ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis [Economics and Business Economics]

English Tax in the Premier League

Name student: Wouter van Puffelen

Student ID number: 573679

Supervisor: Prof. dr. ir. Jan van Ours

Second assessor: xxx

Date final version: 27-06-2023

I. Introduction

The transfer market in football is a fascinating phenomenon; twice during the season it opens for clubs to buy or sell their assets: the players. The prices that are paid, the transfer fees, have evolved over the years just as prices of products in other markets. In the summer of 2017, Paris Saint Germain paid FC Barcelona 222 million euros for the services of Neymar (Transfermarkt, 2023). This is still the highest transfer fee ever to be recorded in the history of football. With fees this high, many are led to wonder as to how the fees are build up. Most football fans would agree that Neymar is one of the most established names in modern football and was so already during his time at FC Barcelona, so it does not surprise that big money was involved in his transfer. The biggest spenders in the market are the English Premier League clubs. The Premier League has an estimated market value of 10,42 billion euros (Transfermarkt, 2023). Which is more than double the 4,78 billion euros of the Spanish La Liga, which is estimated to be second by Transfermarkt (2023). It is therefore no surprise that some of the biggest transfers are moves to or within the Premier League. One of the moves that is interesting is the move of Harry Maguire to Manchester United Football Club in 2019. He moved from Leicester City Football Club for 87 million euros (Transfermarkt, 2023). The transfer was ridiculed by many, including British ex-football player and pundit Paul Merson, who labelled the fee as "ridiculous". This is where the so-called 'English Tax' comes into the picture. It is believed that English clubs pay more for English players simply because of their nationality. Many transfers over the years have been accused by the internet for containing some sort of English Tax. Some articles online touch on this phenomenon. Some say that the tax is a myth and that it is all attached to other factors like age and experience (sub. Editors, 2022). Diego Pereira also touched on the myth of the English Tax on a column for 'over the line sports' and believes that there is indeed an English Tax after comparing some of the summer moves of 2021. Another article argues that there is not so much an English Tax, but rather a 'Premier League Tax' (*The Misunderstanding of the 'English Tax' in the Transfer Market*, 2022). Meaning that Premier League teams tend to overvalue players. However, there has never been academical research on this topic and thus the English tax remains a myth. I aim to find evidence by studying recent transfers, to answer the question: Do English top clubs pay higher transfer fees for English players?

I believe this research is academically relevant as it adds to the theory of whether sports teams are win maximisers or profit maximisers. When teams are win maximisers and they behave in a rational manner, then it makes no sense that teams pay higher fees just for a specific nationality. By using this argumentation and reversing it, we can conclude that teams that pay higher fees just because of nationality are in fact profit maximisers. English players at English clubs may be preferred by fans and sponsors, making it lucrative to invest in them.

This research is socially relevant as it may influence the behaviour of football clubs and their sporting directors in negotiations and valuations of players. And it hopefully will solve the myth of the English Tax.

II. Theory

Recent developments have seen economic theory being applied in sports in many ways. For many economic theories that have been developed over the years, strict conditions are in order: firms focus on profit, economic agents are rational, everybody acts in their own best interest, and many more. Taking these restrictions and applying them to the world of sports is not so straightforward. Are sports teams really profit maximisers or do they maximise the utility of the people involved with the team, such as the fans? Rottenberg (1956) was one of the first academics that applied economic theory to the world of sports. In this application, he assumed that sports teams are just like any other rational business and therefore maximize their profits. He argues that when teams want to maximise their wins and do whatever it takes to achieve this, the profits in the league as a whole will decrease due to the lack of competition; people want to see exciting games, not one-sided ones. Rottenberg (1956) stated: "in baseball no team can be successful unless its competitors also survive and prosper sufficiently so that the differences in the quality of play among teams are not "too great."" Rottenberg based his whole analysis on the assumption that teams maximize their profits. This assumption was questioned by Sloane (1971) after he focussed on clubs in the English Football League, which is also the niche which we will be looking at, and concluded that these teams can have four different targets. The first potential target is profit maximisation, but this seems highly unlikely since most clubs make losses (Sloane, 1971). The second possible target is security, which seems more plausible for teams in the lower divisions which operate at heavy losses. These teams often create funds for long-term survival of the club by selling their top players (Sloane, 1971). Thirdly, clubs can aim to maximise sales, or rather attendance. New players would be brought in when the club notices a decrease in attendance (Sloane 1971). The final objective that can be pursued by clubs is the maximization of utility. Where it is difficult to see whose utility is maximised in regular businesses, football clubs and the people involved in them have more aligned preferences and managers have little influence on the business side of the club and focus solely on performance, meaning that the utility that is being maximised is that of the fans (Sloane 1971). Sloane also argues in his paper that utility maximisation is not unconditional: clubs have different departments with different targets. The managers will maximise performance, but there are also financial constraints put up by directors in order to ensure the existence of the club: clubs cannot spend unlimited amounts of money to acquire the best possible teams; the board of directors will not allow this.

Neale (1964) argued that the goal of profit maximisation could be justified by assuming that the individual clubs are not the firms, but rather the league which these individual clubs form together. The league is the decision-maker and maximises profits, which will be divided over the clubs. Sloane (1971) rejects this assumption for the Football League since the clubs in this league take their own decisions within the rules that have been set by the league.

Later, empirical evidence was used to identify the goals of sports teams. Ferguson et al. (1991) use a simple model in which they analysed club behaviour by looking at the attendance and the ticket pricing of NHL teams over three seasons. The results from their test offer support for the profit-maximisation theory (Ferguson et al., 1991). A study by Atkinson et al. (1988) on the NFL found no convincing evidence for profit maximisation as main goal for owners and even suggested that the motivation of owners could be explained by utility maximisation. They provide support for this hypothesis by measuring salaries that are higher than the marginal revenue product (Atkinson et al., 1988). Késenne and Pauwels (2006) doubt these types of analysis involving ticket pricing, since Fort and Quirk (2004) state that both profit maximisers and win maximisers use the same pricing rules and therefore Késenne and Pauwels (2006) use a model that contains not one decision variable, which most of the studies use, but two decision variables: ticket pricing and talent demand. From their model they conclude that win maximising teams hire more talent and they ask more money for tickets. However, they still are unable to label the targets of sports teams in general. Garcia-del-Barrio and Szymanski (2009) also try to identify the goal of sports teams, specifically English and Spanish football teams. They use data on performance in order to label the behaviour of these clubs as profit-maximising or win-maximising. Their findings support that these clubs act as win maximisers subject to a zero-profit condition (Garcia-del-Barrio & Szymanski, 2009). They do, however, acknowledge that the period which they studied was, 1994-2004, was one of significant change (Garcia-del-Barrio & Szymanski, 2009). They conclude that the behaviour of win maximisation, which seemed most plausible for this period, may not be observed permanently.

It is important to identify whether clubs maximise their profits or their wins as both assumptions may lead to very different outcomes when analysis is done. Fort and Quirke (1995) find that salary caps can be expected to keep weaker teams financially viable while also improving the competitive balance in a league where clubs maximize profits. Besides that, in a profit maximizing model, demand for talent will drop together with the wages when revenue sharing is increased (Rascher, 1997). Rascher also finds in his models that in leagues with pure profit maximization, all teams would be in favour of increasing the revenue sharing and lowering the salary cap. Vrooman (1997) shows that players have an increased effort in win maximising leagues and the competitive balance is also greater. Szymanski (2003) shows that revenue sharing in a profit maximising league will worsen the competitive balance, which he later supported together with Késenne (2004) when they showed that

this statement is true for any revenue function that is concave. Késenne (2000) addressed that in the case of a win maximising league, revenue sharing will improve the competitive balance in general. Ford and Quirke (2004) show in their paper that, in a league, when teams target to maximise their win percentage, the talent price and talent demand is just as great, if not greater, as when the owners of the clubs are profit maximisers.

It could be argued that the English Tax is rational, even from a win-maximising point of view. This would be because looking for homegrown talent may come with lower scouting costs, allowing teams to spend more on the player fee. When teams look abroad, it is likely that it is more costly to put up a scouting network. When you have two players with all the same characteristics, but one is foreign and had to be scouted abroad and the other one is from the same nationality as the club, then scouting costs are lower. The costs of buying the foreign player would be:
Costs = Transfer Fee + Scouting Costs Abroad.
While the cost function of buying the homegrown player would look the following:
Costs = Transfer Fee + Scouting Costs Domestic.
As I have explained, Scouting costs abroad > Scouting costs domestic. So, clubs are willing to pay a little extra for the transfer itself in order to secure the player, as long as:
Scouting Costs Domestic + Extra Fee – Scouting Costs Abroad =< 0.
One could say that this 'Extra Fee' is the English Tax. The reasoning is rational, but this reasoning should apply to all clubs, and not just English clubs, therefore this theory will only suffice when in other countries a same sort of tax can be found: a Domestic Tax.
For profit maximisers it could be worth it to pay a higher transfer fee when domestic players attract more sponsors or more sales in merchandise, this makes it rational for teams with these goals to invest the extra amount to attract these players. This theory also does not need the support of other competitions as sponsors and sales can vary between countries.

Another assumption that would rationalize the preference for homegrown talent would be the barrier of language and culture. Communication is key in many situations, and this does not differ on the football pitch. The cohesion could also play a big role in performances. Although I do acknowledge that it might play a role, especially for the teams that are near the bottom of the table, I also believe that quality of the player is more important for football teams, and this is backed by the fact that the top teams in modern football have a huge diversity in player nationalities. Also, as is the case in the theory of scouting costs, when differences in fees are drive by language and cultural barriers, one would expect to find the same price differences in different competitions.

Then there is yet another possible explanation for an English Tax. The Premier League enforces the so called 'homegrown rule'. This rule states that the squad, which exists of 25 players in total, must at least contain 8 "home grown players". The Premier League (2016) describes "home grown players" as follows: "A "Home-Grown Player" means a player who, irrespective of nationality or age, has been registered with any club affiliated to The Football Association or the Football Association of Wales for a period, continuous or not, of three entire seasons, or 36 months, before his 21st birthday (or the end of the season during which he turns 21)." This rule increases the demand for English players by nature, and therefore it would be no surprise that there is an English tax, especially for teams that target win maximisation. This is because they do not only want English players, but the players also need to be of good enough quality, thus the demand increases, while the supply decreases.
This rule is not enforced by the French Ligue 1, so this could be the main cause of an English Tax that is not automatically a Domestic Tax. However, the UEFA does enforce this same rule for all their competitions, and since I only study the top teams of each division that play or aim to play European football, it is likely that the French teams in my dataset also create transfer policies by these rules.

From past research it is still difficult to label the clubs in my dataset as profit maximisers or as win maximisers. The research of Garcia-del-Barrio and Szymanski (2009) comes the closest to describing our dataset as their subjects were also European football clubs. This leads me to believe that the clubs in my dataset are win-maximisers and therefore I expect that there cannot be an English Tax as this does contradict the club's main goal. And if there is indeed evidence for an English Tax, then this should be generalized to other leagues for it to be considered rational. It would then no langer be seen as an English Tax, but more as an extra fee to compensate for the lack of scouting costs.

III. Hypotheses

Following the studied literature, and taking the practice into consideration, I have formed 2 hypotheses that will be central to the research.
H1: English top teams, on average, do not pay more money for English players.
This hypothesis follows the literature of Garcia-del-Barrio and Szymanski (2009) where they state that European football clubs are best described as win maximisers, and an English tax would be irrational from this target's perspective.
H2: When there is an English Tax, then this tax should be generalized to a Domestic Tax.
This hypothesis results from the theoretical settings discussed previously, where scouting costs, the "homegrown rule" and language and cultural differences are taken into consideration. In this case, when teams are win maximisers, they are still considered rational. It is when there exists variation

per country on whether there is a Domestic Tax or not, that the win maximisation target can be questioned on its goodness of fit.

IV. Method

In order to expose the English Tax, I will study incoming transfers of the traditional top 6 football clubs in the Premier League: Arsenal FC, Chelsea FC, Liverpool FC, Manchester City FC, Manchester United FC and Tottenham Hotspur FC. I will look at a period of five years: 2018-2023. I have chosen for this recent period as I think more recent data in this department is more relevant for the ever so fast changing football transfer market. Besides that, over time more information about transfers has been recorded and thus it makes it more attractive to research as more data is available.
Not only will I study the incoming transfers of the top 6 teams in the Premier League, but I will also look at the incoming transfers of 6 top clubs in the French Ligue 1: LOSC Lille, AS Monaco, Olympique Lyonnais, Olympique Marseille, Paris Saint Germain and Stade Rennais. I will use this league to compare to the transfers in the Premier League and see whether this English Tax could just be seen as a 'Domestic Tax' in general. I have chosen the French league in particular, because the French people are widely seen as a very nationalistic people, and this may affect the level of this 'Tax'. It would be interesting to observe whether the Tax in more nationalistic countries is higher.

For the empirical research that I conduct, I will use an OLS regression model using the logarithm of the transfer fees as the dependent variable and using various independent variables. As I believe the transfer fee is built up out of many components, there will be many variables to control for. I will control for:

- Year: The dummy of the year in which the transfer happened.
- Age: The age of the player at the time of the move (dummies for different age groups).
- Winter Transfer: Dummy with value 1 when the transfer happened in the January window.
- Contract: Remaining months of contract at the old club.
- Nationality: A dummy that will be either 1 when the player is English/French depending on the league that is studied and 0 when the nationality is different to that of the domestic league.
- Competition: A dummy that will be 1 when the move is within the same competition and 0 otherwise.
- Quality: The rating given by the most recent FIFA game at the time of the transfer.
- Form: The rating of the most recent season on whoscored.com at the time of the transfer.
- Minutes played in competition: in the season on which 'Form' is based.
- Total minutes played: in the season on which 'Form' is based.

- Position: four dummies; Forward, Midfielder, Defender and Goalkeeper.

I will use two different datasets: one of the incoming transfers in the Premier League and one of the incoming transfers in the Ligue 1. With each of the datasets, I will use STATA (2019) to run the multivariate regressions with all of the above variables.

The first regression model, that focusses on the Premier League transfers, will look as follows:

(I) $Log(Transfer\ Fee)_{it} = \beta_0 + \beta_1 Winter\ Transfer_{it} + \beta_2 Contract_{it} + \beta_3 English_{it} + \beta_4 Competition_{it} + \beta_5 Quality_{it} + \beta_6 Form_{it} + \beta_7 Minutes\ in\ Competition_{it} + \beta_8 Total\ Minutes_{it} + \beta_9 Forward_{it} + \beta_{10} Midfielder_{it} + \beta_{11} Defender_{it} + \beta_{12} Age22-26_{it} + \beta_{13} Age27-31_{it} + \beta_{14} Age32_{it} Plus + \varepsilon_{it}$

The second regression model will be similar, only with some slight changes:

(II) $Log(Transfer\ Fee)_{it} = \beta_0 + \beta_1 Winter\ Transfer_{it} + \beta_2 Contract_{it} + \beta_3 French_{it} + \beta_4 Competition_{it} + \beta_5 Quality_{it} + \beta_6 Form_{it} + \beta_7 Minutes\ in\ Competition_{it} + \beta_8 Total\ Minutes_{it} + \beta_9 Forward_{it} + \beta_{10} Midfielder_{it} + \beta_{11} Defender_{it} + \beta_{12} Age23-26_{it} + \beta_{13} Age27-31_{it} + \beta_{14} Age32Plus_{it} + +\varepsilon_{it}$

Besides these models, I will also use a model that focusses on absolute outcomes rather than relative outcomes. The model will use the same independent variables as model I and model II, but as a dependent variable it will use the absolute transfer fee instead of the logarithm of the transfer fee. For the use of these models, it is important that the data are in line with multiple assumptions. I will discuss these assumptions and prove their validity for my data in the appendix, but for now it is important to know that the assumptions hold.

V. Data

As previously mentioned, I will study the incoming transfers over the last 5 years of the clubs stated in the methodology sector. To gather the data needed to create the independent variables stated in the methodology sector, I have used a variety of online websites. For each transfer, I collected the fee; the year; age of the player; whether it was a winter transfer; the remaining contract; the competition; nationality of the player; and the position of the player via Tranfermarkt.nl. Transfermarkt (2023) is a website that gathers data on many football related topics, including transfers. Each transfer can be studied individually, and details of the transfers are provided. The 'form' variable is created by taking the ratings on Whoscored.com of the season before the transfer. Whoscored provides ratings to many footballers in various leagues, but unfortunately not all of them. Some of the moves in the dataset were moves from leagues on which whoscored had no data, I had to remove these moves from the dataset in order to run the regression with only accurate information. The ratings that are provided on Whoscored are separated into competitions; I focus only on the domestic competition. The ratings are on a scale of 1-10. For the 'rating' variable I aim to summarize the overall quality of the player and therefore do not look only at the last season as I do for the 'form' variable. I took the ratings created by the popular video game series "FIFA", produced by EA Sports. FIFA is a game that is released every year and thus every year the ratings, which are on a scale of 1 to 100, change. These changes are, however, not so sensitive to form: players will not drop from 93 to 73 after one bad season. I used the FIFA rating of the game before the move (e.g. Mbappé moved from AS Monaco to PSG on a permanent basis in the summer of 2018, I used his rating on FIFA 18, which was 83). The FIFA Ratings I retrieved from Futhead.com, Futwiz.com and Futbin.com. For players that moved in the January transfer window, I use the FIFA game that was released at the beginning of that season and I use the who scored rating of the first half of the season. The information I gathered on the minutes of playing time came from Footballdatabase.eu. I collected the playing time in the competition on which the Whoscored rating was based and the total amount of minutes played in the season in all competitions.

Table 1: Descriptive Statistics Premier League

Variable	Obs.	Mean	Std. Dev.	Min	Max
Year	104	2020.519	1.526	2018	2023
WinterTransfer	104	0.154	0.363	0	1
Age	104	23.788	3.627	17	36
TransferFee	104	38.477	26.964	1.7	121
RemainingContract	104	27.327	14.069	1	82
English	104	0.154	0.363	0	1
SameCompetition	104	0.212	0.41	0	1
FIFARating	104	77.404	7.2013	56	92
Form	104	7.04	0.403	6.2	8.5
MinPlayedComp	104	2070.856	843.839	113	3420
MinutesPlayed	104	3014.827	1066.001	424	5625
Forward	104	0.269	0.446	0	1
Midfielder	104	0.26	0.441	0	1
Defender	104	0.394	0.491	0	1
Goalkeeper	104	0.077	0.268	0	1

Notes: All position variables, the competition variable, the winter transfer variable and the English variable are dummies. The transfer fee is measured in million euros and the median is 32.75. The remaining contract is measured in months.

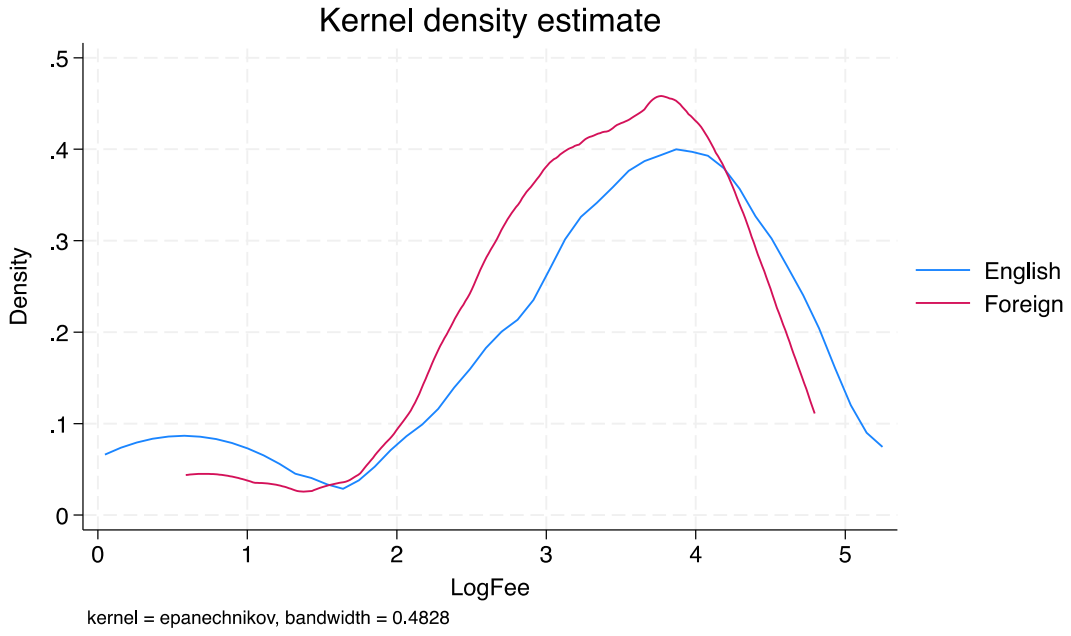

Figure 1: Kernel density estimate for Premier League players.

In table 1, the dataset for the Premier League is summarized with some descriptive statistics. We can see from this table that we observed 104 incoming transfers over the past 5 years and that 15,4% of those were transfers of English players and 21,2% moved within the same competition. We also see that most of the transfers were defenders, which is not surprising since in most line-ups, there are more defenders than other positions, so there is more demand for these players. That also explains why only 7,7% of the all the players that transferred were goalkeepers. We also observe that the biggest amount of money that was paid for a player was 121 million euros, while the lowest transfer fee paid was 1,7 million euros. Another interesting observation in table 1 is that, in my dataset, 15,4% of the transfers occurred during the January transfer window.

In figure 1, we take a first glance at the potential relative price difference between national players and foreign players in the case of the Premier League. Even though we do observe that the plot for the English players is skewed more to the right, which would indicate that English players are more expensive, we also see that the density of low prices is also higher than that of foreign players. Thus, from figure 1 it is still difficult to see whether on average there would be differences in the transfer fee.

Table 2: Descriptive Statistics Ligue 1

Variable	Obs.	Mean	Std. Dev.	Min	Max
Year	113	2020.071	1.522	2018	2023
WinterTransfer	113	0.133	0.341	0	1
Age	113	24.398	3.416	18	36
TransferFee	113	15.914	19.07	0.25	180
RemainingContract	113	25.354	12.906	5	52
French	113	0.283	0.453	0	1
SameCompetition	113	0.319	0.468	0	1
FIFARating	113	74.92	6.011	52	87
Form	113	6.852	0.312	6	7.54
MinPlayedComp	113	1774.159	867.736	17	3420
MinPlayedTotal	113	2420.77	1064.507	17	4906
Forward	113	0.31	0.464	0	1
Midfielder	113	0.301	0.461	0	1
Defender	113	0.301	0.461	0	1
Goalkeeper	113	0.088	0.285	0	1

Notes: All position variables, the competition variable, the winter transfer variable and the English variable are dummies. The transfer fee is measured in million euros and the median is 12. The remaining contract is measured in months.

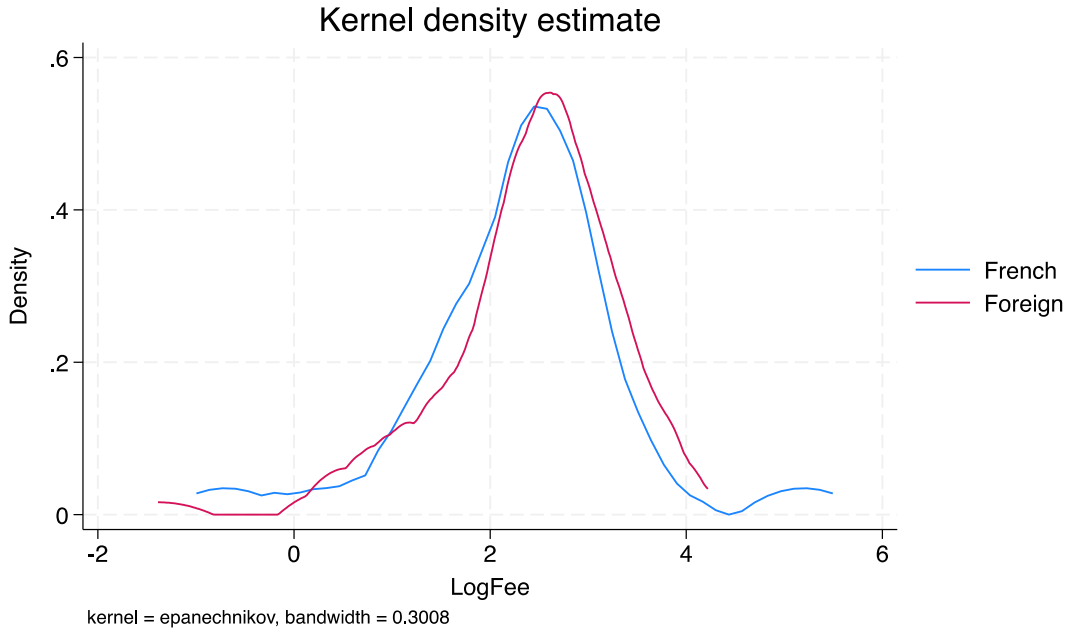

Figure 2: Kernel density estimate of the Log(Fee) for Ligue 1 players.

When we look at the descriptive statistics for the French Ligue 1 in table 2, we see that we have approximately the same number of observations. We observe that the highest transfer fee paid was the one PSG paid for Kylian Mbappé: 180 million euros. This is almost 50% higher than the record transfer of the Premier League over the past 5 years, but we also observe that the lowest fee paid is a lot lower than the lowest fee paid in the Premier League over the past 5 years.
In figure 2, we again plotted the kernel density estimate of the players that play in the French Ligue 1. As was the case for the Premier League in figure 1, we can observe some differences between national and foreign players, but it is difficult to draw any conclusions about average differences in relative price.

Table 3: Mean comparisons between Premier League and Ligue 1

Variable	**Premier League**	**Ligue 1**	**Diff**
Winter Transfer	0.15	0.13	0.02
			(0.05)
Age	23.79	24.40	-0.61
			(0.48)
Transfer Fee	38.48	15.91	22.56***
			(3.20)
Contract	27.33	25.35	1.97
			(1.84)
Nationality	0.15	0.28	-0.13**
			(0.06)
Competition	0.21	0.32	-0.11*
			(0.06)
Quality	77.40	74.92	2.48***
			(0.91)
Form	7.04	6.85	0.19***
			(0.05)
Minutes in Competition	2070.86	1774.16	296.70**
			(116.23)
Total minutes	3014.83	2420.77	594.06***
			(144.76)
Forward	0.27	0.31	-0.04
			(0.06)
Midfielder	0.26	0.30	-0.04
			(0.06)
Defender	0.39	0.30	0.09
			(0.06)
Goalkeeper	0.08	0.09	-0.01
			(0.04)

*Notes: standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

When we compare the means of the most important variables in the datasets in table 3, we can see that the samples are not completely similar, which is in line with expectations. We observe that French teams, on average, attracted more French players than English teams attracted English players over the past five years. Reason for this could be that the Premier League is widely seen as the biggest league in the world, and this attracts international players, so relatively fewer national players are bought. We also observe that the average transfer fees paid by the top teams in England were significantly higher than the fees paid by the French top teams. This again comes back to the fact that the Premier League is just a bigger competition worldwide and this brings more money on the table in terms of sponsorships and television rights (Transfermarkt, 2023), allowing these English teams to spend more money on transfers. It is therefore also no surprise that they buy players with significantly better quality and form as can be seen in table 3. Also, we observe that players that moved to the Premier League played significantly more minutes in the domestic competition, as well as all the competitions combined.

VI. Results

In this section I will present the results from the regressions that I performed with STATA.

Hypothesis 1

In the following regression table, I have presented the results of the linear regression performed using the data of the Premier League transfers over the 5 years with the different independent variables that are discussed in the methodology sector in column 1 and 3. These regressions results are aimed to identify an English Tax and therefore it touches on my first hypothesis. The first column sums all the variables used in the regression and the other columns report the estimated coefficients and the standard errors of these variables.

Table 4: Linear Regression using log(fee) and absolute fee as dependent variables

LogFee / Absolute Fee	(1) Premier League (relative)	(2) Ligue 1 (relative)	(3) Premier League (absolute)	(4) Ligue 1 (absolute)
Winter Transfer	-0.564*	-0.336	-9.084	-6.311
	(0.309)	(0.353)	(7.192)	(4.822)
Contract	0.03***	0.018***	0.903***	0.075
	(0.004)	(0.006)	(0.142)	(0.135)
Nationality	0.186	-0.386*	9.569**	-1.785
	(0.151)	(0.202)	(4.441)	(3.040)
Competition	0.216	0.329*	7.975*	5.821
	(0.15)	(0.19)	(4.565)	(5.718)
Quality	0.058***	0.082***	1.496***	1.581***
	(0.016)	(0.016)	(0.381)	(0.465)
Form	0.02	0.193	3.029	20.323*
	(0.25)	(0.353)	(7.891)	(10.324)
Min.in Comp.	0	0	-0.007	-0.011
	(0.00)	(0.00)	(0.006)	(0.008)
Min. in Total	0	0*	0.007	0.01
	(0.00)	(0.00)	(0.005)	(0.007)
Forward	0.739**	0.357	12.515	-5.998
	(0.371)	(0.434)	(10.678)	(6.309)
Midfielder	0.349	0.393	-1.059	-9.55
	(0.308)	(0.41)	(8.203)	(9.136)
Defender	0.067	0.267	-6.626	-5.159
	(0.321)	(0.425)	(7.476)	(6.072)
AgeGroup2	-0.162	-0.645***	-3.099	-13.531***
	(0.196)	(0.185)	(5.356)	(4.576)
AgeGroup3	-0.354	-1.034***	-8.567	-21.496***
	(0.221)	(0.086)	(7.245)	(7.837)
AgeGroup4	-1.885***	-1.612**	-46.13***	-35.89***
	(0.269)	(0.624)	(11.278)	(12.736)
2019	0.248	0.042	6.449	-6.021
	(0.317)	(0.228)	(7.547)	(6.646)
2020	-0.061	0.08	-7.8	-3.286
	(0.259)	(0.279)	(6.505)	(5.984)
2021	0.136	0.17	3.455	-1.186
	(0.136)	(0.234)	(7.7)	(5.179)
2022	0.141	0.053	0.032	-5.985
	(0.273)	(0.223)	(6.78	(6.613)
2023	0.687	0.088	7.863	-1.598
	(0.472)	(0.487)	(14.732)	(6.514)
Constant	-2.675	-5.858**	-126.435**	-227.221***
	(1.822)	(2.302)	(52.082)	(83.783)
Obs.	104	113	104	113
R-Squared	0.693	0.491	0.661	0.462

Notes: All position variables, year variables, the competition variable, the winter transfer variable, the age variables and the Nationality variable are dummies. Age group 1 = 17-21 for the Premier League and 18-22 for the Ligue 1, Age group 2 = 22-26 for the Premier League and 23-26 for the Ligue 1, Age group 3 = 27-31 for both and Age group 4 = 32+ for both. The transfer fee is measured in millions. The remaining contract is measured in months. Quality and Form are scores between 0 and 100 and 0 and 10 respectively. All other variables are continuous. Standard errors between parentheses.

**** p<0.01, ** p<0.05, * p<0.1*

In table 4, column 1, the relative results for the Premier League regression are reported. This regression shows how much the transfer fee increases or decreases relatively; so, the coefficients should be interpreted as percentages. For example, when we speak of a winter transfer, we see that, on average, the fee is 56.4% lower than for summer transfers. However, the significance level of this coefficient is above the 5%-level for which we aim, and therefore we cannot conclude that the coefficient is different from 0. When we further study column 1, we observe that there are only four significant coefficients. We find that attackers are, on average, more expensive than other positions. We also find that teams are willing to player higher fees for players of better quality, which is in line with expectations. Another conclusion we can draw from this regression column is that the price of a player relatively increases in the number of months that he has left on his contract. The final significant coefficient is the highest age group. We cannot really conclude much from this finding, ideally all age groups were significant, and we could find a decreasing or increasing pattern in the coefficients, but for now I can only say that a player's price drops significantly when they get to the point in their career in which they are close to retiring. This makes sense, players lose fitness ability and also longevity, this gives some understanding as to why the relative drop in price is so severe. For my research and especially my first hypothesis, it is interesting to see what happens with the nationality variable. In the log-linear regression, we see that the coefficient is not significant, and therefore we cannot conclude that there is a relative effect of one being English on the price tag. However, when we look at the absolute regression in column 3, we do find a positive and significant effect: on average Premier League teams paid 9.5 million euros more for players, because they were English. This finding does indicate the existence of an English Tax. However, when comparing the two models, I found that the log-linear model suits the data better than the absolute model. I came to this conclusion after comparing the Akaike Information Criterion (AIC): column 1 of table 4 had a value of 200.471 which was lower than the 907.016 of column 3. The same results were found for column 2, which scored 268.429 and therefore outperformed column 4 which scored 956.007. For my conclusions, I assume that the estimators in my model are complete. In other words: we assume that the price tag of the player is only influenced by the variables that are in the model. Now, even though I did not find convincing evidence for an English Tax, I will still check the French dataset to see how these results compare.

Hypothesis 2

Regression table 4 shows the results of the regression run with the transfer data of the French Ligue 1 in column 2 and 4. All the same independent variables as in the regression for the Premier League are used. These results are used for conclusions on hypothesis 2.

Following the results of the Premier League regression and the conclusion that it is unlikely that there is indeed an English Tax, I still want to check whether a tax can be found in the French competition. Based on the theory, I argued that if there was to be an English Tax, that it could most likely be generalized to a Domestic Tax. When we compare the log-linear regressions in table 4, column 1 and 2, we observe some similarities. Both have significant coefficients for contract and quality. Also, the signs of these coefficients are similar: positive. We do, however, also observe differences. The Ligue 1 regression has more significant coefficients. For example, we see that all age groups have significant coefficients, and we observe that the price tag relatively decreases in age. Also, the constant is significant, but this is not relevant as the coefficient cannot be interpreted in a logical manner for all four the regressions: it represents the transfer fee of the player if all other variables are 0; so it represents a player that has 0 months left on his contract, and these players are free agents and therefore have a transfer fee of 0 euros.

When we shift our focus to the variable of interest in column 2, we see that the coefficient here is statistically significant at a level of 10%. We also observe that the coefficient is negative, which would suggest that French players are actually relatively cheaper. This finding is likely the result of a higher supply. As seen in table 3, there were significantly more transfers of French players to the Ligue 1 than there were transfers of English players to the Premier League. So, it could be that there are simply more French players on the market that possess the quality to play for top football clubs. In economic theory, more supply will decrease the price. It can also be argued that the Ligue 1 is not as attractive for foreign players as the Premier League and therefore the supply of foreign players is lower in France than in England.

When we look at column 4, we observe no significant effect of nationality. In other words: where there does not seem to be an English Tax, there does not seem to be a French Tax either, especially in relative terms. I already concluded that I cannot reject my first hypothesis, and therefore it is also impossible to check and reject my second hypothesis. These conclusions are all based on the same assumption that the variables in the model are the only explaining factors of the dependent variable.

There are, as previously discussed, some differences in the results in table 4 for the different leagues. These differences occur because the two datasets differ significantly for certain variables as shown in table 3. It could be that the samples were not drawn randomly, meaning that the transfers in the

time period we studied, 2018-2023, vary from the average of all transfers. However, it can also be argued that these two leagues are just fundamentally different, and they will always be different, no matter what time period we study. When it is the case that these leagues are fundamentally different, then it would not hurt our analysis, and the differences we observe are just the differences between the two leagues. I suspect that it is indeed the case that these two leagues are different, and when looking at the overall UEFA coefficient ranking of all time, we also see that England ranks higher than France and also for the period of 2018-2023, England got the better of France (*UEFA Coëfficiënten Ranglijst 2022 / 2023 - Overzicht*, 2023). The ranking looks at so-called 'coefficients' that teams can earn by playing in in European competitions such as the Champions League. When teams perform well, they earn these points and can climb the overall ranking which comes with rewards like more qualification places for European competitions. In the European competitions, the top teams of all European leagues compete and since we use top teams in our research, I believe this is a good way to rank these teams. So, when using this ranking, we conclude that the differences we find in the dataset are differences between the leagues and not coincidental differences. This means that we can compare the two regressions and draw some conclusions from it.

Sensitivity analysis

In order to find how sensitive my models are to changes in the data; I have run another 6 regressions. For both leagues, I have run regressions for the specific positions apart from the goalkeeper position since there was too little data for this position in my dataset. I have presented the results in table 5. In table 5, we observe that the results vary a lot between the positions. For example, quality only has a significant effect on the relative price for defenders in the Premier League, and for forwards and defenders in the Ligue 1. So, the average results presented in table 4 are computed from the largely varying results in table 5. When we use a dataset that would have had more attackers, the results would presumably be very different from what they are now in table 4. So, the model is very sensitive to the data. However, as previously argued, the ratio of the positions is likely to be very similar during each transfer window.

It is also interesting to look at the coefficients of the nationality variable in table 5. Where on average there was no sign of an English Tax, we now observe that for defenders, the relative price is significantly higher when the player is also English. This would mean that there is an English Tax for defenders. For the French league, we see that there is still no significant increase in relative price for French players. We do, however, observe that the relative price of French midfielders is significantly lower compared to foreign midfielders.

Table 5: Linear regressions per position using log(Fee) as dependent variable

	Premier League			Ligue 1		
LogFee	Forwards	Midfielders	Defenders	Forwards	Midfielders	Defenders
Winter	-0.201	-0.477**	-2.676***	-0.494	0	-1.273**
Transfer	(0.397)	(0.189)	(0.446)	(0.933)	(0.426)	(0.6)
Contract	0.027**	0.031***	0.033**	0.027	0.029**	0.007
	(0.11)	(0.008)	(0.13)	(0.014)	(0.11)	(0.013)
Nationality	-0.113	0.557	0.983**	-0.312	-0.548**	-0.613
	(0.312)	(0.337)	(0.371)	(0.495)	(0.221)	(0.417)
Competition	0.47	0.329	-0.159	0.649	0.245	0.38
	(0.416)	(0.367)	(0.305)	(0.415)	(0.304)	(0.322)
Quality	0.039	0.021	0.059**	0.121**	0.021	0.071**
	(0.027)	(0.021)	(0.021)	(0.044)	(0.028)	(0.026)
Form	0.138	0.298	0.481	-0.034	0.113	-0.304
	(0.343)	(0.247)	(0.575)	(0.606)	(0.342)	(0.655)
Min.in Comp.	0	-0.001**	0	0	-0.001	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Min. in Total	0	0.001***	-0.001*	0	0	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AgeGroup2	-0.445	0.279	0.428	-0.39	-0.404	-0.767
	(0.537)	(0.347)	(0.279)	(0.398)	(0.392)	(0.489)
AgeGroup3	-0.35	0.03	0.996	-0.381	-0.732*	-1.909***
	(0.409)	(0.49)	(0.588)	(0.445)	(0.417)	(0.456)
AgeGroup4	-2.033***	.	-0.113	.	.	-1.438
	(0.4)		(0.813)			(1.241)
2019	1.026*	-0.367	0.04	-0.554	0.626	-0.005
	(0.565)	(0.352)	(0.839)	(0.48)	(0.362)	(0.509)
2020	0.416	-0.256	-0.141	0.096	0.029	-0.249
	(0.546)	(0.338)	(0.683)	(0.611)	(0.535)	(0.394)
2021	0.775	-0.117	0.218	-0.544	0.276	-0.292
	(0.554)	(0.492)	(0.688)	(0.509)	(0.442)	(0.429)
2022	0.737	-0.144	0.482	-0.262	0.944**	-0.012
	(0.564)	(0.265)	(0.681)	(0.493)	(0.364)	(0.334)
2023	0.397	0.048	3.761***	1.128	-1.262**	0.399
	(0.564	(0.518)	(0.777)	(1.026)	(0.454)	(0.719)
Constant	-1.662	-1.655	-5.256	-7.826	-0.358	-1.185
	(2.968)	(2.618)	(3.453)	(4.378)	(3.086)	(5.175)
Obs.	28	27	41	35	34	34
R-Squared	0.813	0.891	0.768	0.735	0.675	0.794

Notes: All year variables, the competition variable, the winter transfer variable, the age variables and the Nationality variable are dummies. Age group 1 = 17-21 for the Premier League and 18-22 for the Ligue 1, Age group 2 = 22-26 for the Premier League and 23-26 for the Ligue 1, Age group 3 = 27-31 for both and Age group 4 = 32+ for both. The transfer fee is measured in millions. The remaining contract is measured in months. Quality and Form are scores between 0 and 100 and 0 and 10 respectively. All other variables are continuous. Standard errors between parentheses.
**** p<0.01, ** p<0.05, * p<0.1*

VII. Conclusion & Discussion

In this paper, I have researched the so-called English Tax by running a regression on transfer data of the English Premier League and a comparison regression on transfer data of the French Ligue 1 over the same time period. The results in the regressions were promising in favour of the English Tax when using the absolute transfer fee in the regression model: being English is associated with a higher absolute transfer fee, on average, for moves to the Premier League and being French has no significant correlation with the absolute transfer fee, on average, for moves to the Ligue 1. As discussed in the theory segment of this paper, this suggests the existence of an English Tax. However, I also stated that the model that used the logarithm of the transfer fee suited the data better and did not show any evidence for an English Tax. These conclusions are drawn on the assumption that the model I used is complete and does not suffer severely from omitted variable bias. However, the model might not be complete and one variable that could be missing is 'work ethic'. Work ethic is possibly correlated with nationality and with the transfer fee of a player. Even though the model might not be complete, I believe that it does show great evidence for the absence English Tax and therefore I believe I cannot reject either of my hypotheses: English top teams do not pay higher fees for English players and this phenomenon can be generalized to other countries and divisions.

As mentioned before, there might be arguments against the validity of the Conditional Independence Assumption, meaning that my estimations would suffer from the following problem: omitted variable bias. Even though I believe that my results are significant, it might be possible to create a more convincing model. An ideal setting would be a dataset of players that are on average the same in every possible characteristic, apart from their nationality. As this is rather impossible, it might be a good idea to focus on transfers of players in a smaller geographical region. For example, one can focus only on the transfers of players within the United Kingdom. This way the sample will have on average more similar characteristics than when one also adds players from Africa for example. However, gathering enough data for this example will be a tough challenge.
In this paper, I compared the English and the French top divisions, and found no convincing evidence for a Tax in both competitions, but there are many more leagues and they can all differ in whether they have a Domestic Tax or not. Perhaps the French and English top divisions are unique in the fact that they do not pay higher fees for French and English players. Therefore, I would recommend further research on this very topic and broadening it to many different leagues in many different countries.

VIII. Appendix

Assumptions and tests for the linear regression model

1. Linearity

To check for obvious deviations from linearity, I plotted all independent variables with the dependent variable on a scatterplot to observe this manually. The age was an obvious deviation from linearity, as shown in figures 3 and 4, and I solved this problem by making different age group variables.

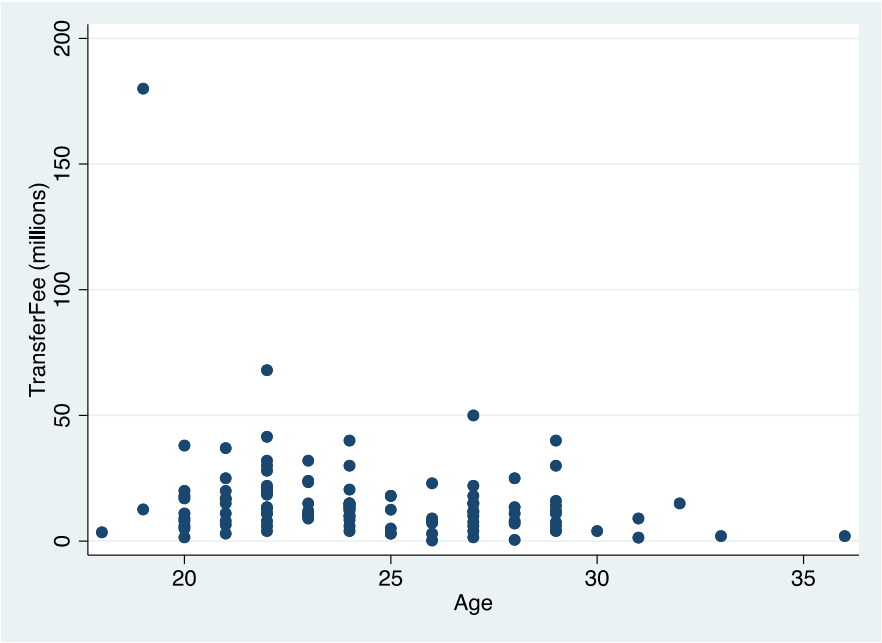

Figure 3: Scatterplot of Age and Transfer Fee using Premier League data.

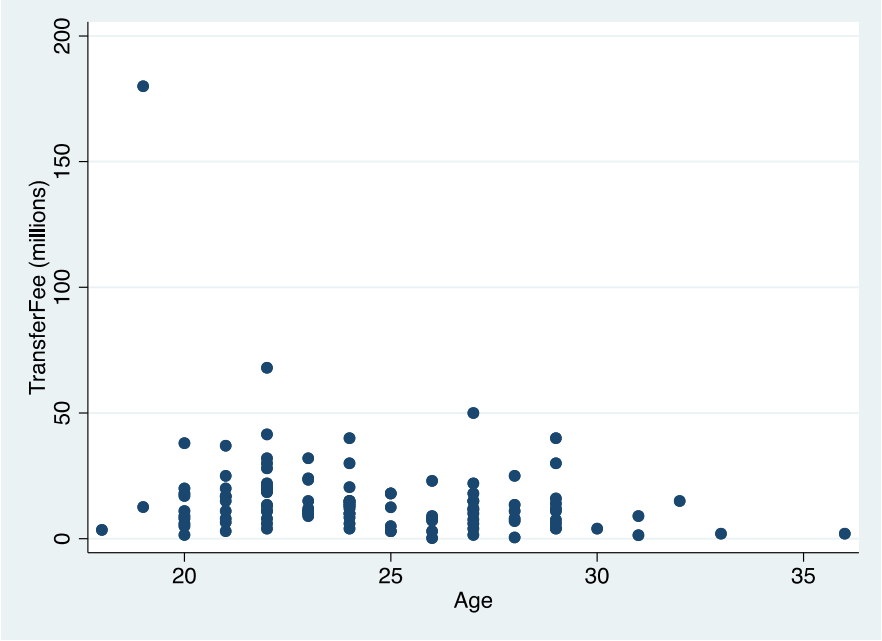

Figure 4: Scatterplot of Age and Transfer Fee using Ligue 1 data.

Apart from the age variable, all other variables suited linearity, meaning that this assumption holds.

2. Random sample

The samples that I have drawn are all transfers from the past five years, I would argue that there is some sort of random factor in there, but the data is recent and therefore it may not be representative for transfers that happened longer ago, but this is of no concern as I am only interested in the present and this is best explained by the near past.

3. Multicollinearity

When looking at the regression functions I and II, there is collinearity between some variables. There is collinearity between the position dummies and between the age dummies. However, this is of no worry, as STATA automatically removes one of the dummies to solve the collinearity problem.

4. Exogeneity

The model uses many independent variables, and even though it is probably not enough to make the model completely exogenous, I do believe that I use the most important factors that drive the price of a football player and thus, I believe that my results can be interpreted as such. Another reason is that the many factors that are difficult to translate into data or even unobservable, are the same for sporting directors that manage the club's transfer policy and so these directors will also not take into account all factors.

5. Homoscedasticity

In order for a dataset to be homoscedastic, the variance should be the same for variables in different groups. In each dataset I have 2 groups: the nationals (English or French, depending on the dataset) and foreigners. I check variance for these groups for every variable using Levene's test.

Table 6: Levene's test for equal variances for Premier League data.

	Mean			
	Foreign (0)	English (1)	Levene Statistic	P-value
Winter Transfer	0.16	0.13	0.51	0.48
Transfer Fee	37.57	43.48	1.58	0.21
Contract	27.65	25.56	0.00	0.98
Competition	0.16	0.5	13.63	0.00***
Quality	77.74	75.56	0.02	0.87
Form	7.07	6.86	1.36	0.25
Min. in Comp.	2086.55	1984.56	2.48	0.12
Minutes Total	3101.61	2537.50	4.37	0.04**
Forward	0.27	0.25	0.15	0.70
Midfielder	0.27	0.19	2.56	0.11
Defender	0.39	0.44	0.39	0.53
Goalkeeper	0.07	0.13	2.25	0.14
Log(Year)	7.61	7.61	0.36	0.55
Age17-21	0.22	0.43	5.89	0.02**
Age22-26	0.57	0.44	0.01	0.94
Age-27-31	0.17	0.06	6.19	0.01**
Age32+	0.03	0.06	1.13	0.29
Obs.	88	16		

Notes: Equality of variances tested for the two groups: Foreign and English.

In table 6, we observe that some of the variances of the variables are unequal, like for some of the age groups, but most of them are equal. This still makes it difficult to draw a conclusion about the entire dataset and therefore I will look for equality of variance in the residuals. The residuals are the difference between the predicted values of my model as described in the methodology sector and the actual values of my dataset. The residuals are described by ε_{it} in the model.

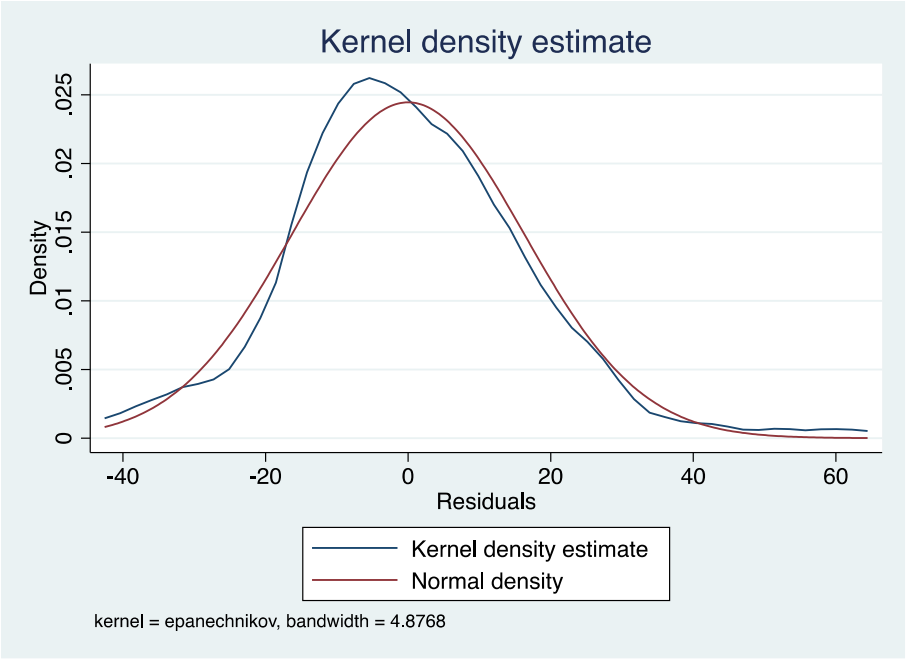

Figure 5: Kernel density estimate for the residuals of the Premier League data.

In figure 5, I created a Kernel density estimate and compare this to what a normal density would look like. The normal density reflects homoscedasticity. As seen in the figure, the estimate does not deviate all that much from the normal density: this would mean that there is indeed homoscedasticity. However, before we draw any definite conclusions, let's look at some other graphical representations of the data.

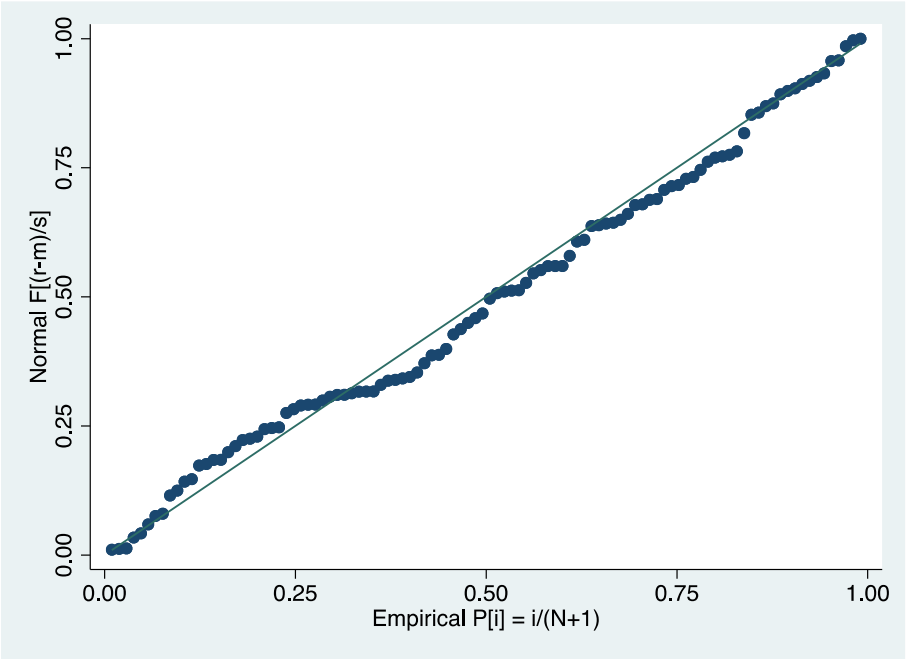

Figure 6: Standardized normal probabilities plot of the residuals of the Premier League data.

In figure 6, we see what the standardized normal probabilities would look like on the line, and we observe the probabilities plot of our own residuals. Again, there is not much deviation from the normal plot.

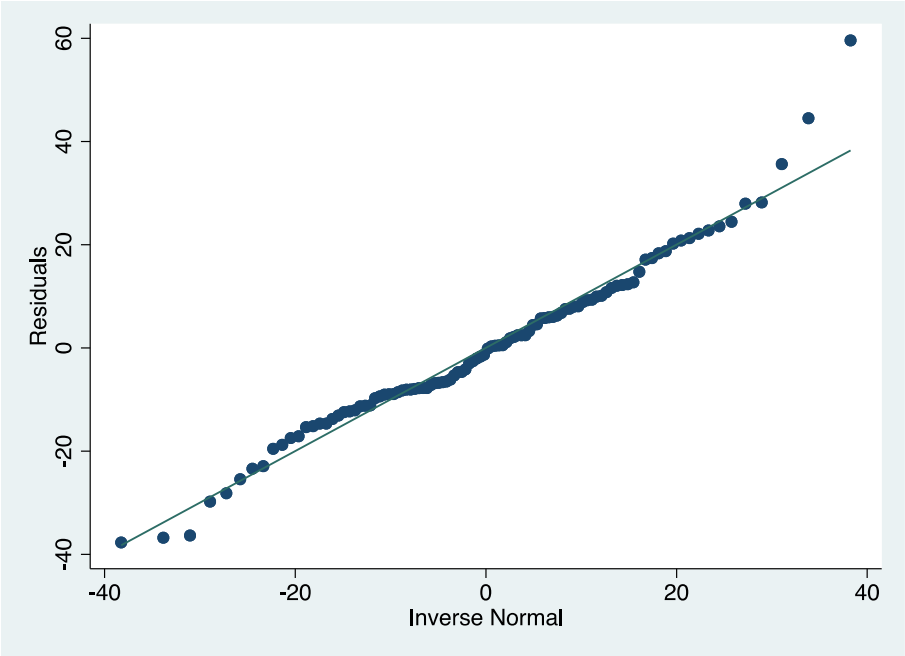

Figure 7: Quantiles of the residuals of the Premier League data against the quantiles of normal distribution.

Also, when looking at figure 7, where the quantiles of the residuals of our data are plotted against the quartiles of a normal distribution, we see little difference. This leads me to conclude that the residuals are normally distributed and therefore the data for the Premier League that is used for the regression, is homoscedastic.

Now, I will perform the same analysis for the data gathered for the Ligue 1.

Table 7: Levene's test for equal variances for Ligue 1 data

	Mean			
	Foreign	**French**	**Levene Statistic**	**P-value**
Winter Transfer	0.14	0.13	0.09	0.76
Transfer Fee	15.31	17.43	1.91	0.17
Contract	26.67	22.03	3.91	0.05*
Competition	0.17	0.69	8.88	0.00***
Quality	74.94	74.88	2.95	0.09*
Form	6.83	6.90	0.09	0.76
Min. in Comp.	1636.12	2123.56	0.16	0.69
Minutes Total	2302.54	2720.03	0.22	0.64
Forward	0.30	0.34	0.85	0.36
Midfielder	0.27	0.38	3.57	0.06*
Defender	0.32	0.25	2.51	0.12
Goalkeeper	0.11	0.03	8.40	0.00***
Log(Year)	7.61	7.61	0.23	0.63
Age18-22	0.33	0.44	2.73	0.10
Age23-26	0.37	0.25	7.68	0.01***
Age27-31	0.26	0.31	1.16	0.28
Age32+	0.04	0	5.23	0.02**
Obs.	81	32		

Notes: Equality of variances tested for the two groups: Foreign and French.

Again, from the tests in table 7, there is no clear conclusion that can be drawn about homoscedasticity and therefore I will expand the analysis the same way that I did for the Premier League.

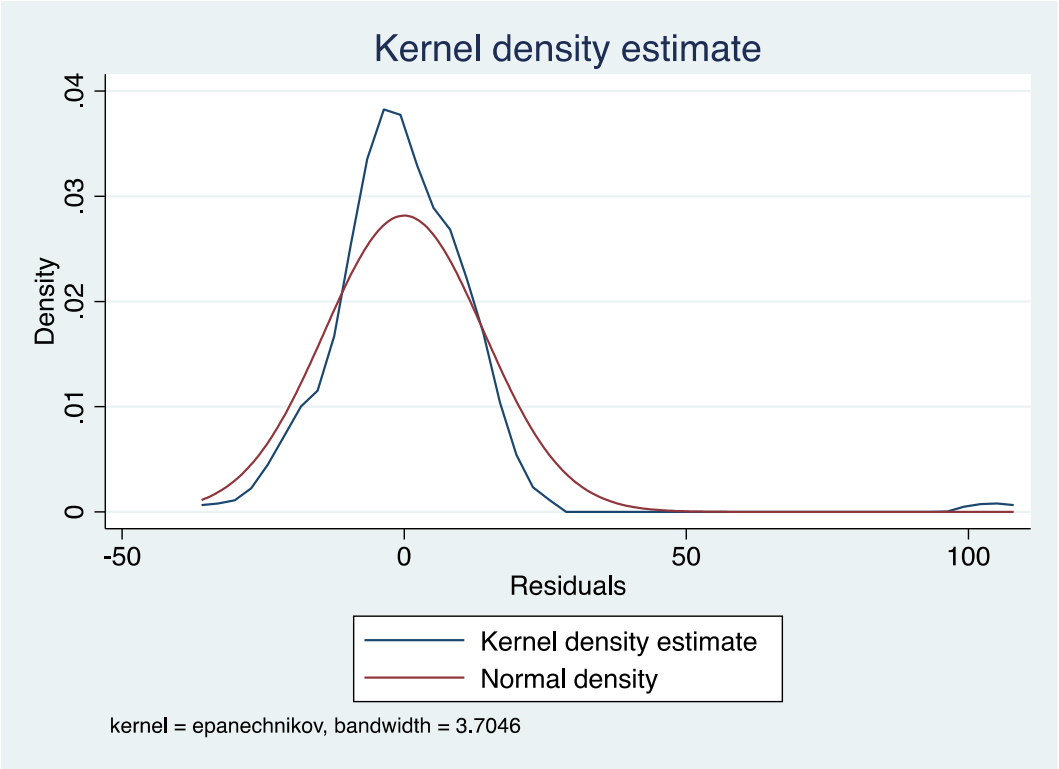

Figure 8: Kernel density estimate for the residuals of the Ligue 1 data.

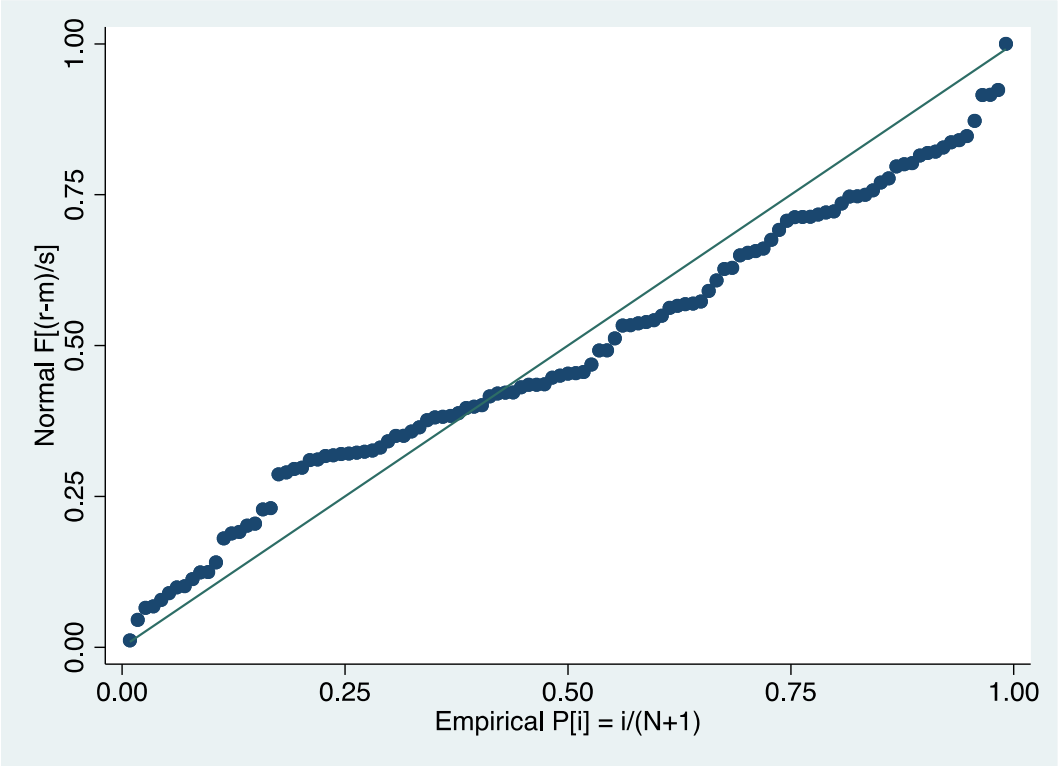

Figure 9: Standardized normal probabilities plot of the residuals of the Ligue 1 data.

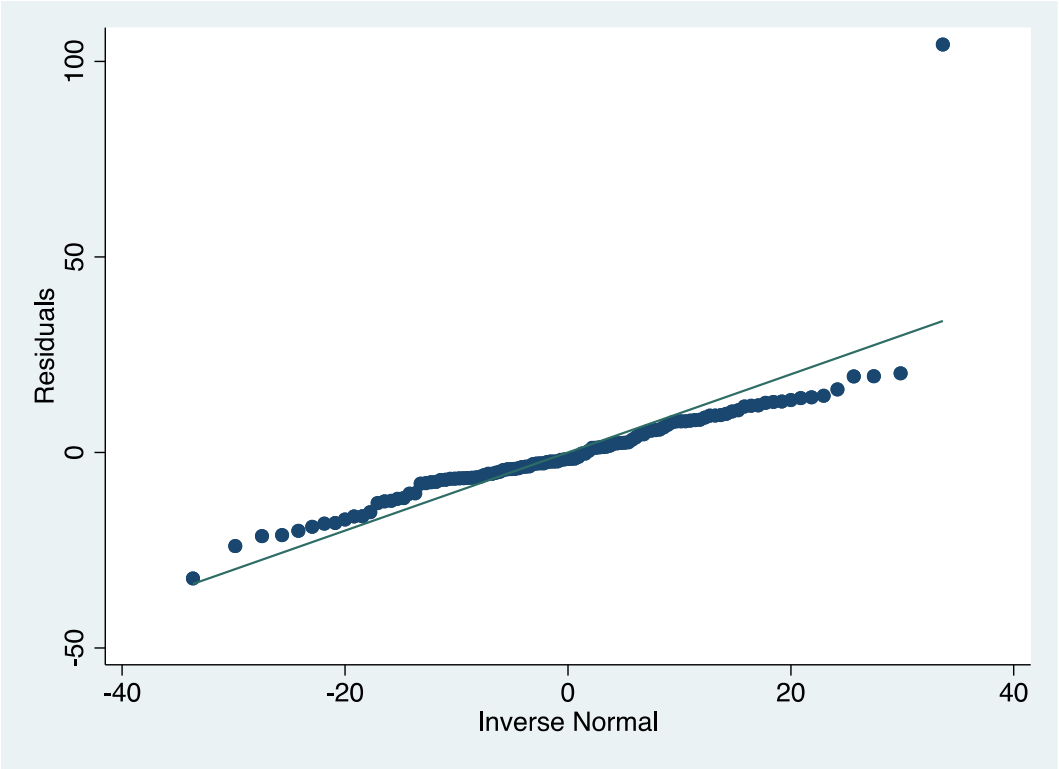

Figure 10: Quantiles of the residuals of the Ligue 1 data against the quantiles of normal distribution.

After observing figures 8, 9 and 10, I also conclude that the data for the Ligue 1 is homoscedastic. The residuals follow a more or less normal distribution.

IX. References

Data:

Footballdatabase : Stats de Joueurs, Transferts, Matchs, Clubs. (n.d.). https://footballdatabase.eu/

Transfermarkt. (n.d.). *Voetbaltransfers, transfergeruchten, marktwaarden & nieuws*. https://transfermarkt.nl/

Football Statistics | Football Live Scores | WhoScored.com. (n.d.). https://www.whoscored.com/

Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC. StataCorp. 2019.

UEFA Coëfficiënten Ranglijst 2022 / 2023 - Overzicht. (2023). UEFA Coëfficiënten. https://coefficienten.nl/uefa-coefficienten-ranglijst/

Literature:

Ferguson, D., Stewart, K., Jones, J. C. H., & Dressay, A. L. (1991). The Pricing of Sports Events: Do Teams Maximize Profit? *Journal of Industrial Economics*, *39*(3), 297. https://doi.org/10.2307/2098521

Fort, R., & Quirk, J. P. (2004). Owner Objectives and Competitive Balance. *Journal of Sports Economics*, *5*(1), 20–32. https://doi.org/10.1177/1527002503261275

Garcia-Del-Barrio, P., & Szymanski, S. (2009b). Goal! Profit Maximization Versus Win Maximization in Soccer. *Review of Industrial Organization*, *34*(1), 45–68. https://doi.org/10.1007/s11151-009-9203-6

Jones, J. C. H. (1969). The Economics of the National Hockey League. *Canadian Journal of Economics*, *2*(1), 1. https://doi.org/10.2307/133568

Kesenne, S. (2000). Revenue Sharing and Competitive Balance in Professional Team Sports. *Journal of Sports Economics*, *1*(1), 56–65. https://doi.org/10.1177/152700250000100105

Késenne, S., & Pauwels, W. (2006). Club Objectives and Ticket Pricing in Professional Team Sports. *Eastern Economic Journal 32*(1), 549-560.

Neale, W. C. (1964). The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition. *Quarterly Journal of Economics*, *78*(1), 1. https://doi.org/10.2307/1880543

Rascher, D. A. (1997). A Model of a Professional Sports League. *International Advances in Economic Research*, *3*(3), 327–328. https://doi.org/10.1007/bf02294925

Rottenberg, S. (1956). The Baseball Players' Labor Market. *Journal of Political Economy*, *64*(3), 242–258. https://doi.org/10.1086/257790

Sloane, P. J. (1971). Scottish Journal of Political Economy: The Economics of Proffessional Football: The Football Club as a Utility Maximiser. *Scottish Journal of Political Economy*, *18*(2), 121–146. https://doi.org/10.1111/j.1467-9485.1971.tb00979.x

Szymanski, S. (2003). The Economic Design of Sporting Contests. *Journal of Economic Literature*, *41*(4), 1137–1187. https://doi.org/10.1257/jel.41.4.1137

Szymanski, S., & Kesenne, S. (2004). Competitive Balance and Gate Revenue Sharing in Team Sports. *Journal of Industrial Economics*, *52*(1), 165–177. https://doi.org/10.1111/j.0022-1821.2004.00220.x

Vrooman, J. (1997). A Unified Theory of Capital and Labor Markets in Major League Baseball. *Southern Economic Journal*, *63*(3), 594. https://doi.org/10.2307/1061097

Non-scientific articles:

Ewanfcraven (2022) *The Misunderstanding of the 'English Tax' in the Transfer Market*. Graph Roots Football. https://graphroots.co.uk/2021/10/07/the-misunderstanding-of-the-english-tax-in-the-transfer-market/

Pereira D. (2021). *A Look at the "English Tax" Issue*. Over the Line Sports. https://overthelinesports.ca/f/a-look-at-the-english-tax-issue?blogcategory=NFL

Sub.editors. (2022). *The 'English Tax' in the Transfer Market: Reality or Myth?* Palatinate. https://www.palatinate.org.uk/the-english-tax-in-the-transfer-market-reality-or-myth/

What is a Premier League Squad List? (n.d.). https://www.premierleague.com/news/102383