.43	158	622
Observations	20			

Table A.6: Summary statistics for Ligue 1

	Ligue 1			
Variables (Avg. per match)	mean	sd	min	max
Points	1.35	.42	.71	2.39
Goals scored	1.28	.50	.74	2.76
Goals received	1.28	.24	.87	1.79
Yellow cards	2.16	.44	1.32	2.97
Red cards	.099	.041	.03	.21
Fouls committed	13.68	1.65	11.29	16.29
Fouls suffered	13.07	1.56	9.97	16.18
Shots	12.91	2.80	9.58	20.18
Shots accurracy	.34	.037	.28	.42
Passes	334.95	74.31	245	539
Observations	20			

Table A.7: Summary statistics for Serie A

	Serie A			
Variables (Avg. per match)	mean	sd	min	max
Points	1.36	.48	.53	2.37
Goals scored	1.34	.37	.66	2.03
Goals received	1.34	.33	.79	1.97
Yellow cards	2.58	.31	1.92	3.26
Red cards	.069	.041	0	.16
Fouls committed	14.82	1.67	12.61	17.87
Fouls suffered	13.99	1.77	11.05	16.89
Shots	15.62	4.31	9.24	23.79
Shots accurracy	.32	.026	.28	.37
Passes	343.55	80.41	228	508
Observations	20			

Table A.8: Summary statistics for La Liga

	La Liga			
Variables (Avg. per match)	mean	sd	min	max
Points	1.36	.37	.84	2.29
Goals scored	1.29	.34	.84	2.37
Goals received	1.29	.29	.76	1.84
Yellow cards	3.01	.54	2.13	3.97
Red cards	.045	.025	0	.11
Fouls committed	16.34	2.014	12.84	20.66
Fouls suffered	15.71	3.42	10.76	22.61
Shots	15.21	4.29	10.71	24.74
Shots accurracy	.35	.030	.3	.43
Passes	326.95	92.50	184	574
Observations	20			

Table A.9: Robustness check model comparison part 1

	Players' value			
	Main model	No position	No continent	No team data
Age	0.65***	0.65***	0.64***	0.59***
	(0.052)	(0.053)	(0.052)	(0.054)
Age2	-0.014***	-0.014***	-0.013***	-0.013***
	(0.00093)	(0.00095)	(0.00093)	(0.00097)
Height	0.011***	0.0087**	0.011***	0.011***
	(0.0028)	(0.0028)	(0.0028)	(0.0030)
La Liga	-0.12	-0.13	-0.11	0.063
	(0.080)	(0.081)	(0.081)	(0.056)
Ligue 1	-0.43***	-0.44***	-0.45***	-0.15**
	(0.069)	(0.069)	(0.069)	(0.055)
Premier League	0.77***	0.77***	0.77***	0.93***
	(0.062)	(0.062)	(0.062)	(0.051)
Serie A	-0.28***	-0.3***	-0.28***	-0.19***
	(0.072)	(0.073)	(0.073)	(0.057)
Hours played	0.028***	0.023***	0.027***	0.015***
	(0.0027)	(0.0024)	(0.0027)	(0.0027)
Goals	0.020**	0.022**	0.020**	0.041***
	(0.0076)	(0.0076)	(0.0076)	(0.0078)
Assists	0.019*	0.021*	0.019*	0.037***
	(0.0083)	(0.0083)	(0.0083)	(0.0088)
Yellow cards	0.0054	0.0025	0.0067	-0.0013
	(0.0076)	(0.0076)	(0.0076)	(0.0079)
Red cards	0.0306	0.016	0.0401	0.053
	(0.046)	(0.046)	(0.046)	(0.050)
Shots	0.0055***	0.0080***	0.0057***	0.0056***
	(0.0014)	(0.0013)	(0.0014)	(0.0015)
N	2370	2370	2370	2370
R^2	0.77	0.77	0.77	0.73

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.10: Robustness check model comparison part 2

	Players' value			
	Main model	No position	No continent	No team data
Shots accuracy	0.0045***	0.0050***	0.0045***	0.0050***
	(0.00096)	(0.00097)	(0.00096)	(0.0011)
Passes	0.000082	0.000095	0.000085	0.00053***
	(0.000070)	(0.000068)	(0.000070)	(0.000063)
Passes accuracy	0.013***	0.013***	0.013***	0.012***
	(0.0011)	(0.0011)	(0.0011)	(0.0012)
Fouls committed	-0.0016	0.00040	-0.0014	-0.0022
	(0.0017)	(0.0016)	(0.0017)	(0.0018)
Fouls suffered	0.0039**	0.0045***	0.0041***	0.0038**
	(0.0012)	(0.0012)	(0.0012)	(0.0013)
Played in CL	0.68***	0.68***	0.68***	1.13***
	(0.063)	(0.063)	(0.063)	(0.044)
Played in EL	0.40***	0.40***	0.39***	0.67***
	(0.064)	(0.064)	(0.064)	(0.043)
Team points	0.0084	0.0074	0.0019	
	(0.17)	(0.17)	(0.17)	
Team goals scored	0.019	-0.00051	0.032	
	(0.13)	(0.13)	(0.13)	
Team goals received	-0.55***	-0.56***	-0.57***	
	(0.13)	(0.13)	(0.13)	
Team yellow cards	-0.034	-0.025	-0.031	
	(0.070)	(0.071)	(0.070)	
Team red cards	2.61***	2.59***	2.66***	
	(0.49)	(0.49)	(0.49)	
Team fouls committed	-0.048***	-0.050***	-0.049***	
	(0.014)	(0.014)	(0.014)	
N	2370	2370	2370	2370
R^2	0.77	0.77	0.77	0.73

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.11: Robustness check model comparison part 3

	Players' value			
	Main model	No position	No continent	No team data
Team fouls suffered	0.044***	0.044***	0.044***	
	(0.011)	(0.011)	(0.011)	
Team shots	0.032**	0.032**	0.033**	
	(0.010)	(0.010)	(0.010)	
Team shots accuracy	0.034***	0.034***	0.034***	
	(0.0082)	(0.0082)	(0.0082)	
Team passes	-0.00035	-0.00035*	-0.00036*	
	(0.00018)	(0.00018)	(0.00018)	
Asia	0.084	0.11		0.11
	(0.16)	(0.16)		(0.16)
Australia	-0.22	-0.24		-0.33
	(0.42)	(0.45)		(0.40)
Central America	0.25	0.23		0.28
	(0.14)	(0.14)		(0.15)
Europe	0.052	0.046		0.086
	(0.048)	(0.048)		(0.051)
North America	0.16	0.18		0.25
	(0.18)	(0.18)		(0.23)
South America	0.20**	0.18**		0.31***
	(0.066)	(0.066)		(0.069)
Defender	-0.26***		-0.25***	-0.22***
	(0.059)		(0.059)	(0.063)
Midfielder	-0.063		-0.067	-0.063
	(0.050)		(0.050)	(0.053)
_cons	-10.72***	-10.33***	-10.52***	-9.42***
	(0.97)	(0.96)	(0.96)	(0.92)
N	2370	2370	2370	2370
R^2	0.77	0.77	0.77	0.73

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.12: Variance inflation factors

	Variance inflation factors
Age	138.24
Age2	135.50
Height	1.24
La Liga	4.16
Ligue 1	3.13
Premier League	2.92
Serie A	3.23
Hours played	10.12
Goals	4.42
Assists	2.16
Yellow cards	3.04
Red cards	1.08
Shots	6.55
Shots accuracy	1.35
Passes	5.89
Passes accuracy	1.55
Fouls committed	4.46
Fouls suffered	3.15
Played in CL	2.07
Played in EL	1.87
Team points	22.68
Team goals scored	13.12
Team goals received	8.51
Team yellow cards	5.39
Team red cards	1.63
Team fouls committed	4.71
Team fouls suffered	3.30
Team shots	6.60
Team shots accuracy	3.51
Team passes	1.04
Asia	1.12
Australia	1.04
Central America	1.10
Europe	2.14
North America	1.10
South America	1.92
Defender	2.77
Midfielder	2.15

Table A.13: Comparison of main model with VIF adjusted model part 1

	Players' value	
	Main model	VIF adjusted model
Age	0.65***	0.66***
	(0.052)	(0.053)
Age2	-0.014***	-0.014***
	(0.00093)	(0.00094)
Height	0.011***	0.011***
	(0.0028)	(0.0029)
La Liga	-0.12	-0.13
	(0.080)	(0.079)
Ligue 1	-0.43***	-0.45***
	(0.069)	(0.067)
Premier League	0.77***	0.81***
	(0.062)	(0.056)
Serie A	-0.28***	-0.32***
	(0.072)	(0.072)
Hours played	0.028***	
	(0.0027)	
Goals	0.020**	0.030***
	(0.0076)	(0.0079)
Assists	0.019*	0.029***
	(0.0083)	(0.0086)
Yellow cards	0.0054	0.015
	(0.0076)	(0.0078)
Red cards	0.0306	0.047
	(0.046)	(0.048)
Shots	0.0055***	0.0093***
	(0.0014)	(0.0014)
N	2370	2370
R^2	0.77	0.76

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.14: Comparison of main model with VIF adjusted model part 2

	Players' value	
	Main model	VIF adjusted model
Shots accuracy	0.0045***	0.0053***
	(0.00096)	(0.00098)
Passes	0.000082	0.00061***
	(0.000070)	(0.000051)
Passes accuracy	0.013***	0.013***
	(0.0011)	(0.0011)
Fouls committed	-0.0016	0.0043**
	(0.0017)	(0.0016)
Fouls suffered	0.0039**	0.0060***
	(0.0012)	(0.0012)
Played in CL	0.68***	0.65***
	(0.063)	(0.062)
Played in EL	0.40***	0.41***
	(0.064)	(0.064)
Team points	0.0084	
	(0.17)	
Team goals scored	0.019	
	(0.13)	
Team goals received	-0.55***	-0.57***
	(0.13)	(0.069)
Team yellow cards	-0.034	-0.015
	(0.070)	(0.071)
Team red cards	2.61***	2.50***
	(0.49)	(0.49)
Team fouls committed	-0.048***	-0.045**
	(0.014)	(0.014)
N	2370	2370
R^2	0.77	0.76

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.15: Comparison of main model with VIF adjusted model part 3

	Players' value	
	Main model	VIF adjusted model
Team fouls suffered	0.044***	0.040***
	(0.011)	(0.010)
Team shots	0.032**	0.021*
	(0.010)	(0.0087)
Team shots accuracy	0.034***	0.029***
	(0.0082)	(0.0063)
Team passes	-0.00035	-0.00036*
	(0.00018)	(0.00018)
Asia	0.084	0.12
	(0.16)	(0.17)
Australia	-0.22	-0.25
	(0.42)	(0.45)
Central America	0.25	0.23
	(0.14)	(0.15)
Europe	0.052	0.053
	(0.048)	(0.048)
North America	0.16	0.16
	(0.18)	(0.18)
South America	0.20**	0.17*
	(0.066)	(0.067)
Defender	-0.26***	-0.14*
	(0.059)	(0.058)
Midfielder	-0.063	-0.080
	(0.050)	(0.050)
_cons	-10.72***	-10.43***
	(0.97)	(0.96)
N	2370	2370
R^2	0.77	0.76

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$