Augsburg Honors Review

Volume 11 Article 5

2018

What Makes a Soccer Player Expensive? Analyzing the Transfer Activity of the Richest Soccer

Lukas Barbuscak
Augsburg University

What Makes a Soccer Player Expensive? Analyzing the Transfer Activity of the Richest Soccer

Lukas Barbuscak
Augsburg University

Abstract

This study focuses on developing a model for transfer fees paid for the players sold to the fifteen most valuable and the richest soccer clubs in the world. Using the data primarily collected from transfermarkt.com, a leading soccer database focusing on transfers and soccer community valuations, we ran a linear regression using five variables measuring popularity, monetary valuation, and productivity. Results show that variables such as number of Google searches, number of years on contract left, community valuation of a player, number of goals and assists, and race of a player significantly influence the transfer fee paid for a player. Moreover, it shows that the transfer fee for the most expensive soccer player ever is not random and can be explained using the right independent variables.

Introduction

Today's soccer transfer market has grown increasingly extreme. The fees paid for players often seem exaggerated, especially in recent years. In summer 2017, the prices for players have skyrocketed. The transfer fee for Neymar exceeded 222 million euros, which was the highest amount ever spent for a soccer player. The general public shook their heads, unable to see the logic in such high fees. In other words, paying such an outrageous price for a soccer player simply did not make any sense (Goal.com, 2017). In this research paper, we examine the rationale for these high transfer payments.

Literature Review

Several studies have analyzed the transfer fees in soccer. Carmichael and Thomas (1993) examined the transfer market in the English Football League and applied a bargaining theory to explain the transfer fees. The player characteristics in their equation were age, goals, general fitness, league appearances, and the position played by the player. Among other conclusions, the authors found that the transfer fee increased by a couple of percentage points by each goal or league appearance (p. 1475). In addition, they concluded that the transfer fees started declining after players reached 25 years of age.
Other studies incorporated bargaining elements into the equation such as those completed by Dobson and Gerrard (1999) and Speight and Thomas (1997). Dobson and Gerrard studied the inflation rate of the transfer fees between the years 1990 and 1996 as well as what factors determine the player's price on the market. The main variables used were age, league starts, goals, number of international starts, or goal rate. Dobson and Gerrard (1999) differentiated between the buying club and the selling club. The authors found that the average rate of inflation for transfer fees was around 11% per year between 1990 and 1996. Speight and Thomas (1997) compared the negotiated transfer fees with arbitration settlements. In their article, the authors used both individual predictors (goals, experience, fitness) as well as club variables (bargaining strength of both selling and buying clubs). Speight and Thomas found that, unfortunately for the selling clubs, arbitration settlements were lower than the potential market value, suggesting that the model of arbitrage was beneficial for the buying clubs who would rather wait for arbitration than negotiate the transfer fee.

Dobson, Gerrard, and Howe (2000) continued the research on transfer fees using the negotiation aspect, this time in non-league English soccer. The player characteristics used were age, the number of previous clubs, the number of appearances, goals, or the position dummy variables. Dobson et al. again used a model which differentiated between the buying and selling clubs. The authors concluded that the transfer market in both non-league and professional soccer is very alike with broadly similar data-generating processes.

Unlike Dobson, Gerrard, and Howe, who focused on non-league soccer, Frick

(2007) analyzed the labor market in the top five European soccer leagues. Frick criticized the researchers, claiming that "most of the statistics used are rather 'indirect' measures that only imperfectly reflect a player's value to his team" (p. 426), with an exception of the number of assists and tackles by the player as well as the superstar effect. Frick also recognized footedness and the contract duration (or number of years of contract left) as important variables in predicting the transfer fees of the players. Frick argued that "the more successful the buying and/or the selling club are (either in economic or in sporting terms), the higher the transfer fee that the two clubs agree on" (p. 431). This paper confirmed conclusions drawn by the authors of the previous studies which incorporated bargaining elements and the power of selling and buying clubs in soccer economics (Carmichael & Thomas, 1993; Dobson & Gerrard, 1999; Speight & Thomas, 1997; Dobson, Gerrard, & Howe, 2000).

Another study that recognized contract duration as an important predictor of the transfer fee was done by Carmichael, Forrest, and Simmons (1999). The authors studied the variables that might predict transfer fees and tried to develop a model to determine the prices of players. Unlike other studies, Carmichael et al. included players that did not get transferred in their data set. They claimed that data on contract duration, time of expiry, and renewal are "private," thus they could not include the variables in their model. The independent variables used in this study were goals, age, age squared, players' positions, international starts, and several team variables. The authors found that some players were more likely to be transferred than others, and if the transfer happened, it was explained by the variables used.

Two different approaches were used by other researchers to determine the transfer fees in soccer. Herm, Callsen-Bracker, and Kreis (2014) focused on the community approach of analysis for transfer fees, measuring how the soccer community calculates the transfer fee and how well the soccer community was able to predict the observed transfer fees. Conversely, Muller, Simons, and Weinmann (2017) focused on a data-driven approach, predicting the market value by objective data on both performance and popularity. Herm et al. (2014) evaluated the accuracy of the online community in predicting transfer values in soccer. The authors used market values from transfermarkt.de, which "have a good reputation in the sports industry and have a high economic relevance; they are used in actual transfer and salary negotiations" (p. 487). The variables used as indicators of the market value were age, precision, scoring, assertion, footedness, and several external attributes such as public attention or other clients of the player's agent. The results showed that the most objective predictors of market value identified by the community were goals, age, percentage of successful passes, duels won, footedness, and the attention of general public measured by the number of the player's Google search results. Muller et al. (2017), on the other hand, used a strictly data-driven approach and criticized the belief that the soccer community could produce similar or better results than using objective data. Using a data set of 4217 players from the top five European leagues over the course of six seasons, Muller et al. found that the data approach used in the model "can overcome several of crowdsourcing's practical limitations while producing comparatively accurate estimates" (p. 621), suggesting that the objective variables measuring performance and popularity were significant in determining the value of the player.

Several studies focused on the determinants of players' salaries instead of the transfer fees. Brandes and Franck (2012) studied the performance reactions of German soccer players to fair and unfair wage allocations. "Fair" wage would be explainable by the previous season's performance as well as by age and experience. The authors used Opta Index, a product of a professional sports company, to analyze players "performance index" and used that number in their model. Brandes and Franck found that the deviations from the fair wage indeed influenced the players' performance on the pitch. Frick (2011) also studied player remuneration and the contract duration. To determine the player's salary, Frick developed an econometric model which included both last season's and career statistics. Frick concluded that recent performances (last season) influence the salary much more than the player's past performances. The author also found that region of birth was important, goalkeepers earned significantly less than the field players, the captaincy of the player matters as well, and, the maximum earnings are enjoyed by players when they are 27–28 years old.

Another study conducted by Bryson, Frick, and Simmons (2013) explored footedness, which is an ability to play soccer with both feet, and connected the ability to players' salaries. The authors used *Kicker Magazine* and transfermarkt.de data sets which included the top 5 European soccer leagues. Bryson et al. found that two-footedness significantly adds to the players' salary. Although, it does not significantly improve the team's performance. Fry, Galanos, and Posso (2014) also found a certain premium of being a left-footed player. They studied the determinants of the players' goal productivity. Fry et al. found a concave relationship between age and productivity, and that young, tall, left-footed strikers have the best probability of scoring a goal.

Productivity, the number of goals and assists of a player over the course of one season, nationality of the player, and popularity seem to be the main predictors of the "superstar effect" in soccer. Lucifora and Simmons (2003) examined the superstar effect in the Italian soccer league, the famous Serie A. Other than using the traditional variables (goals, assists, appearances), Lucifora and Simmons used two new dummy variables — "superstar indicators" — depending on the goals per game. The authors found that "superstar effects are generated by consumer interest in forward players who frequently score goals and/or create goals for their colleagues," and this was confirmed by these players having higher salaries. This finding explains a general consensus in the soccer world: attacking players are more valuable and earn more than other players (Bryson, Frick, & Simmons, 2013).

As previously mentioned, productivity does not seem to be the only component of the superstar effect. Bryson, Rossi, and Simmons (2014) explored how migrant players influence winning. The independent variables used were player's productivity, age, experience, season dummies, and nationality dummies. Bryson et al. used three dependent variables: wages, attendance, and team points. The authors found that there was a substantial wage premium for migrant soccer players which is only partially explained by their productivity. They concluded that the "superstar effect" indeed exists, suggesting that the nationality of the player matters.

The popularity component of the superstar effect was explored by Garcia-del-Barrio and Pujol (2004; 2007). The data set used was *Marca*, a Spanish newspaper, and the variables used, in addition to the traditionally used goals, assists, or games played, included the revenue-generating capability of the player (expressed by the total number of links reported by Google). The authors concluded that the transfer value of players stemmed both from their sporting performance and their economic contribution — which was one of the most surprising and relevant findings of their paper, but also one that absolutely makes sense; a player's popularity and mass media presence increases his marketing potential.

The conclusions by Garcia-del-Barrio and Pujol were tested by Lehmann and Schulze (2008). The authors explored the superstar theories for the German Bundesliga and whether superstar status and talent explain players' salaries. Variables used were goals, assists, various team dummies, players' position, tackles, and Google searches. Lehmann and Schultze found that both individual performance and media presence significantly explained German players' salaries. A very similar conclusion was reached by Franck and Nuesch (2012). The authors studied talent and popularity and their contribution to the team's success. Franck and Nuesch used data from Kicker and transfermarkt. Specifically, they used 20 variables that determined the transfer fee and, thus, talent. In addition to the traditional variable, they also examined yellow and red cards, cross-completion rate, shots, or assists. The authors showed that both talent and popularity contribute to the transfer values of the players.

The studies focusing on nationality had conflicting results. Frick, Pietzner and Prinz (2007) studied several questions related to the career duration in professional soccer. The authors used an extensive data set from Kicker, and included individual characteristics, position dummies, nationality dummies, and institutional environment. Frick et al. concluded that there seemed to be a discrimination based on nationality of the player regarding the career duration. For instance, South American players tend to be eliminated from the top German league more frequently than players from other countries.

Pedace (2008) studied wage discrimination in the English soccer league. The author found that South American players tended to be overpaid even when they did not increase the team performance. However, they did increase attendance. Findings by Wilson and Yung-Hsiang (2003) seemed to suggest the opposite. The authors analyzed nationality discrimination in the five largest soccer leagues. Wilson and Yung-Hsiang concluded that players from Balkan or South America increase performance of the team. However, these players were under-represented in the teams despite having positive influence on attendance and teams' performances. This might indicate bias from owners of the soccer clubs.

When it comes to a race of the player, results were also conflicting. Reilly and Witt (1995) explored the element of race in the transfer fees. The authors included a number of performance variables, differentiated between the divisions of buying and selling clubs, and included a dummy variable if the player was black. Reilly and Witt found that there was very little evidence of discrimination in transfer fees of these players. However, the study did not exclude possible discrimination in other forms.

Szymanski (2000) proposed a market test for racial discrimination in English soccer leagues. The author considered several variables indicating players' performance such as age, career length, goals scored, or player's position on the pitch. The results suggested that the clubs which were underrepresented by black players performed worse than other clubs, thus concluding that owners "have paid a premium in the player market" (p. 601) by discriminating against black players.

Theoretical Model

This study used a community-driven approach rather than a data-driven approach to analyze transfer fees for soccer players and to focus solely on the top 15 richest soccer clubs in the world; the choice of variables was highly influenced by this focus (which is a big difference between our study and studies done before despite sharing a similarity of researching transfer fees and related variables). Players coming to big clubs are usually well known before making the transfer — thus their market value, the specifics of their contract, or the players' popularity measured by their exposure in media are also well known and were suggested in our literature review to be crucial in determining the players' transfer fees. Our literature review often discussed goals and assists as variables included in various studies; they essentially suggest productivity in soccer, thus inclusion of such variables are crucial. The race variable is also significant in our analysis since past research suggests that clubs enjoy significant premium in team performance from non-white players despite their underrepresentation in soccer clubs. Due to the size of our sample, inclusion of certain performance (tackles, yellow and red cards, or passing accuracy) and non-performance (international starts, nationality) data would not bring anything to the model since the community might include that in their market value estimates.

Therefore, our general model is as follows:

$$\boldsymbol{Fee = \beta 1 + \beta 2 MarketValue + \beta 3 ContractLeft + \beta 4 GoogleResults + \beta 5 Productivity + \beta 6 Race + \varepsilon,}$$

where our dependent variable is the actual transfer fee of the player as reported on transfermarkt.com.

Table 1: Independent Variables

Independent Variable	Explanation
MarketValue	Market Value reported by transfermarkt.com at the time of the transfer
ContractLeft	Number of years a player had on his contract with the former club
GoogleResults	Number of Google results when searching the player
Productivity	Number of goals and assists from the previous season
Race	Dummy variable, race of the player

There are 5 independent variables in our model. MarketValue indicates the value of the player at the time of the transfer (not at time of collecting data), as reported by transfermarkt.com. The market value decided by the community members according to the judge principle — a method of deciding the market value of the player hierarchically by selecting a judge who evaluates inputs from community members and has the final say on what market value is reported on the website — can be considered as a good estimate of the players' true values (Herm et al., 2014; Muller et al., 2017). Transfermarkt.com market values are based on the community judgment (Herm et al., 2014). The expected sign is positive since we expect a positive correlation between the transfer fee and the community-determined transfer value.

ContractLeft is a number in years suggesting how many years a player has left on his contract; in other words, it shows how long a player has to play for his current club before becoming a free agent. Usually, however, players never reach free agency as the players are sold for at least some kind of monetary compensation before their contracts expire, circumventing a player leaving "for free." The selling clubs hold more power when they have the player signed for four years rather than for one — the expected sign is therefore positive. The more years the players have on their contract, the higher a transfer fee can be demanded for the player.

GoogleResults is the number of Google results when searching for a specific player on Google. The method for collecting these searches was previously used in various research articles (Garcia-del-Barrio, Pujol, 2004; 2007) and is as follows: "[name of the player]" AND "[the name of the selling club]" AND "football." We always searched by the name in the headline reported on the player's page on transfermarkt.com. Likewise, in all cases, the name of the selling club used was not the official full name of the club (for instance, Real Madrid, Club Futbol), but rather the headline name on the club's official page on transfermarkt.com. This number is supposed to measure popularity of players in our data set as well as their potential marketing value for the new club. The expected sign is positive since the more popular a player is, the more value on the market he might be expected to have in reality.

Productivity measures the sum of goals and assists of a player the season before being transferred to a new club. Although assists are often not given the same importance as goals, they are equally important for the productivity of a team, as generally there would not be a goal without a pass preceding it. Moreover, while some attacking players score more, some provide assists, and both types of players are equally valuable to the team. As a result, we decided to create one variable from goals and assists, giving them equal importance. The expected sign is positive — the more goals and assists the player has, the more expensive he will be on the transfer market.

Lastly, race is a dummy variable with two possible values — 1 if a player is originated from the African continent or is a descendant of people from Africa (further referred to as "black"), and 0 if a player is "non-black," which includes players that do not fit the definition of the Race variable having value of 1. The rationale behind this grouping is the reason that it was used similarly in previous studies by Szymanski (2000) and Reilly and Witt (1995). The expected sign is positive, since it was suggested by Szymanski (2000) that enjoy significant premium in team performance by signing black players (p. 602).

All of the independent variables are expected to have a positive linear relationship with the dependent variable, the fee paid for the player.

Data

The main source of data for this study was the website transfermarkt.com which is widely recognized as one of the most reliable sources for soccer data and has been used as a data source for numerous past studies (Bryson et al., 2013; 2014; Franck and Nuesch, 2012; Herm et al., 2014; Muller et al., 2017). All data was obtained in the month of October 2017, and the data did not include the current season which is reasonable since we need to measure player's performance, productivity, and popularity before the transfer, not after. For the GoogleResults variable, we used Google.com, and the data was collected on November 7, 2017.

There are 49 players in our sample, and all made a transfer in the 2017 summer transfer window which took place between July 1,t 2017 and August 31, 2017. All of the players made a move to one of the top-15 most valuable and richest clubs in the world (Forbes.com, 2017), namely: Manchester United, Barcelona, Real Madrid, Bayern Munich, Manchester City, Arsenal, Chelsea, Liverpool, Juventus, Tottenham Hotspur, Paris Saint-Germain, Borussia Dortmund, AC Milan, Atletico de Madrid, and West Ham United. All of these clubs compete in the five prominent European soccer leagues which are in England, Spain, Italy, Germany, and France. The descriptive statistics are as follows:

Table 2: Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Fee (mil. €)	49	222.0	.0	222.0	32.653	34.3906
Contract left (years)	49	4.0	.0	4.0	2.306	1.3103
MV at time of transfer	49	99.65	.35	100.00	20.9153	16.02200
Google results	49	1217.8200	2.1800	1220.0000	61.141633	173.5696119
Product	49	31	0	31	8.84	7.641
Race (1 = black, 0 = non-black)	49	1	0	1	.27	.446
Valid N (listwise)	49					

There were four main conditions to include a player in the data set. First, the player had to play in a senior league prior to the transfer; this allowed an accurate representation of his productivity compared to other players. Second, the transfer fee had to represent the true price of the player in the summer transfer window. This means no buyback clauses or no fees already agreed on for already loaned players waiting to be transferred. Third, transfermarkt.com had to have a record of the player's previous contract length as well as his market value; in other words, we had to have complete data on the player to include him in the data set. Lastly, due to the sample size, we were unable to include goalkeepers at all since there were very few of them in the data set.

There were some irregularities in the data set. The players who transferred "for free" also had expired contracts from their previous club, which means that both variables "fee" and "contractleft" are 0 in that instance. However, they still had a market value and a record of their previous season's performance. There was one obvious outlier which is probably very apparent to all the soccer fans who are familiar with the summer transfer window. Neymar transferred for an astronomical fee of 222 million euros while the second highest fee in the transfer window paid was "only" 105 million euros for Ousmane Dembele from Borussia Dortmund to FC Barcelona. However, since Neymar is a superstar and one of the most valuable players in the world, his inclusion in the model makes sense, especially in a data set full of superstars. The complete list of players included is in the appendix.

Empirical Model and Results

We did two regression runs in our study. The first one, which also included aforementioned outlier Neymar, and the second one which did not include him in the regression. The results are as follows.

Table 3: Regression Results

	Regression 1	Regression 2
Constant	-9.942	-9.874
MarketValue	.529** (.181)	.534** (.187)
ContractLeft	7.264*** (1.153)	7.243*** (1.178)
GoogleResults	.111*** (.014)	.106* (.046)
Productivity	.585* (.238)	0.591* (.245)
Race	10.581** (3.236)	10.686** (3.376)
F	112.085***	33.500***
R^2	.929	.800
Adjusted R^2	.920	.776
Data sources: transfermarkt.com, Google		
a. Dependent Variable: Fee		
b. Predictors: (Constant), Race, GoogleResults, ContractLeft, Productivity, MarketValue		

Table 3 shows the model summary. R^2 of 0.929 and adjusted R^2 of 0.92 indicate great closeness of the data to the regression line. In other words, approximately 92% of the variability in the dependent variable around its mean is explained by this model and the independent variables used. In the context of this analysis, it means that our model makes a relatively precise prediction of the fee for the player given the independent variables used in our linear regression. Table 3 shows our analysis of variance of the model. With an F-statistic of 112.085, the model's p-value is 0.000, which means that the whole model is indeed significant. In other words, our table shows that the relationship between the transfer fee and all the independent variables is statistically significant at a >99% level.

When it comes to independent variables, all seem to be significant at the 98%

level with ContractLeft and GoogleResults showing as the most significant variables with the highest t-values — 6.302 and 8.043 respectively. The lowest significance amongst all variables is 98.2%, suggesting that productivity is the least significant out of all, but is still 3.2% above than the 95% critical t-value, making the variable still statistically significant.

The intercept of -9.942 says that when all dependent variables hold a value of zero, the player has a negative value. However, since our conditions clearly state that all the buying clubs are in top-15 most valuable and richest clubs in the world, and only the players with at least a year of professional career are included in the data set, a negative number in the intercept is not a problem. It is highly improbable that the player would hold a value of zero in all five independent variables, especially popularity (Google results) or his market value.

The beta for market value indicates that for every million that the soccer community from transfermarkt.com predicts, the value of the player rises by 0.529 million in real transfer fee. This number suggests that the community might have a real power over the transfer fee of the player, as Herm et al. (2014) explained, since the number is used in the actual negotiation between the buying and selling clubs. The potential implication for the soccer managers around the world might be huge — the fans can influence the transfer.

With every year of the contract, the player's price rises by 7.264 million. This heavily supports our prediction that the buying club is in a position of disadvantage since the premium for every year of the player's current contract is more than 7 million euros. In other words, it seems to be very advantageous for a club to offer long-term contracts to its players. If the selling club decides to hold on to the player instead of selling him, the premium for doing that is 7.264 million per year. On the top of that, provided that the selling club will not bring in a replacement, wages are a significant expense as well. Per every 10,000 Google results, the player's price increases by 0.111 million. Popularity seems to significantly increase a player's cost on the market. With a mean value of approximately 611,416 Google searches, an average player's cost in one of the richest clubs in the world rises by 6.78 million euros. Mass media and reporting seem to have a real impact on the future transfer fee of the player.

The last independent variable in our regression equation is race. If the player is black, his transfer fee increases by 10.581 million, which supports Szymanski's (2000) findings, but in a different way than maybe expected. Szymanski concluded that the clubs with more black players perform better, and those clubs who do not have as many black players pay the premium for not being successful by not having black players in the team. In our model, the situation seems to be the other way around — and the clubs realize that black players cost more possibly because of their better performances or physical ability (which, however, we cannot confirm by our model).

Exclusion of Neymar produced surprising results. To compare, the following empirical model regression equation includes Neymar, the record-holder in the highest transfer fee in history, in the data set:

$$Fee = -9.942 + 0.529MarketValue + 7.264ContractLeft + 0.111GoogleResults + 0.585Productivity + 10.581Race + \varepsilon,$$

and the following regression equation does not include Neymar in the data set:

$$Fee = -9.874 + 0.534MarketValue + 7.243ContractLeft + 0.106GoogleResults + 0.591Productivity + 10.686Race + \varepsilon.$$

The R^2 decreased from 0.929 to 0.8, and the adjusted R^2 from 0.92 to 0.776. P-value stayed the same, but F decreased by a great amount from 112.085 to 33.5. The model still makes a prediction of the fee and is still significant. The individual Beta values did not change outstandingly; the constant value increased, market value increased as well, contract left and Google results decreased, and both productivity and race increased; none of the changes deform the model. However, the t-value for Google results decreased from 8.043 to 2.309, which is surprising, and the significance decreased somewhat as well.

These changes might be explained by the fact that although Neymar is an outlier, he (and his values of independent variables) fit the model perfectly. The values did not increase or decrease individual Betas significantly. Thus, inclusion or exclusion of Neymar did not influence the model itself by great measures. A possible interpretation is that although his transfer fee might have been marked absolutely illogical and outrageous in general opinion of the soccer world, fans, managers, or coaches, his transfer value makes sense.

Conclusion

Based on a linear regression analysis of 49 players transferring to the richest and most valuable soccer clubs in the world, this paper has shown that the high amounts of money spent on soccer players in the top European teams is justified either by their productivity or their popularity. Race of the player and the years left on his contract seem to play a significant part in explaining the transfer fees as well. Moreover, the community valuation predictor and its significance to the regression equation suggests that the opinion of the soccer community really matters, and that the community prediction is a strong indicator of a real market value of the player, or how much any club is willing to pay for him, in the transfer fee.

The results are consistent with previous research. As Herm et al. (2014) pointed out, “The community has become the main source for reporting market values in the media and has a strong impact on the real sports economy. In fact, it is used in real market transactions and wage negotiations, indicating the power of crowd wisdom in the sports management context” (p. 490). The results seem to suggest that this is true, and the soccer community has much more power over the transfer fees than people that are a part of it might think. Not only does the market value on transfermarkt.com successfully predict the transfer fees, it also plays an important role in an actual transfer.

As several researchers pointed out, one variable, years on the contract left, seems to be crucial in any regression model predicting transfer fees (Carmicheal et al, 1999; Frick, 2007), but the details of these contracts often remain private (Carmicheal et al, 1999). This fact has made a complete analysis with this variable almost impossible in the past, especially when players from lesser-known leagues are included in the data set. However, this study analyzed only the players coming into the biggest clubs in the world, and the data on such players' contracts are usually available to the public. Our research suggests that the time until contract expiration is significant and directly influences the transfer fee the selling club demands from the buying club.

One of the most crucial findings of this study is the fact that we can confirm Garcia-del-Barrio's and Pujol's (2004; 2007) theory that popularity, as measured by the number of results on Google, influences the economics of soccer. Since then, the variable "Google results" as a predictor of popularity has been used in several studies (Lehmann & Schulze, 2008; Herm et al., 2014; Muller et al., 2017) in some form, and all of these studies found a connection between the player's popularity and his transfer fee and/or salary. Our findings support this conclusion with the number of Google searches being one of the main predictors of the transfer fee for the player in our model.

If the soccer community thought that the amount of money paid for Neymar did not make sense, our study hints that they might have been mistaken. With the coefficients having almost the same values with or without his inclusion in our sample, and with both lower R^2 and lower adjusted R^2 model significance without his inclusion, the model seems to justify Neymar's high transfer fee. In addition, the model itself seems to be weaker if Neymar is excluded as hinted by the F-test. The model seems to be stronger with his inclusion even though he is technically a sole outlier in our data set.

To confirm the findings of this study, more research should be done on the topic, ideally one combining a data-driven approach and a community valuation approach in the model for the richest and the most valuable clubs in the world. In addition, a study with a larger sample size would be welcome. However, since the top clubs buy only a few players each year, such a data set would be rather difficult to obtain, especially as there is clearly inflation present (Dobson & :Gerrard, 1999) in the soccer transfer market making it difficult to compare data from several transfer windows. This fact also hints at even higher transfer fees in the future. Neymar's record $250 million transfer might not be unusual anymore.

References

Brandes, L., & Franck, E. (2012). Social preferences or personal career concerns? Field evidence on positive and negative reciprocity in the workplace. *Journal of Economic Psychology*, 33 (5), 925-925.

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

Bryson, A., Rossi, G., & Simmons, R. (2014). The migrant wage premium in professional football: A superstar effect? *Kyklos*, 67 (1), 12–28.

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.

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

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

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

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

Frick, B., Pietzner, G., & Prinz, J. (2007). Career duration in a competitive environment: The labor market for soccer players in Germany. *Eastern Economic Journal*, 33 (3), 429-442.

Frick, B. (2011). Performance, salaries, and contract length: Empirical evidence from German soccer. *International Journal of Sport Finance*, 6, 87-118.

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

Garcia-del-Barrio, P., & Pujol, F. (2004). Pay and performance in the Spanish soccer league: Who gets the expected monopsony rents? Working Paper 05/04 available at: www.unav.es/econom/investigacion/working/wp0504.pdf.

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.

Goal.com. Neymar to PSG, what a crazy world football has become - Fans divided over world record transfer. *www.goal.com*, 3 Aug. 2017, www.goal.com/en/news/neymar-to-psg-what-a-crazy-world-football-has-become-fans/sqkeuif4g6701faa3nneggake.

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

Lehmann, E., & Schulze, 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.

Lucifora, C., & Simmons, R. (2003). Superstar effects in sport: Evidence from Italian soccer. *Journal of Sports Economics*, 4 (1), 35-55.

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.

Pedace, R. (2008). Earnings, performance, and nationality discrimination in a highly competitivelabor market as an analysis of the English professional soccer league. *Journal of Sports Economics*, 9 (2), 115-140.

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

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

Szymanski, S. (2000). A market test for discrimination in the English professional soccer leagues. *The Journal of Political Economy*, 108 (3), 590-603.

Tavana, M., Azizi, F., Azizi, F., & Behzadian, M. (2013). A fuzzy inference system with application to player selection and team formation in multi-player sports. *Sport Management Review*, 16 (1), 97-110.

Wilson, D., & Yung-Hsiang, Y. (2003). Nationality preferences for labour in the international football industry. *Applied Economics*, 35 (14), 1551-1559.

Appendix

Appendix A: Data Set (sources: transfermarkt.com, google.com)

Fee	MarketValue	ContractLeft	GoogleResults	Productivity	Race
222	100	4	1,220.0000	24	0
105	33	4	75.9000	18	1
84.7	50	2	207.0000	31	1
57.5	13	4	39.0000	5	1
53	40	2	67.3000	31	1
51	30	4	59.2000	5	0
50	40	3	43.6000	17	0
44.7	35	2	102.0000	8	0
42	35	2	75.4000	26	0
42	45	4	76.3000	3	0
41.5	22	3	16.3000	13	1
40	25	4	14.3000	8	1
40	30	2	16.2000	15	0
40	15	3	20.0000	9	0
40	16	2	22.1000	3	1
38	22	1	134.0000	9	1
38	22	4	35.7000	21	0
37.9	9	4	34.5000	2	0
35	25	3	22.7000	2	1
35	22	4	20.9000	1	0
30.5	20	4	15.9000	7	1
30	15	4	13.6000	4	0
30	16	4	103.0000	3	0
26	25	3	37.3000	12	0
25	15	2	54.7000	3	1
25	8.5	2	6.3000	6	0
25	10	4	7.4500	13	0
22.3	15	3	24.3000	11	0
22	20	2	14.0000	9	0
20	30	1	53.4000	8	1
20	15	2	5.3100	11	0
20	25	2	11.8000	4	0
18	18	1	7.5700	2	0
18	17	2	14.0000	5	0
17.8	18	1	46.3000	14	0
17	20	1	17.0000	7	0
16.5	10	3	15.9000	4	0
15.1	6	1	23.6000	16	0
12	8.5	1	17.7000	2	0
12	15	1	11.1000	8	0
12	16	1	7.6400	2	0
10.5	6	2	4.7100	3	0
9	7	1	7.6700	3	0
7	3.5	1	4.5900	2	1
2	0.35	3	2.1800	0	0
0	15	0	15.5000	8	0
0	6	0	103.0000	4	0
0	8	0	9.1200	9	0
0	7	0	38.9000	2	0

Appendix B: The List of Players in the Data Set

Neymar da Silvo Santos	Alex Oxlade-Chamberlain	Chicharito
Ousmane Dembélé	Andre Silva	Lucas Biglia
Romelu Lukaku	Danny Drinkwater	Dani Ceballos
Benjamin Mendy	Antonio Rüdiger	Fernando Llorente
Alexandre Lacazette	Victor Lindelöf	Mattia De Sciglio
Kyle Walker	Nélson Semedo	Mahmoud Dahoud
Bernardo Silva	Theo Hernández	Ömer Toprak
Nemanja Matic	Danilo	Rodrigo Bentancur
Mohamed Salah	Vitolo	Andrew Robertson
Leonardo Bonucci	Serge Aurier	Jeremy Toljan
Corentin Tolisso	Davide Zappacosta	Matheus Pereira
Davinson Sánchez	Andrea Conti	Sead Kolasinac
Federico Bernardeschi	Marko Arnautovic	Dani Alves
Paulinho	Hakan Calhanoglu	Sebastian Rudy
Tiemoué Bakayoko	Blaise Matuidi	Pablo Zabaleta
Ricardo Rodriguez	Maximilian Philipp	Romelu Lukaku
Chicharito	Niklas Süle	Benjamin Mendy
Lucas Biglia	Mateo Musacchio	

Appendix C: Regression 1 Output

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.964[a]	.929	.920	9.6995
a. Predictors: (Constant), Race, GoogleResults, ContractLeft, Productivity, MarketValue				

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	52724.668	5	10544.934	112.085	.000[b]
	Residual	4045.414	43	94.079		
	Total	56770.082	48			
a. Dependent Variable: Fee						
b. Predictors: (Constant), Race, GoogleResults, ContractLeft, Productivity, MarketValue						

Table 4: Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-9.942	3.382		-2.940	.005
	MarketValue	.529	.181	.247	2.931	.005
	ContractLeft	7.264	1.153	.277	6.302	.000
	GoogleResults	.111	.014	.561	8.043	.000
	Productivity	.585	.238	.130	2.461	.018
	Race	10.581	3.236	.137	3.270	.002
a. Dependent Variable: Fee						

Appendix D: Regression 2 Output

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.894[a]	.800	.776	9.8123
a. Predictors: (Constant), Race, ContractLeft, Productivity, GoogleResults, MarketValue				

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16127.045	5	3225.409	33.500	.000[b]
	Residual	4043.852	42	96.282		
	Total	20170.897	47			
a. Dependent Variable: Fee						
b. Predictors: (Constant), Race, ContractLeft, Productivity, GoogleResults, MarketValue						

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-9.874	3.463		-2.852	.007
	MarketValue	.534	.187	.290	2.856	.007
	ContractLeft	7.243	1.178	.455	6.148	.000
	GoogleResults	.106	.046	.204	2.309	.026
	Productivity	.591	.245	.211	2.409	.020
	Race	10.686	3.376	.232	3.165	.003
a. Dependent Variable: Fee						